# *Appendices*

# Appendix A
# DOE Human Genome Program: Early History, Enabling Legislation

A brief history of the U.S. Department of Energy (DOE) Human Genome Program will be useful in a discussion of the objectives of the DOE program as well as those of the collaborative U.S. Human Genome Project. The DOE Office of Biological and Environmental Research (OBER) of DOE and its predecessor agencies—the Atomic Energy Commission and the Energy Research and Development Administration—have long sponsored research into genetics, both in microbial systems and in mammals, including basic studies on genome structure, replication, damage, and repair and the consequences of genetic mutations. (See Appendix E for a discussion of the DOE Biological and Environmental Research Program.)

In 1984, OBER [then named Office of Health and Environmental Research (OHER)] and the International Commission on Protection Against Environmental Mutagens and Carcinogens cosponsored a conference in Alta, Utah, which highlighted the growing roles of recombinant DNA technologies. Substantial portions of the meeting's proceedings were incorporated into the Congressional Office of Technology Assessment report, *Technologies for Detecting Heritable Mutations in Humans*, in which the value of a reference sequence of the human genome was recognized.

Acquisition of such a reference sequence was, however, far beyond the capabilities of biomedical research resources and infrastructure existing at that time. Although the small genomes of several microbes had been mapped or partially sequenced, the detailed mapping and eventual sequencing of 24 distinct human chromosomes (22 autosomes and the sex chromosomes X and Y) that together comprise an estimated 3 billion subunits was a task some thousandsfold larger.

DOE OHER was already engaged in several multidisciplinary projects contributing to the nation's biomedical capabilities, including the GenBank DNA sequence repository, which was initiated and sustained by DOE computer and data-management expertise. Several major user facilities supporting microstructure research were developed and are maintained by DOE. Unique chromosome-processing resources and capabilities were in place at Los Alamos National Laboratory and Lawrence Livermore National Laboratory. Among these were the fluorescence-activated cell sorter (called FACS) systems to purify human chromosomes within the National Laboratory Gene Library Project for the production of libraries of DNA clones. The availability of these monochromosomal libraries opened an important path—a practical means of subdividing the huge total genome into 24 much more manageable components.

With these capabilities, OHER began in 1986 to consider the feasibility of a dedicated human genome program. Leading scientists were invited to the March 1986 international conference at Santa Fe, New Mexico, to assess the desirability

## Enabling Legislation

In the United States, the first federal support for genetics research was through the Atomic Energy Commission. In the early days of nuclear energy development, the focus was on radiation effects and later broadened under the Energy Research and Development Administration (ERDA) and the Department of Energy to include the health implications of all energy technologies and their by-products. Major enabling legislation follows.

**Atomic Energy Act of 1946** (P.L. 79-585): Provided the initial charter for a comprehensive program of research and development related to the utilization of fissionable and radioactive materials for medical, biological, and health purposes.

**Atomic Energy Act of 1954** (P.L. 83-703): Further authorized AEC "to conduct research on the biologic effects of ionizing radiation."

**Energy Reorganization Act of 1974** (P.L. 93-438): Provided that responsibilities of ERDA should include "engaging in and supporting environmental, biomedical, physical, and safety research related to the development of energy resources and utilization technologies."

**Federal Non-Nuclear Energy Research and Development Act of 1974** (P.L. 93-577): Authorized ERDA to conduct a comprehensive non-nuclear energy research, development, and demonstration program to include the environmental and social consequences of the various technologies.

**DOE Organization Act of 1977** (P.L. 95-91): Instructed the department "to assure incorporation of national environmental protection goals in the formulation and implementation of energy programs; and to advance the goal of restoring, protecting, and enhancing environmental quality, and assuring public health and safety," and to conduct "a comprehensive program of research and development on the environmental effects of energy technology and programs."

and feasibility of implementing such a project. With virtual unanimity, participants agreed that ordering and eventually sequencing DNA clones representing the human genome were desirable and feasible goals. With the receipt of this enthusiastic response, OHER initiated several pilot projects. Program guidance was further sought from the DOE Health Effects Research Advisory Committee (HERAC).

## HERAC Recommendation

The April 1987 HERAC report recommended that DOE and the nation commit to a large, multidisciplinary scientific and technological undertaking to map and sequence the human genome. DOE was particularly well suited to focus on resource and technology development, the report noted; HERAC further recommended a leadership role for DOE because of its demonstrated expertise in managing complex and long-term multidisciplinary projects involving both the development of new technologies and the coordination of efforts in industries, universities, and its own laboratories.

Evolution of the nation's Human Genome Project further benefited from a 1988 study by the National Research Council (NRC) entitled *Mapping and Sequencing the Human Genome*, which recommended that the United States support this research effort and presented an outline for a multiphase plan.

## DOE and NIH Coordination

The National Institutes of Health (NIH) was a necessary participant in the large-scale effort to map and sequence the human genome because of its long history of support for biomedical research and its vast community of scientists. This was confirmed by the NRC report, which recommended a major role for NIH. In 1987, under the leadership of Director James Wyngaarden, NIH established the Office of Genome Research in the Director's Office. In 1988, DOE and NIH signed a Memorandum of Understanding in which the agencies agreed to work together, coordinate technical research and activities, and share results. In 1990, DOE and NIH submitted a joint research plan outlining short- and long-term goals of the project.

# Appendix B
# DOE-NIH Guidelines for Sharing Data and Resources

*At its December 7, 1992, meeting, the DOE-NIH Joint Subcommittee on the Human Genome approved the following sharing guidelines, developed from the DOE draft of September 1991.**

The information and resources generated by the Human Genome Project have become substantial, and the interest in having access to them is widespread. It is therefore desirable to have a statement of philosophy concerning the sharing of these resources that can guide investigators who generate the resources as well as those who wish to use them.

A key issue for the Human Genome Project is how to promote and encourage the rapid sharing of materials and data that are produced, especially information that has not yet been published or may never be published in its entirety. Such sharing is essential for progress toward the goals of the program and to avoid unnecessary duplication. It is also desirable to make the fruits of genome research available to the scientific community as a whole as soon as possible to expedite research in other areas.

Although it is the policy of the Human Genome Project to maximize outreach to the scientific community, it is also necessary to give investigators time to verify the accuracy of their data and to gain some scientific advantage from the effort they have invested. Furthermore, in order to assure that novel ideas and inventions are rapidly developed to the benefit of the public, intellectual property protection may be needed for some of the data and materials.

After extensive discussion with the community of genome researchers, the advisors of the NIH and DOE genome programs have determined that consensus is developing around the concept that a 6-month period from the time the data or materials are generated to the time they are made available publicly is a reasonable maximum in almost all cases. More rapid sharing is encouraged.

Whenever possible, data should be deposited in public databases and materials in public repositories. Where appropriate repositories do not exist or are unable to accept the data or materials, investigators should accommodate requests to the extent possible.

The NIH and DOE genome programs have decided to require all applicants expecting to generate significant amounts of genome data or materials to describe in their application how and when they plan to make such data and materials available to the community. Grant solicitations will specify this requirement. These plans in each application will be reviewed in the course of peer review and by staff to assure they are reasonable and in conformity with program philosophy. If a grant is made, the applicant's sharing plans will become a condition of the award and compliance will be reviewed before continuation funding is provided. Progress reports will be asked to address the issue.

---

*Reprinted from *Human Genome News* **4**(5), 4 (1993).

# Appendix C
# NIH-DOE Guidance on Human Subjects Issues in Large-Scale DNA Sequencing

*Date issued: August 9, 1996*

## Introduction

The Human Genome Project (HGP) is now entering into large-scale DNA sequencing. To meet its complete sequencing goal, it will be necessary to recruit volunteers willing to contribute their DNA for this purpose. The guidance provided in this document is intended to address ethical issues that must be considered in designing strategies for recruitment and protection of DNA donors for large-scale sequencing.

Nothing in this document should be construed to differ from, or substitute for, the policies described in the Federal Regulations for the Protection of Human Subjects [45CFR46 (NIH) and 10CFR745 (DOE)]. Rather, it is intended to supplement those policies by focusing on the particular issues raised by large-scale human DNA sequencing. This statement addresses six topics: (1) benefits and risks of genomic DNA sequencing; (2) privacy and confidentiality; (3) recruitment of DNA donors as sources for library construction; (4) informed consent; (5) IRB approval; and (6) use of existing libraries.

The guidance provided in this statement is intended to afford maximum protection to DNA donors and is based on the belief that protection can best be achieved by a combination of approaches including:

- ensuring that the initial version of the complete human DNA sequence is derived from multiple donors;
- providing donors with the opportunity to make an informed decision about whether to contribute their DNA to this project; and
- taking effective steps to ensure the privacy and confidentiality of donors.

### 1. Benefits and Risks of Genomic DNA Sequencing

The HGP offers great promise for the improvement of human health. As a consequence of the HGP, there will be a more thorough understanding of the genetic bases of human biology and of many diseases. This, in turn, will lead to better therapies and, perhaps more importantly, prevention strategies for many of those diseases. Similarly, as the technology developed by the HGP is applied to understanding the biology of other organisms, many other human activities will be affected including agriculture, environmental management, and biologically based industrial processes.

While the HGP offers great promise to humanity, there will be no direct benefit, in either clinical or financial terms, to any of the individuals who choose to donate DNA for large-scale sequencing. Rather, the motivation for donation is likely to be an altruistic willingness to contribute to this historic research effort.

However, individuals who donate DNA to this effort may face certain risks. Information derived from the donors will become available in public databases. Such information may reveal, for example, DNA sequence-based information about disease susceptibility. If the donor becomes aware of such information, it could lead to emotional distress on her/his part. If such health-related information becomes known to others, discrimination against the donor (e.g., in insurance or in employment) could result. Unwanted notoriety is another potential risk to donors. Therefore, those engaged in large-scale sequencing must be sensitive to the unique features of this type of research and ensure that both the protections normally afforded research subjects and the special issues associated with human genomic DNA sequencing are thoroughly addressed.

While some risks to donors can already be identified, the probability of adverse events materializing appears to be low. However, the risks of harm to individuals will increase if confidentiality is not maintained and/or the number of donors is limited to a very few individuals. Either, or both, of these situations would increase the possibility of a donor's identity being revealed without his/her knowledge or permission.

A final issue to consider is characterized in a statement taken from the OPRR Guidebook[1] which points out that "some areas [of genetic research] present issues for which no clear guidance can be given at this point, either because enough is not known about the risks presented by the research, or because no consensus on the appropriate resolution of the problem exists." It is anticipated that the DNA sequence information produced by the Human Genome Project will be used in the future for types of research which cannot now be predicted and the risks of which cannot be assessed or disclosed.

### 2. Privacy and Confidentiality

In general, one of the most effective ways of protecting volunteers from the unexpected, unwelcome or unauthorized use of information about them is to ensure that there are no opportunities for linking an individual donor with information about him/her that is revealed by the research. By not collecting information about the identity of a research subject and any biological material or records developed in the course of the research, or by subsequently removing all

identifiers ("anonymizing" the sample), the possibility of risk to the subject stemming from the results of the research is greatly reduced. Large-scale DNA sequence determination represents an exception because each person's DNA sequence is unique and, ultimately, there is enough information in any individual's DNA sequence to absolutely identify her/him. However, the technology that would allow the unambiguous identification of an individual from his/her DNA sequence is not yet mature. Thus, for the foreseeable future, establishing effective confidentiality, rather than relying on anonymity, will be a very useful approach to protecting donors.

Investigators should introduce as many disconnects between the identity of donors and the publicly available information and materials as possible. There should not be any way for anyone to establish that a specific DNA sequence came from a particular individual, other than resampling an individual's DNA and comparing it to the sequence information in the public database. In particular, no phenotypic or demographic information about donors should be linked to the DNA to be sequenced.[2] For the purposes of the HGP such information will rarely be useful, and recording such information could result in possible misuse and compromise donor confidentiality.

Confidentiality should be "two way." Not only should others be unable to link a DNA sequence to a particular individual, but no individual who donates DNA should be able to confirm directly that a particular DNA sequence was obtained from their DNA sample.[3] This degree of confidentiality will preclude the possibility of re-contacting DNA donors, providing another degree of protection for them. It should be clear to both investigators and to donors that the contact involved in obtaining the initial specimen will be the only contact.[4]

Another approach for protecting all DNA donors is to reduce the incentive for wanting to know the identities of particular donors. If the initial human sequence is a "mosaic" or "patchwork" of sequenced regions derived from a number of different individuals, rather than that of a single individual, there would be considerably less interest in who the specific donors were. Although there may be scientific justification that each clone library used for sequencing should be derived from one person, there is no scientific reason that the entire initial human DNA sequence should be that of a single individual. As approximately 99.9% of the human DNA sequence is common between any two individuals, most of the fundamental biological information contained in the human DNA sequence is common to all people.

To increase the likelihood that the first human DNA sequence will be an amalgam of regions sequenced from different sources, a number of clone libraries must be made available. Although a number of large insert libraries have been made, most do not meet all of the standards set in this document; therefore, these libraries should be used as substrates for large-scale sequencing only under circumscribed conditions (see section 6, p. 79). Starting immediately, new libraries will be developed that have the advantage of being constructed in accordance with the ethical principles discussed in this document; they may also confer some additional scientific benefit. Such libraries are critical for the long-range needs of the HGP.

## 3. Source/Recruitment of DNA Donors for Library Construction

Another implication of the fact that 99.9% of the human DNA sequence is shared by any two individuals is that the backgrounds of the individuals who donate DNA for the first human sequence will make no scientific difference in terms of the usefulness and applicability of the information that results from sequencing the human genome. At the same time, there will undoubtedly be some sensitivity about the choice of DNA sources. There are no scientific reasons why DNA donors should not be selected from diverse pools of potential donors.[5]

There are two additional issues that have arisen in considering donor selection. These warrant particular discussion:

- It is recognized that women have historically been underrepresented in research, so it can be anticipated that concerns might arise if males (sperm DNA) were used exclusively as the source of DNA for large-scale sequencing. Although there would be no scientific basis for concern, because even in the case of a male source, half of the donor's DNA would have come from his mother and half from his father, nevertheless perceptions are not to be dismissed. While the choice of donors will not be dictated to investigators, it is expected that, because multiple libraries will be produced, a number of them will be made from female sources while others will be made from male sources.
- Staff of laboratories involved in library construction and DNA sequencing may be eager to volunteer to be donors because of their interest and belief in the HGP. However, proximity to the research may create some special vulnerabilities for laboratory staff members. It is also possible that they will feel pressure to donate and there may be an increased likelihood that confidentiality would be breached. Finally, there is a potential that the choice of persons so closely involved in the research may be interpreted as elitist. For all of these reasons, it is recommended that donors should not be recruited from laboratory staff, including the principal investigator.

## 4. Informed Consent

Obtaining informed consent specifically for the purpose of donating DNA for large-scale sequencing raises some unique concerns. Because anonymity cannot be guaranteed and confidentiality protections are not absolute, the disclosure process to potential donors must clearly specify what the process of DNA donation involves, what may make it different from other types of research, and what the implications are of one's DNA sequence information being a public scientific resource.

Federal regulations (45CFR46 and 10CFR745) require the disclosure of a number of issues in any informed consent document. They include such issues as potential benefits of the research, potential risks to the donor, control and ownership of donated material, long-term retention of donated material for future use, and the procedures that will be followed. In addition, there are several other disclosures that are of special importance for donors of DNA for large-scale sequencing. These include:

- the meaning of confidentiality and privacy of information in the context of large-scale DNA sequencing, and how these issues will be addressed;
- the lack of opportunity for the donor to later withdraw the libraries made from his/her DNA or his/her DNA sequence information from public use;
- the absence of opportunity for information of clinical relevance to be provided to the donor or her/his family;
- the possibility of unforeseen risks; and
- the possible extension of risk to family members of the donor or to any group or community of interest (e.g., gender, race, ethnicity) to which a donor might belong.

Many academic human genetics units have considerable experience in dealing with research subjects and obtaining informed consent, while the laboratories that are likely to be involved in making the libraries for sequencing have, in general, much less experience of this type. Therefore, library makers are encouraged to establish a collaboration with one or more human genetics units, with the latter being responsible for recruiting donors, obtaining informed consent, obtaining the necessary biological samples, and providing a blinded sample to the library maker. Collaboration with tissue banks may be considered as long as these banks are collecting tissues in accordance with this guidance. The library maker should have no contact with the donor and no opportunity to obtain any information about the donor's identity.

## 5. IRB Approval

Effective immediately, projects to construct libraries for large-scale DNA sequencing must obtain Institutional Review Board (IRB) approval before work is initiated. IRBs should carefully consider the unique aspects of large-scale sequencing projects. Some of the informed consent provisions outlined may be somewhat at odds with the usual and customary disclosures found in most protocols involving human subjects and which IRBs usually consider. For example, research subjects usually are given the opportunity to withdraw from a research project if they change their minds about participating. In the case of donors for large-scale sequencing, it will not be possible to withdraw either the libraries made from their DNA or the DNA sequence information obtained using those libraries once the information is in the public domain. By the time a significant amount of DNA sequence data has been collected, the libraries, as well as individual clones from them, will have been widely distributed and the sequence information will have been deposited in and distributed from public databases. In addition, there will be no possibility of returning information of clinical relevance to the donor or his/her family.

## 6. Use of Existing Libraries for Large-Scale Sequencing

Many of the existing libraries (including those derived from anonymous donors) were not made in complete conformity with the principles elaborated above. The potential risks that may result from their use will be minimized by the rapid introduction of several new libraries constructed in accordance with this guidance, which NCHGR and DOE are taking steps to initiate. This will ensure that the existing libraries will only contribute small amounts to the first complete human DNA sequence. In the interim, existing libraries can continue to be used for large-scale sequencing, only if IRB approval and consent for "continued use" are obtained[6] and approval by the funding agency is granted.

It is important that in obtaining consent for contined use of existing libraries, no coercion of the DNA donor occur. It is therefore recommended that consideration be given to whether it is appropriate for the individual who previously recruited the donor to recontact him/her to obtain this consent. In some cases an IRB may determine that the recontact should be made by a third party to assure that the donors are fully informed and allowed to choose freely whether their DNA can continue to be used for this purpose.

# Conclusion

This document is intended to provide guidance to investigators and IRBs who are involved in large-scale sequencing efforts. It is designed to alert them to special ethical concerns that may arise in such projects. In particular, it provides guidance for the use of existing and the construction of new DNA libraries. Adhering to this guidance will ensure that the initial version of the complete human sequence is derived from multiple, diverse donors; that donors will have the opportunity to make an informed decision about whether to contribute their DNA to this project; and that effective steps will be taken by investigators to ensure the privacy and confidentiality of donors.

Investigators funded by NCHGR and DOE to develop new libraries for large-scale human DNA sequencing will be required to have their plans for the recruitment of DNA donors, including the informed consent documents, reviewed and approved by the funding agency before donors are recruited. Investigators involved in large-scale human sequencing will also be asked to observe those aspects of this guidance that pertain to them.

Approved August 17, 1996, by:

Francis S. Collins, M.D., Ph.D., Director, National Center for Human Genome Research, National Institutes of Health

Aristides N. Patrinos, Ph.D., Associate Director, Office of Health and Environmental Research, U.S. Department of Energy

## Footnotes

1. Office of Protection from Research Risks, Protecting Human Research Subjects: Institutional Review Board Guidebook (OPRR: U.S. Government Printing Office, 1993).

2. It is recognized that it will be trivially easy to determine the sex of the donor of the library, by assaying for the presence or absence of Y chromosome in the library.

3. There are a number of approaches to preventing a DNA donor from knowing that his/her DNA was actually sequenced as part of the HGP. For example, each time a clone library is to be made, an appropriately diverse pool of between five and ten volunteers can be chosen in such a way that none of them knows the identity of anyone else in the pool. Samples for DNA preparation and for preparation of a cell line can be collected from all of the volunteers (who have been told that their specimen may or may not eventually be used for DNA sequencing) and one of those samples is randomly and blindly selected as the source actually used for library construction. In this way, not only will the identity of the individual whose DNA is chosen not be known to the investigators, but that individual will also not be sure that s/he is the actual source.

4. Although recontacting donors should not be possible, investigators will potentially want to be able to resample a donor's genome. Thus, at the time the initial specimen is obtained, in addition to making a clone library representing the donor's genome, it should also be used to prepare an additional aliquot of high molecular weight DNA for storage and a permanent cell line. Either resource could then be used as a source of the donor's genome in case additional DNA were needed or comparison with the results of the analysis of the cloned DNA were desired.

5. There has been discussion in the scientific community about the sex of DNA donors. A library prepared from a female donor will contain DNA from the X chromosome in an amount equivalent to the autosomes, but will completely lack Y chromosomal DNA. Conversely, a library prepared from a male donor will contain Y DNA, but both X and Y DNA will only be present at half the frequency of the DNA from the other chromosomes. Scientifically, then, there are both advantages and disadvantages inherent in the use of either a male or a female donor. The question of the sex of the donor also involves the question of the use of somatic or germ line DNA to make libraries. For making libraries, useful amounts of germ line DNA can only be obtained from a male source (i.e., from sperm); it is not possible to obtain enough ova from a female donor to isolate germ line DNA for this purpose. Opinion is divided in the scientific community about whether germ line or somatic DNA should be used for large-scale sequencing. Somatic DNA is known to be rearranged, relative to germ line DNA, in certain regions (e.g., the immunoglobulin genes) and the possibility has been raised that other developmentally based rearrangements may occur, although no example of the latter has been offered. While some believe that the sequence product should not contain any rearrangements of this sort, others consider this potential advantage of germ line DNA to be relatively minor in comparison to the need to have the X chromosome fully represented in sequencing efforts and prefer the use of somatic DNA.

6. Individuals whose DNA was used for library construction (with the exception of those created from deceased or anonymous individuals) should be fully informed about the risks and benefits described above, should freely choose whether they would like their DNA to continue to be used for this purpose, and their decision should be documented.

# Executive Summary of Joint NIH-DOE Human Subjects Guidance

1. Those engaged in large-scale sequencing must be sensitive to the unique features of this type of research and ensure that both the protections normally afforded research subjects and the special issues associated with human genomic DNA sequencing are thoroughly addressed.

2. For the foreseeable future, establishing effective confidentiality, rather than relying on anonymity, will be a very useful approach to protecting donors.

3. Investigators should introduce as many disconnects between the identity of donors and the publicly available information and materials as possible.

4. No phenotypic or demographic information about donors should be linked to the DNA to be sequenced.

5. There are no scientific reasons why DNA donors should not be selected from diverse pools of potential donors.

6. While the choice of donors will not be dictated to investigators, it is expected that, because multiple libraries will be produced, a number of them will be made from female sources while others will be made from male sources.

7. It is recommended that donors should not be recruited from laboratory staff, including the principal investigator.

8. The disclosure process to potential donors must clearly specify what the process of DNA donation involves, what may make it different from other types of research, and what the implications are of one's DNA sequence information being a public scientific resource.

9. Library makers are encouraged to establish a collaboration with one or more human genetics units [or tissue banks].

10. The library maker should have no contact with the donor and no opportunity to obtain any information about the donor's identity.

11. Effective immediately, projects to construct libraries for large-scale DNA sequencing must obtain Institutional Review Board (IRB) approval before work is initiated.

12. Existing libraries can continue to be used for large-scale sequencing, only if IRB approval and consent for continued use are obtained and approval by the funding agency is granted.

13. It is important that in obtaining informed consent for continued use of existing libraries, no coercion of the DNA donor occur.

# Appendix D
# Human Genome Project and Genetics on the World Wide Web

**August 1997**

The World Wide Web offers the easiest path to information about the Human Genome Project and related genetics topics. Some useful sites to visit are included in the list below.

## Human Genome Project

**DOE Human Genome Program**
*http://www.er.doe.gov/production/ober/hug_top.html*

Devoted to the DOE component of the U.S. Human Genome Project and to the DOE Microbial Genome Program. Links to many other sites.

**Human Genome Project Information**
*http://www.ornl.gov/hgmis*

Comprehensive site covering topics related to the U.S. and worldwide Human Genome Projects. Useful for updating scientists and providing educational material for nonscientists, in support of DOE's commitment to public education. Developed and maintained for DOE by the Human Genome Management Information System (HGMIS) at Oak Ridge National Laboratory.

**NIH National Human Genome Research Institute**
*http://www.nhgri.nih.gov*

Site of the NIH sector of the U.S. Human Genome Project.

## DOE Human Genome Program Publications

***Human Genome News**
*http://www.ornl.gov/hgmis/publicat/publications.html*

Quarterly newsletter reporting on the worldwide Human Genome Project.

**Biological Sciences Curriculum Study (BSCS) Teaching Modules**

Online versions in preparation; hardcopies available from 719/531-5550

- "Genes, Environment, and Human Behavior," tentative title, in preparation
- "Mapping and Sequencing the Human Genome: Science, Ethics, and Public Policy" (1992)
- "The Human Genome Project: Biology, Computers, and Privacy" (1996)
- "The Puzzle of Inheritance: Genetics and the Methods of Science" (1997)

***Primer on Molecular Genetics, 1992**
*http://www.ornl.gov/hgmis/publicat/publications.html#primer*

Explains the science behind the genome research.

***To Know Ourselves, 1996**
*http://www.ornl.gov/hgmis/tko*

Booklet reviewing DOE's role, history, and achievements in the Human Genome Project and introducing the science and other aspects of the project.

## Ethical, Legal, and Social Issues Related to Genetics Research

**HGMIS Gateways Web page**
*http://www.ornl.gov/hgmis/links.html*

Choose "Ethical, Legal, and Social Issues."

**Center for Bioethics, University of Pennsylvania**
*http://www.med.upenn.edu/~bioethic*

Full-text articles about such ethical issues as human cloning; includes a primer on bioethics.

**Courts and Science On-Line Magazine (CASOLM)**
*http://www.ornl.gov/courts*

Coverage of genetic issues affecting the courts.

**ELSI in Science**
*http://www.lbl.gov/Education/ELSI/ELSI.html*

Teaching modules designed to stimulate discussion on implications of scientific research.

**Eubios Ethics Institute**
*http://www.biol.tsukuba.ac.jp/~macer/index.html*

Site includes newsletter summarizing literature in bioethics and biotechnology.

**Genetic Privacy Act**
*http://www.ornl.gov/hgmis/resource/elsi.html*

Model legislation written with support of the DOE Human Genome Program.

**MCET—The Human Genome Project**
*http://phoenix.mcet.edu/humangenome/index.html*

ELSI issues for high school students.

*Print copy available from HGMIS (see p. 87 or inside front cover for contact information).

**National Bioethics Advisory Committee**
*http://www.nih.gov/nbac/nbac.htm*

The bioethics committee offers advice to the National Science and Technology Council and others on bioethical issues arising from research related to human biology and behavior.

**National Center for Genomic Resources**
*http://www.ncgr.org*

Comprehensive Genetics and Public Issues page; includes congressional bills related to genetic privacy.

**The Gene Letter**
*http://www.geneletter.org/genetalk.html*

Bimonthly newsletter to inform consumers and professionals about advances in genetics and encourage discussion about emerging policy dilemmas.

**Your Genes, Your Choices**
*http://www.nextwave.org/ehr/books/index.html*

Booklet written in simple English, describing the Human Genome Project; the science behind it; and how ethical, legal, and social issues raised by the project may affect people's everyday lives.

## General Genetics and Biotechnology

Many of the following sites contain links to both educational and technical material.

**HGMIS Community Education and Outreach Gateways Web Page**
*http://www.ornl.gov/hgmis/links.html*

**Access Excellence**
*http://outcast.gene.com/ae/index.html*

Extensive genetic and biotechnology resources for teachers and nonscientists.

**BIO Online (Biotechnology Industry Organization)**
*http://www.bio.com*

Comprehensive directory of biotechnology sites on the Internet.

**Biospace**
*http://www.biospace.com*

Biotech industry site; profiles biotech companies by region.

**BioTech**
*http://biotech.chem.indiana.edu*

An interactive educational resource and biotech reference tool; includes a dictionary of 6000 life science terms.

**Biotechnology Information Center, USDA National Agricultural Library**
*http://www.nal.usda.gov/bic*

Comprehensive agricultural biotechnology resource; includes a bibliography on patenting biotechnology products and processes (*http://www.nal.usda.gov/bic/Biblios/patentag.htm*).

**Bugs 'N Stuff**
*http://www.ncgr.org/microbe*

List of microbial genomes being sequenced, research groups, genome sizes, and facts about selected organisms. Links to related sites.

**Careers in Genetics**
*http://www.faseb.org/genetics/gsa/careers/bro-menu.htm*

Online booklet from the Genetics Society of America, including several profiles of geneticists. See also career sections of sites specified above, such as Access Excellence.

**Carolina Biological Supply Company**
*http://www.carosci.com/Tips.htm*

Teaching materials for all levels. Includes mini-lessons on selected scientific topics, two online magazines, What's New, software, catalogs, and publications.

**Cell & Molecular Biology Online**
*http://www.tiac.net/users/pmgannon*

Links to electronic publications, current research, educational and career resources, and more.

**CERN Virtual Library, Genetics section, Biosciences Division**
*http://www.ornl.gov/TechResources/Human_Genome/genetics.html*

Includes an organism index linking to other pertinent databases, information on the U.S. and international Human Genome Projects, and links to research sites.

**Classic Papers in Genetics**
*http://www.esp.org*

Covers the early years, with introductory notes. See also Access Excellence site above for genetics history.

**Community of Science Web Server**
*http://cos.gdb.org/best.html*

Links to Medline, U.S. Patent Citation Database, Commerce Business Daily, The Federal Register, and other resources.

**Database of Genome Sizes**
*http://www.cbs.dtu.dk/databases/DOGS/index.html*

Lists numerous organisms with genome sizes, scientific and common names, classifications, and references.

**Genetic and biological resources links**
*http://www.er.doe.gov/production/ober/bioinfo_center.html*

**Genetics Education Center, University of Kansas Medical Center**
*http://www.kumc.edu/instruction/medicine/genetics/homepage.html*

Educational information on human genetics, career resources.

**Genetics Glossary**
*http://www.ornl.gov/hgmis/publicat/glossary.html*

Glossary of terms related to genetics.

**Genetics Webliography**
*http://www.dml.georgetown.edu/%7Edavidsol/len.html*

Extensive links for researchers and nonscientists from Georgetown University Library.

**Genomics: A Global Resource**
*http://www.phrma.org/genomics/index.html*

Many links. Website a joint project of the Pharmaceutical Research and Manufacturers of America and the American Institute of Biological Sciences; includes Genomics Today, a daily update on the latest news in the field.

**Hispanic Educational Genome Project**
*http://vflylab.calstatela.edu/hgp*

Designed to educate high school students and their families about genetics and the Human Genome Project. Links to other projects.

**Howard Hughes Medical Institute**
*http://www.hhmi.org*

Home page of major U.S. philanthropic organization that supports research in genetics, cell biology, immunology, structural biology, and neuroscience. Excellent introductory information on these topics.

**Library of Congress**
*http://lcweb.loc.gov/homepage/lchp.html*

**Microbial Database**
*http://www.tigr.org/tdb/mdb/mdb.html*

Lists completed and in-progress microbial genomes, with funding sources.

**MIT Biology Hypertextbook**
*http://esg-www.mit.edu:8001/esgbio/7001main.html*

All the basics.

**Science and Mathematics Resources**
*http://www-sci.lib.uci.edu*

More than 2000 Web references, including Frank Potter's Science Gems and Martindale's Health Science Guide. For teachers at all levels.

**Virtual Courses on the Web**
*http://lenti.med.umn.edu/~mwd/courses.html*

Links to Web tutorials in biology, genetics, and more.

**Welch Web**
*http://www.welch.jhu.edu*

Links to many Internet biomedical resources, dictionaries, encyclopedias, government sites, libraries, and more, from the Johns Hopkins University Welch Library.

**Why Files**
*http://whyfiles.news.wisc.edu*

Illustrated explanations of the science behind the news.

## Images on the Web

**Biochemistry Online**
*http://biochem.arach-net.com*

Essays, courses, 3-D images of biomolecules, modeling, software.

**Bugs in the News!**
*http://falcon.cc.ukans.edu/~jbrown/bugs.html*

Microbiology information and a nice collection of images of biological molecules.

**Cells Alive!**
*http://www.cellsalive.com*

Images (some moving) of different types of cells.

**Cn3D (See in 3-D)**
*http://www3.ncbi.nlm.nih.gov/Entrez/Structure/cn3d.html*

3-D molecular structure viewer allowing the user to visualize and rotate structure data entries from Entrez. Highly technical, for researchers.

**Cytogenetics Gallery**
*http://www.pathology.washington.edu:80/Cytogallery*

Photos (karyotypes) of normal and abnormal chromosomes.

**DNA Learning Center, Cold Spring Harbor Laboratory**
*http://darwin.cshl.org/index.html*

Animated images of PCR and Southern Blotting techniques.

**Gene Map from the 1996 Genome Issue of Science**
*http://www.ncbi.nlm.nih.gov/SCIENCE96*

Click on particular areas of chromosomes and find genes.

**Images of Biological Molecules**
*http://www.cc.ukans.edu/~micro/picts.html*

3-D structures of proteins and nucleic acids obtained from Brookhaven National Laboratory Protein Database and others.

**Lawrence Livermore National Laboratory Chromosome 19 Physical Map**
*http://www-bio.llnl.gov/bbrp/genome/genome.html*

**Los Alamos National Laboratory Chromosome 16 Physical Map**
*http://www-ls.lanl.gov/DBqueries/QueryPage.html*

## Journals and Magazines

**HGMIS Journals Gateways Web page**
*http://www.ornl.gov/hgmis/links.html*

Choose "Journals, Books, Periodicals."

**Biochemistry and Molecular Biology Journals**
*http://biochem.arach-net.com/beasley/journals.html*

Comprehensive list.

**Nature, Nature Genetics, and Nature Biotechnology**
*http://www.nature.com*

Abstracts of articles, full text of letters and editorials.

**Science Magazine**
*http://www.sciencemag.org*

Abstracts and some full-text articles.

**Science Magazine Genome Issue (10/96)**
*http://www.sciencemag.org/science/content/vol274/issue5287*

Full text includes a "clickable" gene map.

**Science News**
*http://www.sciencenews.org*

Online version of weekly popular science magazine with full text of selected articles.

## Medical Genetics

**Blazing a Genetic Trail**
*http://www.hhmi.org/GeneticTrail*

Illustrated booklet from the Howard Hughes Medical Institute on hunting for disease genes.

**Directory of National Genetic Voluntary Organizations and Related Resources**
*http://medhlp.netusa.net/agsg/agsgsup.htm*

Support groups for people with genetic diseases and their families.

**GeneCards**
*http://bioinformatics.weizmann.ac.il/cards*

A database of more than 6000 genes; describes their functions, products, and biomedical applications.

**Gene Therapy**
*http://www.mc.vanderbilt.edu/gcrc/gene/index.html*

Web course covering the basics, with links to other sites.

**Inherited-Disease Genes Found by Positional Cloning**
*http://www.ncbi.nlm.nih.gov/Baxevani/CLONE/index.html*

Links to OMIM.

**NIH Office of Recombinant DNA Activities**
*http://www.nih.gov/od/orda*

Includes a database of human gene therapy protocols.

**Online Mendelian Inheritance in Man (OMIM)**
*http://www.ncbi.nlm.nih.gov/Omim*

A comprehensive, authoritative, and up-to-date human gene and genetic disorder catalog that supports medical genetics and the Human Genome Project.

**Promoting Safe and Effective Genetic Testing in the United States (1997)**
*http://www.med.jhu.edu/tfgtelsi*

Principles and recommendations by a joint NIH-DOE Human Genome Project group that examined the development and provision of gene tests in the United States.

**Understanding Gene Testing**
*http://www.gene.com/ae/AE/AEPC/NIH/index.html*

Illustrated brochure from the National Cancer Institute.

## Science in the News

**EurekAlert!** *http://www.eurekalert.org*

**InScight:** *http://www.apnet.com/inscight*

**SciWeb:** *http://www.sciweb.com/news.html*

Short summaries of major stories, some with links to related articles in other sources.

**HMS Beagle**
*http://biomednet.com/hmsbeagle*

Biweekly electronic journal featuring major science stories, profiles, book reviews, and other items of interest.

**Science Daily**
*http://www.sciencedaily.com*

Headline stories, articles, and links to news services, newspapers, magazines, broadcast sources, journals, and organizations. Also offers weekly bulletins for updates by e-mail.

**Science Guide**
*http://www.scienceguide.com*

Daily news and information service and free science news e-mailer. Also contains directories of newsgroups, grant and funding resources, employment, and online journals.

**ScienceNow**
*http://www.sciencenow.org*

Daily online news service from Science magazine offers articles on major science news.

## Web Search Tools

**Biosciences Index to WWW Virtual Library**
*http://golgi.harvard.edu/htbin/biopages*

**Metacrawler**
*http://www.metacrawler.com*

**"Search the Net"**
*http://metro.turnpike.net/adorn/search.html*

Comprehensive list of search tools, libraries, world fact books, and other useful information.

**Search.com**
*http://www.search.com*

**Yahoo!**
*http://www.yahoo.com*

Prepared August 1997 by
Human Genome Management Information System
Oak Ridge National Laboratory
1060 Commerce Park, MS 6480
Oak Ridge, TN 37830
423/576-6669, *caseydk@ornl.gov*
*http://www.ornl.gov/hgmis*

# Appendix E
# 1996 Human Genome Research Projects

**Research abstracts of these projects appear in Part 2 of this report.**

## *Sequencing*

### Advanced Detectors for Mass Spectrometry
**W.H. Benner** and **J.M. Jaklevic**
Lawrence Berkeley National Laboratory, Berkeley, California

### Mass Spectrometer for Human Genome Sequencing
**Chung-Hsuan Chen**
Oak Ridge National Laboratory, Oak Ridge, Tennessee

### Genomic Sequence Comparisons
**George Church**
Harvard Medical School, Boston, Massachusetts

### A PAC/BAC End-Sequence Data Resource for Sequencing the Human Genome: A 2-Year Pilot Study
**Pieter de Jong**
Roswell Park Cancer Institute, Buffalo, New York

### Multiple-Column Capillary Gel Electrophoresis
**Norman Dovichi**
University of Alberta, Edmonton, Canada

### DNA Sequencing with Primer Libraries
**John J. Dunn** and **F. William Studier**
Brookhaven National Laboratory, Upton, New York

### Rapid Preparation of DNA for Automated Sequencing
**John J. Dunn** and **F. William Studier**
Brookhaven National Laboratory, Upton, New York

### A PAC/BAC End-Sequence Database for Human Genomic Sequencing
**Glen A. Evans**
University of Texas Southwestern Medical Center, Dallas, Texas

### Automated DNA Sequencing by Parallel Primer Walking
**Glen A. Evans**
University of Texas Southwestern Medical Center, Dallas, Texas

### *Parallel Triplex Formation as Possible Approach for Suppression of DNA-Viruses Reproduction
**V.L. Florentiev**
Russian Academy of Sciences, Moscow, Russia

### Advanced Automated Sequencing Technology: Fluorescent Detection for Multiplex DNA Sequencing
**Raymond F. Gesteland**
University of Utah, Salt Lake City, Utah

### Resource for Molecular Cytogenetics
**Joe Gray** and **Daniel Pinkel**
University of California, San Francisco

### DNA Sample Manipulation and Automation
**Trevor Hawkins**
Whitehead Institute and Massachusetts Institute of Technology, Cambridge, Massachusetts

### Construction of a Genome-Wide Characterized Clone Resource for Genome Sequencing
**Leroy Hood, Mark D. Adams,**[1] and **Melvin Simon**[2]
University of Washington, Seattle
[1]The Institute for Genomic Research, Rockville, Maryland
[2]California Institute of Technology, Pasadena, California

### DNA Sequencing Using Capillary Electrophoresis
**Barry L. Karger**
Northeastern University, Boston, Massachusetts

### Ultrasensitive Fluorescence Detection of DNA
**Richard A. Mathies** and **Alexander N. Glazer**
University of California, Berkeley

### Joint Human Genome Program Between Argonne National Laboratory and the Engelhardt Institute of Molecular Biology
**Andrei Mirzabekov**
Argonne National Laboratory, Argonne, Illinois, and
Engelhardt Institute of Molecular Biology, Moscow, Russia

### High-Throughput DNA Sequencing: SAmple SEquencing (SASE) Analysis as a Framework for Identifying Genes and Complete Large-Scale Genomic Sequencing
**Robert K. Moyzis**
Los Alamos National Laboratory, Los Alamos, New Mexico

### One-Step PCR Sequencing
**Barbara Ramsay Shaw**
Duke University, Durham, North Carolina

*Projects designated by an asterisk were funded through small emergency grants to Russian scientists following December 1992 site reviews by David Galas (formerly of OHER, renamed OBER in 1997), Raymond Gesteland (University of Utah), and Elbert Branscomb (LLNL).

### Automation of the Front End of DNA Sequencing

**Lloyd M. Smith** and **Richard A. Guilfoyle**
University of Wisconsin, Madison

### High-Speed DNA Sequence Analysis by Matrix-Assisted Laser Desorption Mass Spectrometry

**Lloyd M. Smith**
University of Wisconsin, Madison

### Analysis of Oligonucleotide Mixtures by Electrospray Ionization-Mass Spectrometry

**Richard D. Smith**
Pacific Northwest National Laboratory, Richland, Washington

### High-Speed Sequencing of Single DNA Molecules in the Gas Phase by FTICR-MS

**Richard D. Smith**
Pacific Northwest National Laboratory, Richland, Washington

### Characterization and Modification of DNA Polymerases for Use in DNA Sequencing

**Stanley Tabor**
Harvard University, Boston, Massachusetts

### Modular Primers for DNA Sequencing

**Levy Ulanovsky**[1,2]
[1]Argonne National Laboratory, Argonne, Illinois
[2]Weizmann Institute of Science, Rehovot, Israel

### Time-of-Flight Mass Spectroscopy of DNA for Rapid Sequence

**Peter Williams**
Arizona State University, Tempe, Arizona

### Development of Instrumentation for DNA Sequencing at a Rate of 40 Million Bases Per Day

**Edward S. Yeung**
Iowa State University, Ames, Iowa

## *Mapping*

### Resolving Proteins Bound to Individual DNA Molecules

**David Allison** and **Bruce Warmack**
Oak Ridge National Laboratory, Oak Ridge, Tennessee

### *Improved Cell Electrotransformation by Macromolecules

**Alexandre S. Boitsov**
St. Petersburg State Technical University, St. Petersburg, Russia

### Overcoming Genome Mapping Bottlenecks

**Charles R. Cantor**
Boston University, Boston, Massachusetts

### Preparation of PAC Libraries

**Pieter J. de Jong**
Roswell Park Cancer Institute, Buffalo, New York

### Chromosomes by Third-Strand Binding

**Jacques R. Fresco**
Princeton University, Princeton, New Jersey

### Chromosome Region-Specific Libraries for Human Genome Analysis

**Fa-Ten Kao**
Eleanor Roosevelt Institute for Cancer Research, Denver, Colorado

### *Identification and Mapping of DNA-Binding Proteins Along Genomic DNA by DNA-Protein Crosslinking

**V.L. Karpov**
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

### A PAC/BAC Data Resource for Sequencing Complex Regions of the Human Genome: A 2-Year Pilot Study

**Julie R. Korenberg**
Cedars Sinai Medical Center, Los Angeles, California

### Mapping and Sequencing of the Human X Chromosome

**D. L. Nelson**
Baylor College of Medicine, Houston, Texas

### *Sequence-Specific Proteins Binding to the Repetitive Sequences of High Eukaryotic Genome

**Olga Podgornaya**
Institute of Cytology, Russian Academy of Sciences, St. Petersburg, Russia

### *Protein-Binding DNA Sequences

**O.L. Polanovsky**
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

### *Development of Intracellular Flow Karyotype Analysis

**A.I. Poletaev**
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

### Mapping and Sequencing with BACs and Fosmids

**Melvin I. Simon**
California Institute of Technology, Pasadena, California

### Towards a Globally Integrated, Sequence-Ready BAC Map of the Human Genome

**Melvin I. Simon**
California Institute of Technology, Pasadena, California

### Generation of Normalized and Subtracted cDNA Libraries to Facilitate Gene Discovery

**Marcelo Bento Soares**
Columbia University, New York, New York

### Mapping in Man-Mouse Homology Regions

**Lisa Stubbs**
Oak Ridge National Laboratory, Oak Ridge, Tennessee

### Positional Cloning of Murine Genes

**Lisa Stubbs**
Oak Ridge National Laboratory, Oak Ridge, Tennessee

### Human Artificial Episomal Chromosomes (HAECS) for Building Large Genomic Libraries

**Jean-Michel H. Vos**
University of North Carolina, Chapel Hill

### *Cosmid and cDNA Map of a Human Chromosome 13q14 Region Frequently Lost at B Cell Chronic Lymphocytic Leukemia

**N.K. Yankovsky**
N.I. Vavilov Institute of General Genetics, Moscow, Russia

## Informatics

### BCM Server Core

**Daniel Davison**
Baylor College of Medicine, Houston, Texas

### A Freely Sharable Database-Management System Designed for Use in Component-Based, Modular Genome Informatics Systems

**Nathan Goodman**
The Jackson Laboratory, Bar Harbor, Maine

### A Software Environment for Large-Scale Sequencing

**Mark Graves**
Baylor College of Medicine, Houston, Texas

### Generalized Hidden Markov Models for Genomic Sequence Analysis

**David Haussler**
University of California, Santa Cruz

### Identification, Organization, and Analysis of Mammalian Repetitive DNA Information

**Jerzy Jurka**
Genetic Information Research Institute, Palo Alto, California

### *TRRD, GERD and COMPEL: Databases on Gene-Expression Regulation as a Tool for Analysis of Functional Genomic Sequences

**N.A. Kolchanov**
Institute of Cytology and Genetics, Novosibirsk, Russia

### Data-Management Tools for Genomic Databases

**Victor M. Markowitz** and **I-Min A. Chen**
Lawrence Berkeley National Laboratory, Berkeley, California

### The Genome Topographer: System Design

**T. Marr**
Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

### A Flexible Sequence Reconstructor for Large-Scale DNA Sequencing: A Customizable Software System for Fragment Assembly

**Gene Myers**
University of Arizona, Tucson

### The Role of Integrated Software and Databases in Genome Sequence Interpretation and Metabolic Reconstruction

**Ross Overbeek**
Argonne National Laboratory, Argonne, Illinois

### Database Transformations for Biological Applications

**G. Christian Overton**, **Susan B. Davidson**, and **Peter Buneman**
University of Pennsylvania, Philadelphia

### Las Vegas Algorithm for Gene Recognition: Suboptimal and Error-Tolerant Spliced Alignment

**Pavel A. Pevzner**
University of Southern California, Los Angeles, California

### Foundations for a Syntactic Pattern-Recognition System for Genomic DNA Sequences: Languages, Automata, Interfaces, and Macromolecules

**David B. Searls**
SmithKline Beecham Pharmaceuticals, King of Prussia, Pennsylvania

### Analysis and Annotation of Nucleic Acid Sequence

**David J. States**
Washington University, St. Louis, Missouri

### Gene Recognition, Modeling, and Homology Search in GRAIL and genQuest

**Edward C. Uberbacher**
Oak Ridge National Laboratory, Oak Ridge, Tennessee

### Informatics Support for Mapping in Mouse-Human Homology Regions

**Edward Uberbacher**
Oak Ridge National Laboratory, Oak Ridge, Tennessee

### SubmitData: Data Submission to Public Genomic Databases

**Manfred D. Zorn**
Lawrence Berkeley National Laboratory, University of California, Berkeley

## ELSI

### The Human Genome: Science and the Social Consequences; Interactive Exhibits and Programs on Genetics and the Human Genome

**Charles C. Carlson**
The Exploratorium, San Francisco, California

### Documentary Series for Public Broadcasting

**Graham Chedd** and **Noel Schwerin**
Chedd-Angier Production Company, Watertown, Massachusetts

### Human Genome Teacher Networking Project

**Debra L. Collins** and **R. Neil Schimke**
University of Kansas Medical Center, Kansas City, Kansas

### Human Genome Education Program

**Lane Conn**
Stanford Human Genome Center, Palo Alto, California

### *Your World/Our World–Biotechnology & You*: Special Issue on the Human Genome Project

**Jeff Davidson** and **Laurence Weinberger**
Pennsylvania Biotechnology Association, State College, Pennsylvania

### The Human Genome Project and Mental Retardation: An Educational Program

**Sharon Davis**
The Arc of the United States, Arlington, Texas

### Pathways to Genetic Screening: Molecular Genetics Meets the High-Risk Family

**Troy Duster**
University of California, Berkeley

### Intellectual Property Issues in Genomics

**Rebecca S. Eisenberg**
University of Michigan Law School, Ann Arbor, Michigan

### AAAS Congressional Fellowship Program

**Stephen Goodman**
The American Society of Human Genetics, Bethesda, Maryland

### A Hispanic Educational Program for Scientific, Ethical, Legal, and Social Aspects of the Human Genome Project

**Margaret C. Jefferson** and **Mary Ann Sesma**[1]
California State University and [1]Los Angeles Unified School District, Los Angeles, California

### Implications of the Geneticization of Health Care for Primary Care Practitioners

**Mary B. Mahowald**
University of Chicago, Chicago, Illinois

### Nontraditional Inheritance: Genetics and the Nature of Science; Instructional Materials for High School Biology

**Joseph D. McInerney** and **B. Ellen Friedman**
Biological Sciences Curriculum Study, Colorado Springs, Colorado

### The Human Genome Project: Biology, Computers, and Privacy: Development of Educational Materials for High School Biology

**Joseph D. McInerney** and **Lynda B. Micikas**
Biological Sciences Curriculum Study, Colorado Springs, Colorado

### Involvement of High School Students in Sequencing the Human Genome

**Maureen M. Munn**, **Maynard V. Olson**, and **Leroy Hood**
University of Washington, Seattle

### *The Gene Letter*: A Newsletter on Ethical, Legal, and Social Issues in Genetics for Interested Professionals and Consumers

**Philip J. Reilly**, **Dorothy C. Wertz**, and **Robin J.R. Blatt**
The Shriver Center for Mental Retardation, Waltham, Massachusetts

### *The DNA Files*: A Nationally Syndicated Series of Radio Programs on the Social Implications of Human Genome Research and Its Applications

**Bari Scott**
Genome Radio Project, KPFA-FM, Berkeley, California

### Communicating Science in Plain Language: The Science+ Literacy for Health: Human Genome Project

**Maria Sosa**, Judy Kass, and Tracy Gath
American Association for the Advancement of Science, Washington, D.C.

### The Community College Initiative

**Sylvia J. Spengler** and Laurel Egenberger
Lawrence Berkeley National Laboratory, Berkeley, California

### Genome Educators

**Sylvia Spengler** and **Janice Mann**
Lawrence Berkeley National Laboratory, Berkeley, California

### Getting the Word Out on the Human Genome Project: A Course for Physicians

**Sara L. Tobin** and **Ann Boughton**[1]
Stanford University, Palo Alto, California
[1]Thumbnail Graphics, Oklahoma City, Oklahoma

### The Genetics Adjudication Resource Project

**Franklin M. Zweig**
Einstein Institute for Science, Health, and the Courts, Bethesda, Maryland

## Infrastructure

### Alexander Hollaender Distinguished Postdoctoral Fellowships

**Linda Holmes** and **Eugene Spejewski**
Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee

### Human Genome Management Information System

**Betty K. Mansfield** and **John S. Wassom**
Oak Ridge National Laboratory, Oak Ridge, Tennessee

### Human Genome Program Coordination

**Sylvia J. Spengler**
Lawrence Berkeley National Laboratory, Berkeley, California

### Support of Human Genome Program Proposal Reviews

**Walter Williams**
Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee

### Former Soviet Union Office of Health and Environmental Research Program

**James Wright**
Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee

## SBIR

## 1996 Phase I

### An Engineered RNA/DNA Polymerase to Increase Speed and Economy of DNA Sequencing

**Mark W. Knuth**
Promega Corporation, Madison, Wisconsin

### Directed Multiple DNA Sequencing and Expression Analysis by Hybridization

**Gualberto Ruano**
BIOS Laboratories, Inc., New Haven, Connecticut

## 1996 Phase II

### A Graphical Ad Hoc Query Interface Capable of Accessing Heterogeneous Public Genome Databases

**Joseph Leone**
CyberConnect Corporation, Storrs, Connecticut

### Low-Cost Automated Preparation of Plasmid, Cosmid, and Yeast DNA

**William P. MacConnell**
MacConnell Research Corporation, San Diego, California

### GRAIL-GenQuest: A Comprehensive Computational Framework for DNA Sequence Analysis

**Ruth Ann Manning**
ApoCom, Inc., Oak Ridge, Tennessee

# *Appendix F: DOE BER Program*

***Text and photos in this appendix first appeared in a brochure prepared by the Human Genome Management Information System for the DOE Office of Biological and Environmental Research to announce a symposium celebrating 50 years of achievements in the Biological and Environmental Research Program. "Serving Science and Society into the New Millennium" was held on May 21–22, 1997, at the National Academy of Sciences in Washington, D.C. The color brochure and other recent publications related to BER research, including the historically comprehensive* A Vital Legacy*, may be obtained from HGMIS at the address on the inside front cover.***

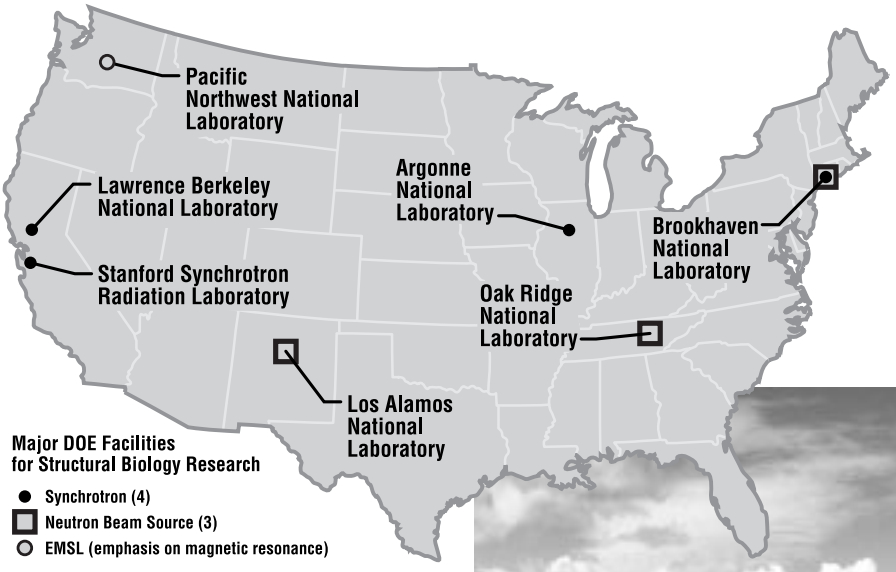
# DOE Biological and Environmental Research Program

## *An Extraordinary Legacy*

To exploit the boundless promise of energy technologies and shed light on their consequences to public health and the environment, the Biological and Environmental Research program of the U.S. Department of Energy's (DOE) Office of Health and Environmental Research (OHER) has engaged in a variety of multidisciplinary research activities:

- Establishing the world's first Human Genome Program.
- Developing advanced medical diagnostic tools and treatments for human disease.
- Assessing the health effects of radiation.

**Biological and Environmental Research Program**
**Aristides Patrinos, Ph.D.**
**Associate Director for Energy Research**
**for the**
**Office of Biological and Environmental Research**
**U.S. Department of Energy**
**301/903-3251, Fax: 301/903-5051**
***http://www.er.doe.gov/production/ober/ober_top.html***




### National User Facilities

**Dedicated biomedical resources, such as those maintained by BER at several DOE laboratories, are available at little or no charge. These resources enable scientists to gain an understanding of relationships between biological structures and their functions, study disease processes, develop new pharmaceuticals, and conduct basic research in molecular biology and environmental processes.**

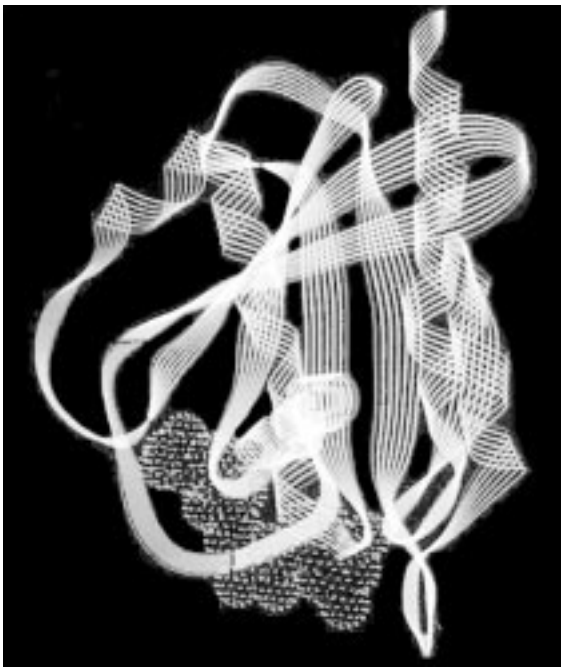
**William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) is a national collaborative user facility for providing innovative approaches to meet the needs of DOE's environmental missions.**

# *An Enduring Mandate*

DOE is carrying forward Congressional mandates that began with its predecessors, the Atomic Energy Commission and the Energy Research and Development Agency:

## Contribute to a Healthy Citizenry

- Develop innovative technologies for tomorrow's biomedical sciences.
- Provide the basis for individual risk assessments by determining the human genome's fine structure by the year 2005.
- Conduct research into advanced medical technologies and radiopharmaceuticals.
- Build and support national user facilities for determining biological structure, and ultimately function, at the molecular and cellular level.



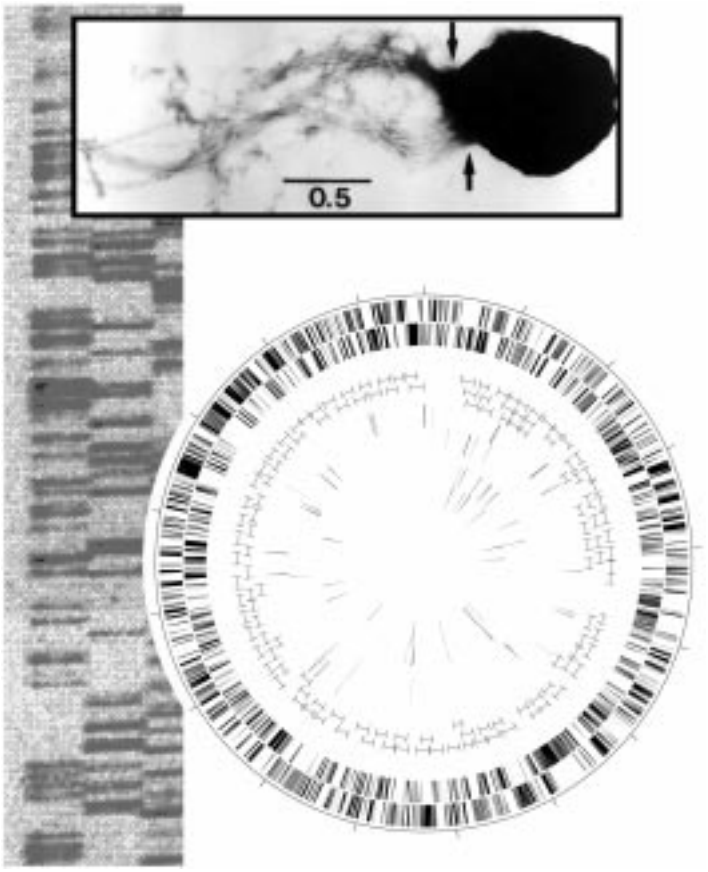
**DOE user facilities are revealing the molecular details of life. Knowing the 3-D structure of the ras protein (above), an important molecular switch governing human cell growth, will enable interventions to shut off this switch in cancer cells.**

## Understand Global Climate Change

Predict the effects of energy production and its use on the regional and global environment by acquiring data and developing the necessary understanding of environmental processes.

## Contribute to Environmental Cleanup

Conduct fundamental research to establish a better scientific basis for remediating contaminated sites.



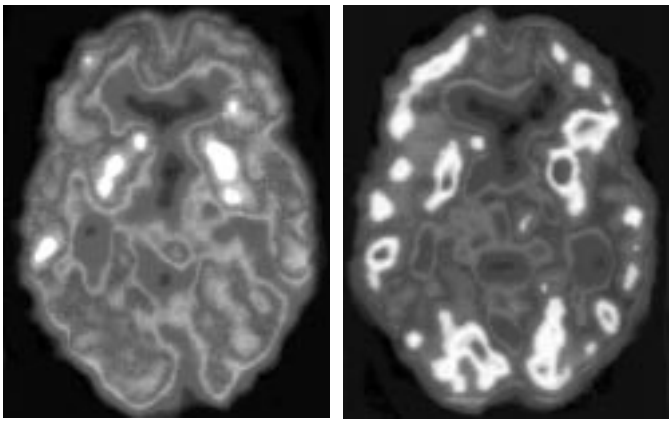

**Determining the fine structure—DNA sequence—of the microorganism *Methanococcus jannaschii* (pictured at right, top) and other minimal life forms in DOE's Microbial Genome Program will benefit medicine, agriculture, industrial and energy production, and environmental bioremediation. The circular representation of the single *M. jannaschii* chromosome, which was fully sequenced in 1996, illustrates the location of genes and other important features. (Vertical bar represents a portion of a sequencing experiment.)**

# *Fifty Years of Achievements. . .*
# *Leading to Innovative Solutions*

## Tools for Medicine and Research

Radioisotopes developed for medicine and medical imaging are being merged with current knowledge in biology and genetics to discover new ways of diagnosing and treating cancer and other disorders, detecting genes in action, and understanding normal development and function of human organ systems.

- Radioactive molecules used in medical imaging for positron emission tomography (PET) and magnetic resonance imaging (MRI) allow noninvasive diagnosis, monitoring, and exploration of human disorders and their treatments.
- Isotopes and other tracers of brain activity are being used to explore drug addiction, the effects of smoking, Alzheimer's disease, Parkinson's disease, and schizophrenia.
- Technetium-99m is used to diagnose diseases of the kidney, liver, heart, brain, and other organs in about 13 million patients per year.
- Striking successes have been achieved using charged atomic particles to treat thyroid diseases, pituitary tumors, and eye cancer, among other disorders.



**One-quarter of all patients in U.S. hospitals undergo tests using descendants of cameras developed by BER to follow radioactive tracers in the body. PET scanning has been key to a generation of brain metabolism studies as well as diagnostic tests for heart disease and cancer. PET studies above reveal brain metabolism differences in recovering alcoholics (left, 10 days, and right, 30 days, after withdrawal from alcohol).**

**The laser-based flow cytometer developed at DOE national laboratories enables researchers to separate human chromosomes for analysis.**

## Genome Projects

A legacy of DOE research on genetic effects paved the way for the world's first Human Genome Program. Now new genomic technologies are being applied to environmental cleanup through the DOE Natural and Accelerated Bioremediation Research and Microbial Genome programs, healthcare and risk assessment, and such other national priorities as industrial processes and agriculture.

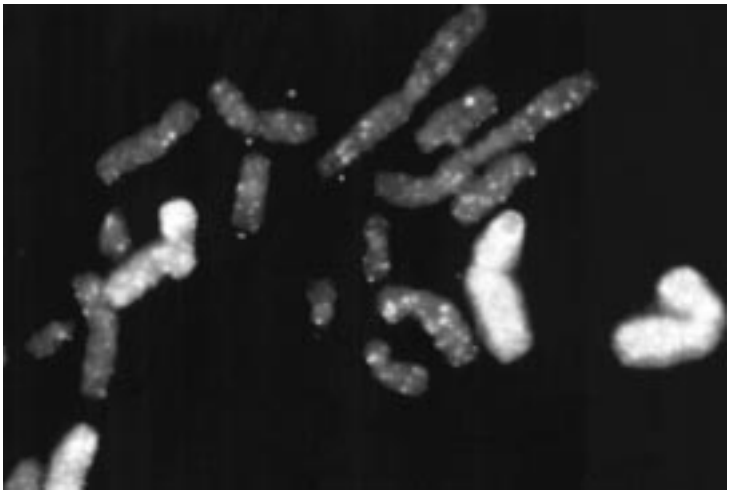
**Discover the breadth of current activities and recent accomplishments via the BER Web Site:**

***http://www.er.doe.gov/production/ober/ober_top.html***

## Radiation Risks and Protection Guidelines

BER studies have become the foundation for laws and standards that protect the population, including workers exposed to radiological sources:

- Guidelines for the safe use of diagnostic X rays and radiopharmaceuticals.
- Safety standards for the presence of radionuclides in food and drinking water.
- Radiation-detection systems and dosimetry techniques.



**Human chromosomes "painted" by fluorescent dyes to detect abnormal exchange of genetic material frequently present in cancer. Chromosome paints also serve as valuable resources for other clinical and research applications.**

## Finding a Link Between DNA Damage and Cancers

Studies of DNA damage have uncovered similar mechanisms at work in damage caused by radiation exposure, X rays, ultraviolet light, and cancer-causing chemicals. A screening test for such chemicals is now one of the first hurdles a new compound must clear on its way to regulatory and public acceptance.

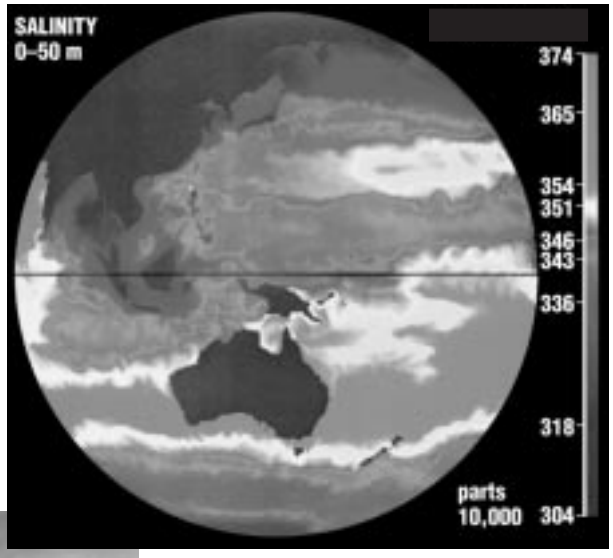
***"... (it's) not so much where we stand as in what direction we are moving. [Oliver Wendell Homes, Sr.]"***

## Tracking the Regional and Global Movement of Pollutants

BER research helped to establish the earliest and most authoritative monitoring network in the world to detect airborne radioisotopes. The use of atmospheric tracers has led to the improved ability to predict the dispersion of pollutants.

## Understanding Global Change

Important achievements in environmental research have led to enhanced capabilities in studying global change, including more accurate predictions of global and regional climate changes induced by increasing atmospheric concentrations of greenhouse gases.




**High-performance computing is promoting faster and more realistic solutions to long-term climate change.**

**The Unmanned Aerospace Vehicle (above) conducts measurements to quantify the fate of solar radiation falling on the earth.**

## Creating a New Science of Ecology

BER achievements in using radioactive tracers to follow the movements of animals, routes of chemicals through food chains, decomposition of forest detritus, together with the program's introduction of computer simulations, created the new field of radioecology.

**This glossary was adapted from definitions in the DOE *Primer on Molecular Genetics* (1992).**

***http://www.ornl.gov/hgmis/publicat/primer/intro.html***

# A

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

**Allele:** Alternative form 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).

# B

**BAC:** See bacterial artificial chromosome.

**Bacterial artificial chromosome (BAC):** A vector used to clone DNA fragments (100- to 300-kb insert size; average, 150 kb) in *Escherichia coli* cells. Based on naturally occurring F-factor plasmid found in the bacterium *E. coli*. Compare cloning vector.

**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.

# C

**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.

**Chromosome:** The self-replicating genetic structure 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.

**Clone:** 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 vectors 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 sequence:** Nucleic acid base sequence that can form a double-stranded structure by matching base pairs with another sequence; 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:** Group of clones representing overlapping regions of a genome.

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

**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).

# D

**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 probe:** 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

***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 chromosome.

**Evolutionarily conserved:** See conserved sequence.

**Exogenous DNA:** DNA originating outside an organism.

**Exon:** The protein-coding DNA sequence of a gene. Compare intron.

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

**Expressed gene:** See gene expression.

# F

**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 separate chromosomes on the basis of their DNA content.

# G

**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 family:** Group 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 technology:** See recombinant DNA technology.

**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 project**: Research and technology development effort 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).

# H

**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.

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

**Homologous chromosome:** Chromosome containing the same linear gene sequences as another, 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.

# I

**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.

**Intron:** The DNA base sequence interrupting the protein-coding sequence of a gene; this sequence is transcribed into RNA but is cut out of the message before it is translated into protein. Compare exon.

**In vitro:** Outside a living organism.

# K

**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.

# L

**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.

# M

**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 disorder:** See polygenic disorder.

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

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

# N

**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.

# O

**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.

# P

**P1-derived artificial chromosome (PAC):** A vector used to clone DNA fragments (100- to 300-kb insert size; average, 150 kb) in *Escherichia coli* cells. Based on bacteriophage (a virus) P1 genome. Compare cloning vector.

**PAC:** See P1-derived artificial chromosome.

**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 disorder:** Genetic disorder 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 chromosome.

**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 bodys 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.

# R

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

**Recombinant clone:** Clone containing recombinant DNA molecules. See recombinant DNA technology.

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

**Recombinant DNA technology:** Procedure 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 region or sequence:** 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.

**Ribonucleotide:** 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.

# S

**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 chromosome:** The X or Y chromosome in human beings that determines 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 cell:** 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.

# T

**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 end of a chromosome. This specialized structure is 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.

## U

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

## V

**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 more than 100,000 transistors on a chip.

## Y

**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.