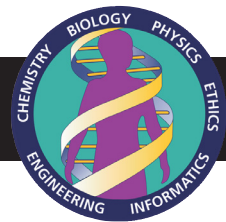# The Human Genome Project, 1990–2003

A Brief Overview

Though surprising to many, the Human Genome Project (HGP) traces its roots to an initiative in the U.S. Department of Energy (DOE). Since 1947, DOE and its predecessor agencies have been charged by Congress with developing new energy resources and technologies and pursuing a deeper understanding of potential health and environmental risks posed by their production and use. Such studies, for example, have provided the scientific basis for individual risk assessments of nuclear medicine technologies.

In 1986, DOE took a bold step in announcing the Human Genome Initiative, convinced that its missions would be well served by a reference human genome sequence. Shortly thereafter, DOE joined with the National Institutes of Health (NIH) to develop a plan for a joint HGP that officially began in 1990. During the early years of the HGP, the Wellcome Trust, a private charitable institution in the United Kingdom, joined the effort as a major partner. Important contributions also came from other collaborators around the world, including Japan, France, Germany, and China.

Ambitious Goals

The HGP's ultimate goal was to generate a high-quality reference DNA sequence for the human genome's 3 billion base pairs and to identify all human genes. Other important goals included sequencing the genomes of model organisms to interpret human DNA, enhancing computational resources to support future research and commercial applications, exploring gene function through mouse-human comparisons, studying human variation, and training future scientists in genomics.

The powerful analytic technology and data arising from the HGP raise complex ethical and policy issues for individuals and society. These challenges include privacy, fairness in use and access of genomic information, reproductive and clinical issues, and commercialization (see p. 8). Programs that identify and address these implications have been an integral part of the HGP and have become a model for bioethics programs worldwide.

A Lasting Legacy

In June 2000, to much excitement and fanfare, scientists announced the completion of the first working draft of the entire human genome. First analyses of the details appeared in the February 2001 issues of the journals *Nature* and *Science*. The high-quality reference sequence was completed in April 2003, marking the end of the Human Genome Project—2 years ahead of the original schedule. Coincidentally, this was also the 50th anniversary of Watson and Crick's publication of DNA structure that launched the era of molecular biology.

Available to researchers worldwide, the human genome reference sequence provides a magnificent and unprecedented biological resource that will serve throughout the century as a basis for research and discovery and, ultimately, myriad practical applications. The sequence already is having an impact on finding genes associated with human disease (see p. 3). Hundreds of other genome sequence projects—on microbes, plants, and animals—have been completed since the inception of the HGP, and these data now enable detailed comparisons among organisms, including humans.

Many more sequencing projects are under way or planned because of the research value of DNA sequence, the tremendous sequencing capacity now available, and continued improvements in technologies. Sequencing projects on the genomes of many microbes, as well as the honeybee, cow, and chicken are in progress.

Beyond sequencing, growing areas of research focus on identifying important elements in the DNA sequence responsible for regulating cellular functions and providing the basis of human variation. Perhaps the most daunting challenge is to begin to understand how all the "parts" of cells—genes, proteins, and many other molecules—work together to create complex living organisms. Future analyses on this treasury of data will provide a deeper and more comprehensive understanding of the molecular processes underlying life and will have an enduring and profound impact on how we view our own place in it.

Early Insights from the Human DNA Sequence

What We've Learned Thus Far

The first panoramic views of the human genetic landscape have revealed a wealth of information and some early surprises. Much remains to be deciphered in this vast trove of information; as the consortium of HGP scientists concluded in their seminal paper, ". . .the more we learn about the human genome, the more there is to explore." A few highlights from the first publications analyzing the sequence follow.

- The human genome contains 3 billion chemical nucleotide bases (A, C, T, and G).

- The average gene consists of 3000 bases, but sizes vary greatly, with the largest known human gene being dystrophin at 2.4 million bases.

- The functions are unknown for more than 50% of discovered genes.

- The human genome sequence is almost (99.9%) exactly the same in all people.

- About 2% of the genome encodes instructions for the synthesis of proteins.

- Repeat sequences that do not code for proteins make up at least 50% of the human genome.

- Repeat sequences are thought to have no direct functions, but they shed light on chromosome structure and dynamics. Over time, these repeats reshape the genome by rearranging it, thereby creating entirely new genes or modifying and reshuffling existing genes.

- The human genome has a much greater portion (50%) of repeat sequences than the mustard weed (11%), the worm (7%), and the fly (3%).

- Over 40% of the predicted human proteins share similarity with fruit-fly or worm proteins.

- Genes appear to be concentrated in random areas along the genome, with vast expanses of noncoding DNA between.

- Chromosome 1 (the largest human chromosome) has the most genes (2968), and the Y chromosome has the fewest (231).

- Genes have been pinpointed and particular sequences in those genes associated with numerous diseases and disorders including breast cancer, muscle disease, deafness, and blindness.

- Scientists have identified about 3 million locations where single-base DNA differences (see p. 9) occur in humans. This information promises to revolutionize the processes of finding DNA sequences associated with such common diseases as cardiovascular disease, diabetes, arthritis, and cancers.

How Does the Human Genome Stack Up?

Organism	Genome Size (Bases)	Estimated Genes
Human (*Homo sapiens*)	3 billion	30,000
Laboratory mouse (*M. musculus*)	2.6 billion	30,000
Mustard weed (*A. thaliana*)	100 million	25,000
Roundworm (*C. elegans*)	97 million	19,000
Fruit fly (*D. melanogaster*)	137 million	13,000
Yeast (*S. cerevisiae*)	12.1 million	6,000
Bacterium (*E. coli*)	4.6 million	3,200
Human immunodeficiency virus (HIV)	9700	9

The estimated number of human genes is only one-third as great as previously thought, although the numbers may be revised as more computational and experimental analyses are performed.

Scientists suggest that the genetic key to human complexity lies not in gene number but in how gene parts are used to build different products in a process called alternative splicing. Other underlying reasons for greater complexity are the thousands of chemical modifications made to proteins and the repertoire of regulatory mechanisms controlling these processes.
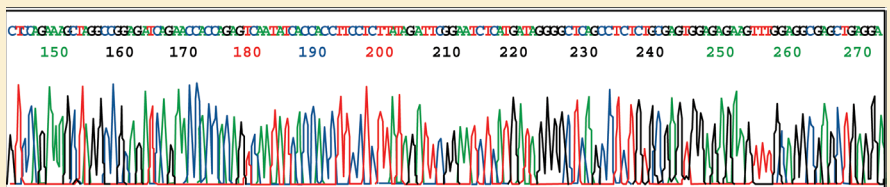
Gene Gateway

A User-Friendly Web Guide to the Human and Other Genomes

All Human Genome Project data and much related information are freely available on the Web, but how does a novice find and use these rich resources? Gene Gateway is a new, nontechnical online guide developed to increase public understanding of and accessibility to this labyrinth of genome science resources. A more technical guide to genomic databases can be found on the *Nature Genetics* site (*www.nature.com*).

Gene Gateway introduces various tools that anyone can use to investigate genetic disorders, chromosomes, genome maps, genes, sequence data, genetic variants, and molecular structures. Major features are described on this page.

Use Gene Gateway tutorials and major scientific Web databases to

a. Explore databases of genes associated with diseases

- Online Mendelian Inheritance in Man (OMIM)
- NCBI LocusLink
- Genome Database (GDB)
- GeneCards

b. Find a gene's location on a chromosome map

- NCBI Map Viewer
- GDB Mapview

c. View the DNA sequence of a gene or amino acid sequence of a protein

- NCBI Sequence Databases
- SWISS-PROT and TrEMBL
- Protein Information Resource–Protein Sequence Database (PIR-PSD)

d. Learn about mutations that cause genetic disorders

- OMIM Allelic Variants
- Human Gene Mutation Database
- HGVbase
- NCBI dbSNP

**e. Examine protein structures**

- NCBI Molecular Database
- Protein Data Bank (PDB)

f. Access related resources such as those on disease support groups, gene tests, and current clinical trials

- Tools for Sequence-Similarity Searching
- Databases of Biomedical Literature
- Genetic Disorder Information
    - Support Groups
    - Genetic Testing
    - Clinical Trials
    - Genetic Health Professionals

Gene Gateway was created as a companion to the Human Genome Landmarks wall poster (see back page)

Major Features of Gene Gateway
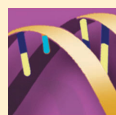
Genome Database Guide
Overviews of databases containing technical information on genes, proteins, and genetic disorders and some search tips for using them.

Bioinformatics Tools
Tips and tutorials for successfully using the databases and other resources described in the Genome Database Guide.
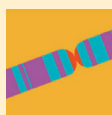
Genetic Disorder Guide
Nontechnical resources on disorder descriptions and treatments, availability of gene tests or clinical trials, support groups, and other general material.

Sample Profiles of Genes and Genetic Disorders
Compilations of information generated for three genetic disorders using the Web resources described above.

Chromosome Viewer
Resource for learning about the physical makeup of human chromosomes and some of the genes that have been found on each one.

Evaluating Medical Information on the Web
General tips and resources for judging the quality of health-related Web sites.

Gene Testing, Pharmacogenomics, and Gene Therapy

DNA underlies almost every aspect of human health, both in function and dysfunction. Obtaining a detailed picture of how genes and other DNA sequences function together and interact with environmental factors ultimately will lead to the discovery of pathways involved in normal processes and in disease pathogenesis. Such knowledge will have a profound impact on the way disorders are diagnosed, treated, and prevented and will bring about revolutionary changes in clinical and public health practice. Some of these transformative developments are described below.

Gene Testing

DNA-based tests are among the first commercial medical applications of the new genetic discoveries. Gene tests can be used to diagnose disease, confirm a diagnosis, provide prognostic information about the course of disease, confirm the existence of a disease in asymptomatic individuals, and, with varying degrees of accuracy, predict the risk of future disease in healthy individuals or their progeny.

Currently, several hundred genetic tests are in clinical use, with many more under development, and their numbers and varieties are expected to increase rapidly over the next decade. Most current tests detect mutations associated with rare genetic disorders that follow Mendelian inheritance patterns. These include myotonic and Duchenne muscular dystrophies, cystic fibrosis, neurofibromatosis type 1, sickle cell anemia, and Huntington's disease.

Recently, tests have been developed to detect mutations for a handful of more complex conditions such as breast, ovarian, and colon cancers. Although they have limitations, these tests sometimes are used to make risk estimates in presymptomatic individuals with a family history of the disorder. One potential benefit to using these gene tests is that they could provide information to help physicians and patients manage the disease or condition more effectively. Regular colonoscopies for those having mutations associated with colon cancer, for instance, could prevent thousands of deaths each year.

Some scientific limitations are that the tests may not detect every mutation associated with a particular condition (many are as yet undiscovered), and the ones they do detect may present different risks to different people and populations. Another important consideration in gene testing is the lack of effective treatments or preventive measures for many diseases and conditions now being diagnosed or predicted.

Revealing information about the risk of future disease can have significant emotional and psychological effects as well. Moreover, the absence of privacy and legal protections can lead to discrimination in employment and insurance or other misuse of personal genetic information. Additionally, because genetic tests reveal information about individuals and their families, test results can affect family dynamics. Results also can pose risks for population groups if they lead to group stigmatization.

Other issues related to gene tests include their effective introduction into clinical practice, the regulation of laboratory quality assurance, the availability of testing for rare diseases, and the education of healthcare providers and patients about correct interpretation and attendant risks.

Families or individuals who have genetic disorders or are at risk for them often seek help from medical geneticists (an M.D. specialty) and genetic counselors (graduate-degree training). These professionals can diagnose and explain disorders, review available options for testing and treatment, and provide emotional support. (For more information, see the Medicine and the New Genetics URL, p. 12.)
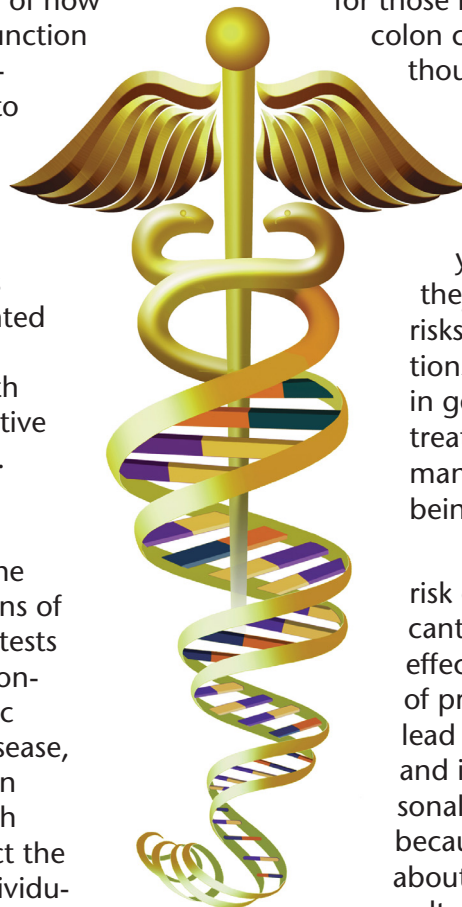
Pharmacogenomics: Moving Away from "One-Size-Fits-All" Therapeutics

Within the next decade, researchers will begin to correlate DNA variants with individual responses to medical treatments, identify particular subgroups of patients, and develop drugs customized for those populations. The discipline that blends pharmacology with genomic capabilities is called pharmacogenomics.

More than 100,000 people die each year from adverse responses to medications that may be beneficial to others. Another 2.2 million experience serious reactions, while others fail to respond at all. DNA variants in genes involved in drug metabolism, particularly the cytochrome P450 multigene family, are the focus of much current research in this area. Enzymes encoded by these genes are responsible for metabolizing most drugs used today, including many for treating psychiatric, neurological, and cardiovascular diseases. Enzyme function affects patient responses to both the drug and the dose. Future advances will enable rapid testing to determine the patient's genotype and guide treatment with the most effective drugs, in addition to drastically reducing adverse reactions.

Genomic data and technologies also are expected to make drug development faster, cheaper, and more effective. Most drugs today are based on about 500 molecular targets; genomic knowledge of the genes involved in diseases, disease pathways, and drug-response sites will lead to the discovery of thousands of new targets. New drugs, aimed at specific sites in the body and at particular biochemical events leading to disease, probably will cause fewer side effects than many current medicines. Ideally, the new genomic drugs could be given earlier in the disease process. As knowledge becomes available to select patients most likely to benefit from a potential drug, pharmacogenomics will speed the design of clinical trials to bring the drugs to market sooner.

Gene Therapy, Enhancement

The potential for using genes themselves to treat disease or enhance particular traits has captured the imagination of the public and the biomedical community. This largely experimental field—gene transfer or gene therapy—holds potential for treating or even curing such genetic and acquired diseases as cancers and AIDS by using normal genes to supplement or replace defective genes or bolster a normal function such as immunity.

More than 600 clinical gene-therapy trials involving about 3500 patients were identified worldwide in 2002.* The vast majority take place in the United States (81%), followed by Europe (16%). Although most trials focus on various types of cancer, studies also involve other multigenic and monogenic, infectious, and vascular diseases. Most current protocols are aimed at establishing the safety of gene-delivery procedures rather than effectiveness.

Gene transfer still faces many scientific obstacles before it can become a practical approach for treating disease. According to the American Society of Human Genetics' Statement on Gene Therapy, effective progress will be achieved only through continued rigorous research on the most fundamental mechanisms underlying gene delivery and gene expression in animals.
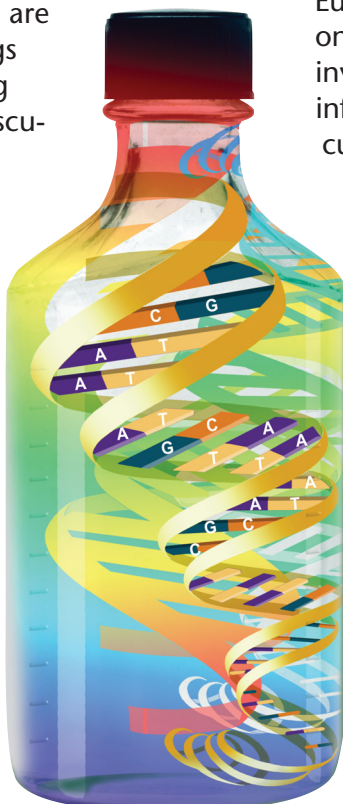
*Source: *Journal of Gene Medicine* Web site (*www.wiley.co.uk*), accessed March 2003.

Other Anticipated Benefits of Genetic Research

Technologies, Resources Having Major Impacts

Rapid progress in genome science and a glimpse into its potential applications have spurred observers to predict that biology will be the foremost science of the 21st Century. Technology and resources generated by the Human Genome Project and other genomic research already are having major impacts on research across the life sciences. Doubling in size in 10 years, the biotechnology industry generated 191,000 direct jobs and 535,000 indirect jobs in 2001. Revenues for that year totaled more than $20 billion directly and $28.5 billion indirectly.*

A list of some current and potential applications of genome research follows. More studies and public discussion are required for eventual validation and implementation of some of these uses (see p. 8).

Molecular Medicine

- Improve diagnosis of disease
- Detect genetic predispositions to disease
- Create drugs based on molecular information
- Use gene therapy and control systems as drugs
- Design "custom drugs" based on individual genetic profiles

Microbial Genomics

- Rapidly detect and treat pathogens (disease-causing microbes) in clinical practice
- Develop new energy sources (biofuels)
- Monitor environments to detect pollutants
- Protect citizenry from biological and chemical warfare
- Clean up toxic waste safely and efficiently

Risk Assessment

- Evaluate the health risks faced by individuals who may be exposed to radiation (including low levels in industrial areas) and to cancer-causing chemicals and toxins

---

*Source: Biotechnology Industry Organization Web site (*www.bio.org*), accessed March 2003.

Bioarchaeology, Anthropology, Evolution, and Human Migration

- Study evolution through germline mutations in lineages
- Study migration of different population groups based on maternal genetic inheritance
- Study mutations on the Y chromosome to trace lineage and migration of males
- Compare breakpoints in the evolution of mutations with ages of populations and historical events

DNA Identification

- Identify potential suspects whose DNA may match evidence left at crime scenes
- Exonerate persons wrongly accused of crimes
- Identify crime, catastrophe, and other victims
- Establish paternity and other family relationships
- Identify endangered and protected species as an aid to wildlife officials (could be used for prosecuting poachers)
- Detect bacteria and other organisms that may pollute air, water, soil, and food
- Match organ donors with recipients in transplant programs
- Determine pedigree for seed or livestock breeds
- Authenticate consumables such as caviar and wine

Agriculture, Livestock Breeding, and Bioprocessing

- Grow disease-, insect-, and drought-resistant crops
- Breed healthier, more productive, disease-resistant farm animals
- Grow more nutritious produce
- Develop biopesticides
- Incorporate edible vaccines into food products
- Develop new environmental cleanup uses for plants like tobacco

Societal Concerns Arising from the New Genetics

Critical Policy and Ethical Issues

Since its inception, the Human Genome Project has dedicated funds toward identifying and addressing the ethical, legal, and social issues surrounding the availability of the new data and capabilities. Examples of such issues follow.*

- **Privacy and confidentiality of genetic information.** *Who owns and controls genetic information? Is genetic privacy different from medical privacy?*

- **Fairness in the use of genetic information by insurers, employers, courts, schools, adoption agencies, and the military, among others.** *Who should have access to personal genetic information, and how will it be used?*

- **Psychological impact, stigmatization, and discrimination due to an individual's genetic makeup.** *How does personal genetic information affect self-identity and society's perceptions?*

- **Reproductive issues including adequate and informed consent and the use of genetic information in reproductive decision making.** *Do healthcare personnel properly counsel parents about risks and limitations? What are the larger societal issues raised by new reproductive technologies?*

- **Clinical issues including the education of doctors and other health-service providers, people identified with genetic conditions, and the general public; and the implementation of standards and quality-control measures.** *How should health professionals be prepared for the new genetics? How can the public be educated to make informed choices? How will genetic tests be evaluated and regulated for accuracy, reliability, and usefulness?* (Currently, there is little regulation.) *How does society balance current scientific limitations and social risk with long-term benefits?*

- **Fairness in access to advanced genomic technologies.** *Who will benefit? Will there be major worldwide inequities?*

- **Uncertainties associated with gene tests for susceptibilities and complex conditions (e.g., heart disease, diabetes, and Alzheimer's disease).** *Should testing be performed when no treatment is available or when interpretation is unsure? Should children be tested for susceptibility to adult-onset diseases?*

- **Conceptual and philosophical implications regarding human responsibility, free will vs genetic determinism, and concepts of health and disease.** *Do our genes influence our behavior, and can we control it? What is considered acceptable diversity? Where is the line drawn between medical treatment and enhancement?*

- **Health and environmental issues concerning genetically modified (GM) foods and microbes.** *Are GM foods and other products safe for humans and the environment? How will these technologies affect developing nations' dependence on industrialized nations?*

- **Commercialization of products including property rights (patents, copyrights, and trade secrets) and accessibility of data and materials.** *Will patenting DNA sequences limit their accessibility and development into useful products?*


*For more information, see the Ethical, Legal, and Social Issues URL, p. 12.

Beyond the Human Genome Project—What's Next?

Genome Sequences Launch a New Level of Scientific Challenges

Building a "Systems Level" View of Life

The DNA sequences generated in hundreds of genome projects now provide scientists with the "parts lists" containing instructions for how an organism builds, operates, maintains, and reproduces itself while responding to various environmental conditions. But we still have very little knowledge of how cells use this information to "come alive." The functions of most genes remain unknown. Nor do we understand how genes and the proteins they encode interact with each other and with the environment. If we are to realize the potential of the genome projects, with far-ranging applications to such diverse fields as medicine, energy, and the environment, we must obtain this new level of knowledge.
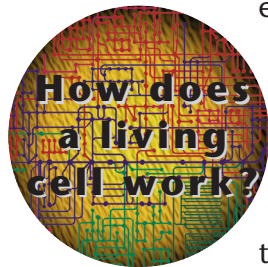
One of the greatest impacts of having whole-genome sequences and powerful new genomic technologies may be an entirely new approach to conducting biological research. In the past, researchers studied one or a few genes or proteins at a time. Because life doesn't operate in such isolation, this inherently provided incomplete—and often inaccurate—views. Researchers now can approach questions systematically and on a much grander scale. They can study all the genes expressed in a particular environment or all the gene products in a specific tissue, organ, or tumor. Other analyses will focus on how tens of thousands of genes and proteins work together in interconnected networks to orchestrate the chemistry of life—a new field called "systems biology" (see "Genomes to Life," p. 10).

Charting Human Variation

Slight variations in our DNA sequences can have a major impact on whether or not we develop a disease and on our responses to such environmental factors as infectious microbes, toxins, and drugs. One of the most common types of sequence variation is the single nucleotide polymorphism (SNP). SNPs are sites in the human genome where individuals differ in their DNA sequence, often by a single base. For example, one person might have the DNA base A where another might have C, and so on. Scientists believe the human genome has at least 10 million SNPs, and they are generating different types of maps of these sites, which can occur both in genes and noncoding regions.

Sets of SNPs on the same chromosome are inherited in blocks (haplotypes). In 2002 a consortium of researchers from six countries established a 3-year effort to construct a map of the patterns of SNPs that occur across populations in Africa, Asia, and the United States. Researchers hope that dramatically decreasing the number of individual SNPs to be scanned will provide a shortcut for tracking down the DNA regions associated with common complex diseases such as cancer, heart disease, diabetes, and some forms of mental illness. The new map also may be useful in understanding how genetic variation contributes to responses to environmental factors. (For more information, see the NIH URL, p. 12.)

How do genetic variations (SNP patterns) differ across populations?

For More Information

Related Web Sites

- **Human Genome Project and Beyond**
  www.ornl.gov/hgmis/
- **Medicine and the New Genetics**
  www.ornl.gov/hgmis/medicine/medicine.html
- **Ethical, Legal, and Social Issues**
  www.ornl.gov/hgmis/elsi/elsi.html
- **Genomes to Life**
  DOEGenomesToLife.org
- **Gene Gateway**
  www.ornl.gov/hgmis/posters/chromosome/
- **Image Gallery (downloadable)**
  www.ornl.gov/hgmis/education/images.html
- **Resources for Teachers**
  www.ornl.gov/hgmis/education/education.html
- **Resources for Students**
  www.ornl.gov/hgmis/education/students.html
- **Careers in Genomics**
  www.ornl.gov/hgmis/education/careers.html
- **DOE Joint Genome Institute**
  www.jgi.doe.gov

- **DOE Genome Research Programs**
  www.ornl.gov/hgmis/
- **NIH National Human Genome Research Institute**
  www.nhgri.nih.gov
- **Genomes OnLine Database (GOLD)**
  wit.integratedgenomics.com/GOLD/
- **National Center for Biotechnology Information**
  www.ncbi.nlm.nih.gov
- **The Institute for Genomic Research**
  www.tigr.org

Genomics and its Impact on Science and Society: The Human Genome Project and Beyond

Produced in March 2003 by the Human Genome Management Information System (HGMIS) at Oak Ridge National Laboratory, Oak Ridge, Tennessee, for the U.S. Department of Energy Office of Biological and Environmental Research.

Free Copies of This Primer

- **Individual print copies or class sets**
  Contact: *caseydk@ornl.gov*, 865/574-0597

- **Downloadable print copies via the Web**
  A PDF version of this document and accompanying PowerPoint slides are accessible via the Web (*www.ornl.gov/hgmis/publicat/primer/*).

Please send comments to e-mail address above.

Free Wall Poster of Human Chromosomes and Genes

A poster depicting ideograms of the human chromosomes and many mapped genes may be ordered via the Web (*www.ornl.gov/hgmis/posters/chromosome/*) or from HGMIS (see contact information below left). Sidebars explain genetic terms and provide URLs for finding more detailed information. [See also Web companion, "Gene Gateway," p. 4]