

Genomics and Its Impact on Science and Society

The Human Genome Project and Beyond

U.S. Department of Energy Genome Research Programs: genomics.energy.gov

A Primer

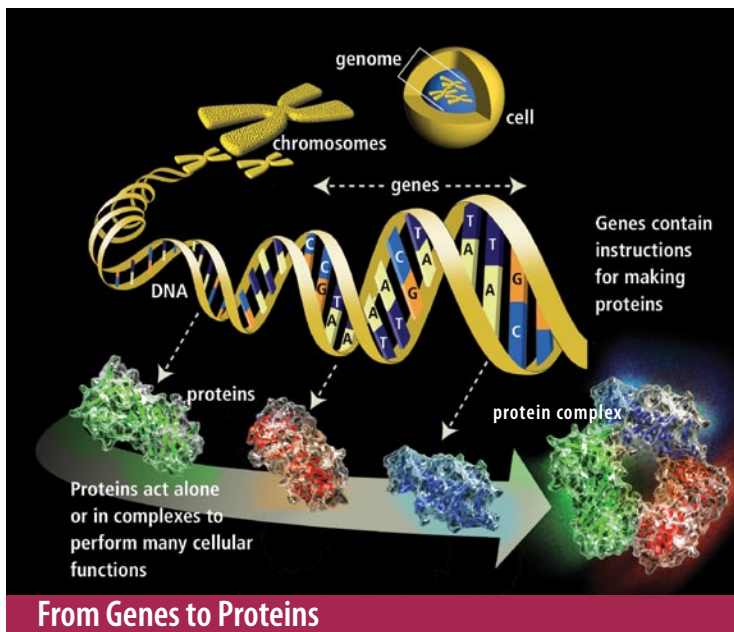
Cells are the fundamental working units of every living system. All the instructions needed to direct their activities are contained within the chemical DNA (deoxyribonucleic acid).

DNA from all organisms is made up of the same chemical and physical components. The **DNA sequence** is the particular side-by-side arrangement of bases along the DNA strand (e.g., ATTCCGGA). This order spells out the exact instructions required to create a particular organism with its own unique traits.

The **genome** is an organism's complete set of DNA. Genomes vary widely in size: The smallest known genome for a free-living organism (a bacterium) contains about 600,000 DNA base pairs, while human and mouse genomes have some 3 billion (see p. 3). Except for mature red blood cells, all human cells contain a complete genome.

DNA in each human cell is packaged into 46 **chromosomes** arranged into 23 pairs. Each chromosome is a physically separate molecule of DNA that ranges in length from about 50 million to 250 million base pairs. A few types of major chromosomal abnormalities, including missing or extra copies or gross breaks and rejoins (translocations), can be detected by microscopic examination. Most changes in DNA, however, are more subtle and require a closer analysis of the DNA molecule to find perhaps single-base differences.

Each chromosome contains many **genes**, the basic physical and functional units of heredity. Genes are specific sequences of bases that encode instructions on how to make proteins. Genes comprise only about 2% of the human genome; the remainder consists of noncoding regions, whose functions may include providing chromosomal structural integrity and regulating where, when, and in what quantity proteins are made. The human genome is estimated to contain some 25,000 genes.



Although genes get a lot of attention, the **proteins** perform most life functions and even comprise the majority of cellular structures. Proteins are large, complex molecules made up of chains of small chemical compounds called amino acids. Chemical properties that distinguish the 20 different amino acids cause the protein chains to fold up into specific three-dimensional structures that define their particular functions in the cell.

The constellation of all proteins in a cell is called its **proteome**. Unlike the relatively unchanging genome, the dynamic proteome changes from minute to minute in response to tens of thousands of intra- and extracellular environmental signals. A protein's chemistry and behavior are determined by the gene sequence and by the number and identities of other proteins made in the same cell at the same time and with which it associates and reacts. Studies to explore protein structure and activities, known as proteomics, will be the focus of much research for decades to come and will help elucidate the molecular basis of health and disease.



U.S. DEPARTMENT OF
ENERGY



**Download
PowerPoint Files**

see p. 12

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 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 analytical technology and data arising from the HGP present 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, it also was 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 chimpanzee, pig, sheep, and domestic cat, 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 of 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.

Insights from the Human DNA Sequence

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 follow from the first publications analyzing the sequence.

- The human genome contains 3.2 billion chemical nucleotide base pairs (A, C, T, and G).
- The average gene consists of 3,000 base pairs, but sizes vary greatly, with the largest known human gene being dystrophin at 2.4 million base pairs.
- Functions are unknown for more than 50% of discovered genes.
- The human genome sequence is almost exactly the same (99.9%) 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 predicted human proteins share similarity with fruit-fly or worm proteins.

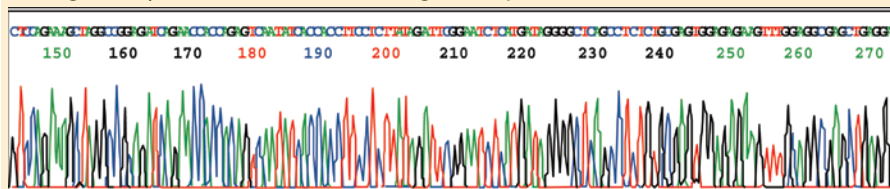
- Genes appear to be concentrated in random areas along the genome, with vast expanses of noncoding DNA between.
- Chromosome 1 (the largest human chromosome) has the most genes (3,168), and Y chromosome has the fewest (344).
- Particular gene sequences have been associated with numerous diseases and disorders, including breast cancer, muscle disease, deafness, and blindness.
- Scientists have identified millions of 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 (Base Pairs)	Estimated Genes
Human (<i>Homo sapiens</i>)	3.2 billion	25,000
Laboratory mouse (<i>M. musculus</i>)	2.6 billion	25,000
Mustard weed (<i>A. thaliana</i>)	100 million	25,000
Roundworm (<i>C. elegans</i>)	97 million	19,000
Fruit fly (<i>D. melanogaster</i>)	137 million	13,000
Yeast (<i>S. cerevisiae</i>)	12.1 million	6,000
Bacterium (<i>E. coli</i>)	4.6 million	3,200
Human immunodeficiency virus (HIV)	9,700	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.



Managing and Using the Data

Bioinformatics Boom

Massive quantities of genomic data and high-throughput technologies are now enabling studies on a vastly larger scale than ever before. Examples include simultaneously monitoring and comparing the activity of tens of thousands of genes in cancerous and noncancerous tissue. Advanced computational tools and interdisciplinary experts are needed to capture, represent, store, integrate, distribute, and analyze the data.

Bioinformatics is the term coined for the new field that merges biology, computer science, and information technology to manage and analyze the data, with the ultimate goal of understanding and modeling living systems. Computing and information demands will continue to rise with the explosive torrent of data from large-scale studies at the molecular, cellular, and whole-organism levels.

Gene Gateway: A User-Friendly Guide to Genome, Gene, and Protein Databases

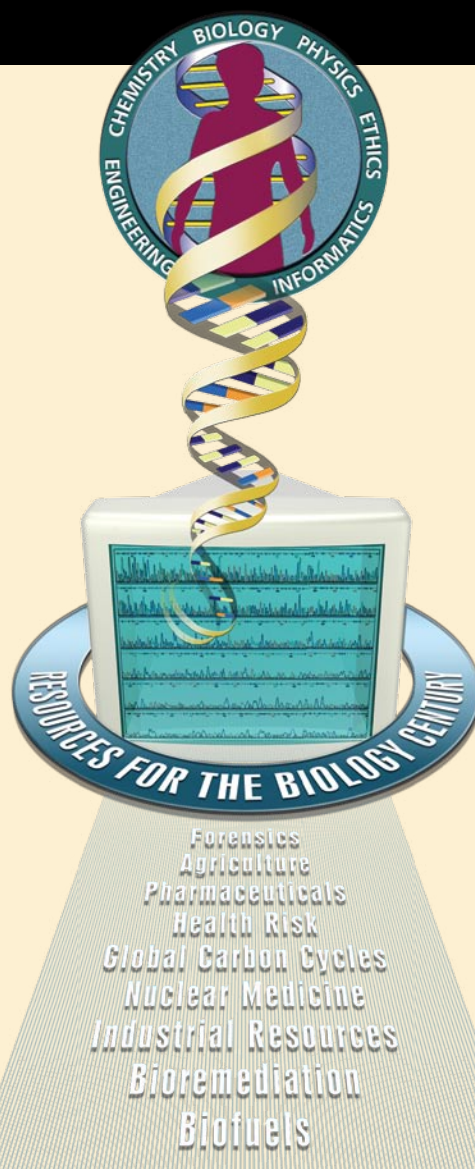
 <http://genomics.energy.gov/genegateway/>

All Human Genome Project data and much related information are freely available on the web, but how do you find and use these rich resources? The Gene Gateway website provides introductory guides and step-by-step tutorials that show how to access and explore genome, gene, and protein databases used by scientists. Gene Gateway demonstrates how to gather information from different databases to gain a better understanding of the molecular biology behind life processes. The site offers a free workbook downloadable in PDF format.

Tutorials

- Identify genes associated with various genetic conditions and biological processes
- Learn about mutations that cause genetic disorders
- Browse a genome and find a gene's location on a chromosome map
- View the DNA sequence of a gene or amino acid sequence of a gene's protein product
- Visualize and modify three-dimensional representations of protein structures

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



Medicine and the New Genetics

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 work 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.

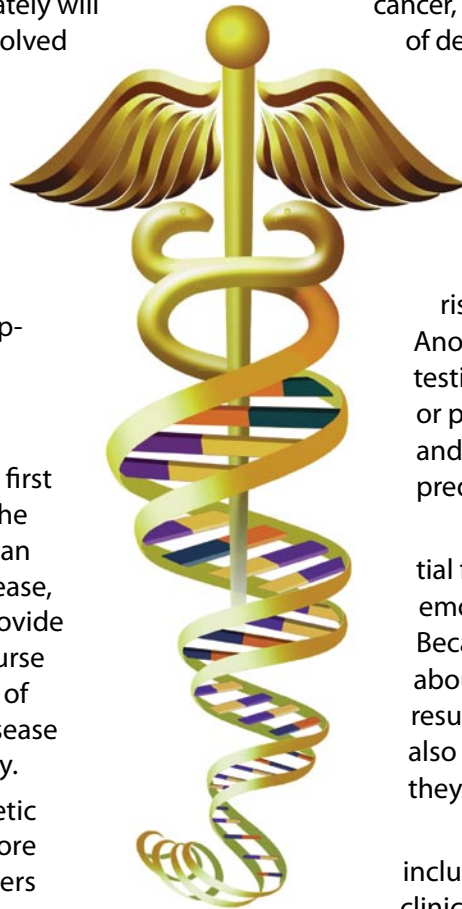
Gene Testing

DNA-based tests are among the first commercial medical applications of the new genetic discoveries. Gene tests can be used to diagnose and confirm disease, even in asymptomatic individuals; provide prognostic information about the course of disease; 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

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 various 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.

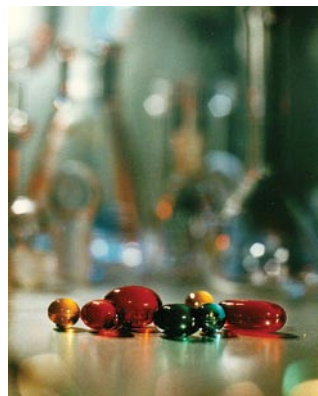
Knowledge about the risk of potential future disease can produce significant emotional and psychological impacts. Because genetic tests reveal information about individuals and their families, test results can affect family dynamics. Results also can pose risks for population groups if they lead to group 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 and 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 URL for Medicine and the New Genetics, 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, but genomic knowledge of genes involved in diseases, disease pathways, and drug-response sites will lead to the discovery of thousands of

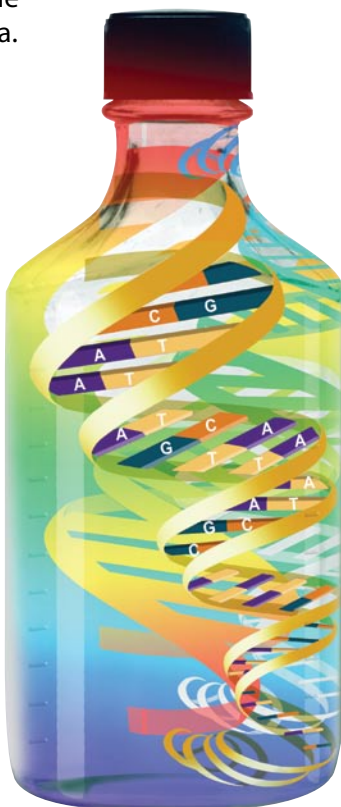
additional targets. New drugs, aimed at specific sites in the body and at particular biochemical events leading to disease, probably will cause fewer side effects than many current medicines. Ideally, genomic drugs could be given earlier in the disease process. As knowledge becomes available to select patients most likely to benefit from a potential drug, pharmacogenomics will speed the design of clinical trials to market the drugs 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 to bolster a normal function such as immunity.

Almost 1,200 clinical gene-therapy trials were identified worldwide in 2006.* The majority (67%) take place in the United States, followed by Europe (29%). 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* website (www.wiley.co.uk/genetherapy/clinical/), August 2006.

Other Anticipated Benefits of Genetic Research

Expanding Impacts of New Technologies, Resources

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 across the life sciences. The biotechnology industry employed more than 250,000 people in 2006, and revenues for 2005 totaled more than \$50.7 billion.* Future revenues are expected to reach trillions of dollars.

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

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 population ages and historical events

DNA Identification

- Identify potential suspects whose DNA may match evidence left at crime scenes
- Exonerate people 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 (e.g., to prosecute poachers)
- Detect bacteria and other organisms that could 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
- Optimize crops for bioenergy production
- 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 such as tobacco

*Source: Biotechnology Industry Organization website (www.bio.org), June 2008.

Societal Concerns Arising from the New Genetics

Critical Policy and Ethical Issues

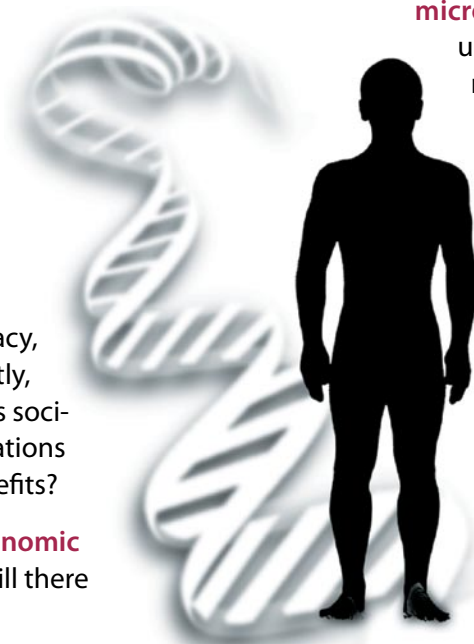
From its inception, the Human Genome Project dedicated funds to identify and address the ethical, legal, and social issues surrounding the availability of new genetic data and capabilities. Examples of such issues follow.*

- **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 health-care personnel properly counsel parents about risks and limitations? What larger societal issues are 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 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?

New Genetics Privacy Act Becomes Law

The Genetic Information Nondiscrimination Act (GINA) became law on May 21, 2008. GINA prohibits U.S. health insurance companies and employers from discrimination on the basis of information derived from genetic tests. In addition, insurers and employers are not allowed under the law to request or demand a genetic test.

- **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 understanding 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?



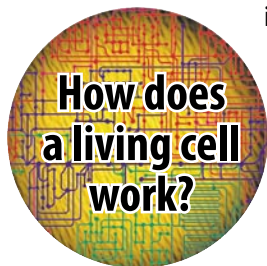
*For more information, see the Ethical, Legal, and Social Issues URL, p. 12.

Beyond the Human Genome Project—What's Next?

Genome Sequences: Paving the Way for a More Comprehensive Understanding

Building a “Systems Level” View of Life

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. We still have very little knowledge of how cells use this information to “come alive,” however, and 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 biological processes are intertwined, these strategies 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. These holistic studies are the focus of a new field called “systems biology” (see DOE Genomics:GTL Program, 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 a 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 in both genes and noncoding regions.

Sets of SNPs on the same chromosome are inherited in blocks (haplotypes). In 2005 a consortium of researchers from six countries completed the first phase of a map of SNP patterns 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 such common complex diseases 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.

How Do Genetic Variations (SNP Patterns) Differ Across Populations?



Exploring Genomes for Energy and Environmental Applications

The Genomics:GTL (formerly Genomes to Life) program of the U.S. Department of Energy (DOE) is using the Human Genome Project's technological achievements to help solve our growing energy and environmental challenges.

Today, genomics is the starting point for a new level of exploration across the life sciences. The GTL research program uses genomic (DNA) sequences of microbes and plants to launch large-scale investigations into their wide-ranging biochemical capabilities having potential applications in bioenergy and the environment (see sidebar below). Before these biological processes can be safely and economically

harnessed for such uses, however, they must be understood in far greater detail and in the context of their operations within a dynamic, living organism.

To obtain this whole-systems knowledge, GTL investigates relevant plant and microbial properties on multiple levels. Starting with the DNA sequence, studies follow its expression (e.g., protein production, interactions, and regulation) in individual cells and populations of cells or organisms in ecosystems. Integrating genomic and many other data types into a computerized knowledgebase will stimulate new research strategies and insights needed for specialized applications.

GTL Investigations of Microbial and Plant Genomes

Microbes and plants have evolved unique biochemistries, offering a rich resource that can be applied to diverse national needs. Some recent projects funded by the DOE Genomics:GTL program highlight the potential wealth of natural capabilities available.

Plants for Biomass, Carbon Storage

Understanding the genes and regulatory mechanisms controlling growth and other traits in the recently sequenced poplar tree may lead to its use for bioethanol production and for sequestration (storage) of carbon.



Microbes Living in Termites: A Potential Source of Enzymes for Bioenergy Production

GTL researchers are investigating bacteria that live in termite hindguts and churn out wood-digesting enzymes. These proteins may be usable for breaking

down plant cellulose into sugars needed for ethanol production. Termites also produce hydrogen as a by-product, a process that potentially could be reproduced on a larger scale.



Synthetic Nanostructures: Harnessing Microbial Enzyme Functions

Enzymes incorporated into synthetic membranes can carry out some of the functions of living cells and may be useful for generating energy, inactivating contaminants, and sequestering atmospheric carbon.

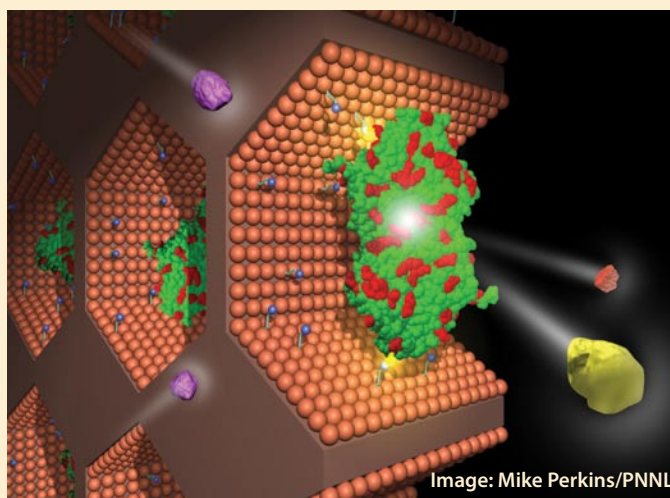


Image: Mike Perkins/PNNL

Genomes for Bioenergy

Cellulosic Biomass: An Abundant, Secure Energy Source to Reduce U.S. Dependence on Gasoline

Bioethanol made from cellulosic biomass—the inedible, fibrous portions of plants—offers a renewable, sustainable, and expandable domestic resource to meet the growing demand for transportation fuels and reduce our dependence on oil.

The United States now produces 7 billion gallons of corn-grain ethanol per year, a fraction of the 142 billion gallons of transportation fuel used annually. Cellulosic ethanol has the potential to dramatically increase the availability of ethanol and help meet the national goal of displacing 30% of gasoline by 2030.

Cellulose is the most abundant biological material on earth. The crops used to make cellulosic ethanol (e.g., postharvest corn plants—not corn grain—and switchgrass) can be grown in most states and often on marginal lands. As with ethanol from corn grain,


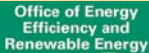
cellulose-based ethanol can be used as a fuel additive to improve gasoline combustion in today's vehicles. Modest engine modifications are required to use higher blends (85% ethanol). Additionally, the amount of carbon dioxide emitted to the atmosphere from producing and burning ethanol is far less than that released from gasoline.

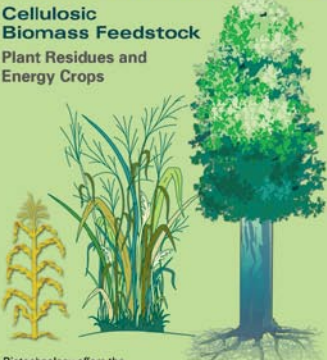
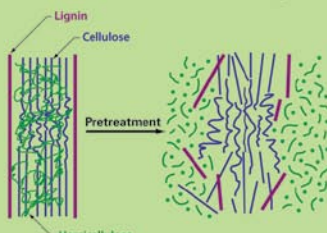
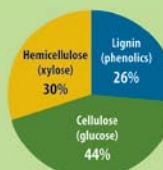
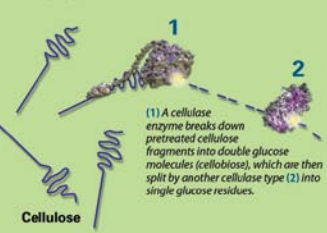
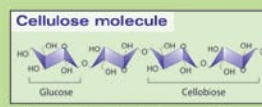
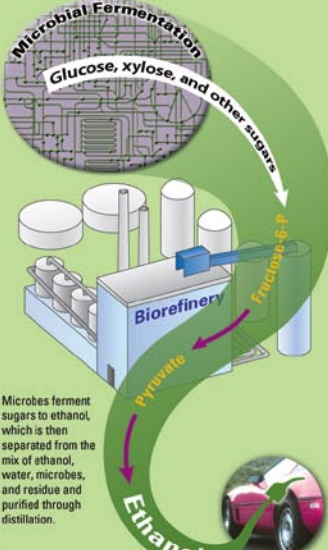
To accelerate technological breakthroughs, the DOE Genomics:GTL program will establish research centers to target specific DOE mission challenges. Three DOE Bioenergy Research Centers are focused on overcoming biological challenges to cellulosic ethanol production. In addition to ethanol, these centers are exploring ways to produce a new generation of petroleum-like biofuels and other advanced energy products from cellulosic biomass.

Download flyer at <http://genomicsgtl.energy.gov/biofuels/placemat.shtml>

U.S. Department of Energy

From BIOMASS to CELLULOSIC ETHANOL: Genomics for Alternative Fuels

<p>Cellulosic Biomass Feedstock Plant Residues and Energy Crops</p>  <p>Biotechnology offers the promise of dramatically increasing ethanol production using cellulose, the most abundant biological material on earth, and other polysaccharides (hemicellulose). Residues including postharvest corn plants (stover) and timber residues could be used, as well as such specialized high-biomass "energy" crops as domesticated poplar trees and switchgrass.</p> <p>Biochemical conversion of cellulosic biomass to ethanol for transportation fuel currently involves three basic steps:</p> <ul style="list-style-type: none"> ▶ Pretreatments to increase the accessibility of cellulose to enzymes and solubilize hemicellulose sugars ▶ Hydrolysis with special enzyme preparations to break down cellulose to sugars ▶ Fermentation to ethanol <p>Making cellulosic biomass conversion to ethanol more economical and practical will require a science base for molecular redesign of numerous enzymes, biochemical pathways, and full cellular systems.</p>	<p>Pretreatment Goal: Make cellulose more accessible to enzymatic breakdown (hydrolysis) and solubilize hemicellulose sugars</p>  <p>Cellulose exists within a matrix of other polymers, primarily hemicellulose and lignin. Pretreatment of biomass with heat, enzymes, or acids removes these polymers from the cellulose core before hydrolysis.</p> <p>Pretreatment, one of the more expensive processing steps, has great potential for improvement through R&D.</p> <p><small>[Figure adapted from N. Mosier et al., 2005. "Features of Promising Technologies for Pretreatment of Lignocellulosic Biomass," <i>Bioresource Technology</i> 96(3): 673-86.]</small></p> <div style="text-align: center;"> <p>Composition of Biomass (lignocellulose)</p>  </div>	<p>Hydrolysis Goal: Break down cellulose into its component sugars using enzyme preparations</p>  <div style="text-align: center;"> <p>Cellulose molecule</p>  <p>Cellulose is made up of double glucose molecules (cellobiose).</p> </div> <p>Enzymes such as cellulases synthesized by fungi and bacteria work together to degrade cellulose and other structural polysaccharides in biomass. Optimizing these complex systems will require a more detailed understanding of their regulation and activity.</p>	<p>Fermentation to Ethanol Goal: Convert sugars to ethanol using microbes</p>  <p>Microbes ferment sugars to ethanol, which is then separated from the mix of ethanol, water, microbes, and residue and purified through distillation.</p>
<p>DOE GTL program contributions needed to</p> <ul style="list-style-type: none"> • Control cell-wall composition for energy production • Develop appropriate model systems for energy crops • Improve quantity and quality of perennial herbaceous and woody biomass crops • Domesticate energy crops for stress tolerance • Develop sustainable management practices 	<p>DOE GTL program contributions needed to</p> <ul style="list-style-type: none"> • Optimize and exploit biological catalysts • Reduce thermochemical treatments and waste • Increase simple sugar yields and concentration <p><small>All recommendations for contributions from the DOE Office of Science's (SC) Genomics:GTL program (formerly Genomes to Life) originate from a December 2005 workshop sponsored by SC and the DOE Office of Energy Efficiency and Renewable Energy. The workshop report and this flyer are available at www.genomicsgtl.energy.gov/biofuels/.</small></p>	<p>DOE GTL program contributions needed to</p> <ul style="list-style-type: none"> • Increase specific activities • Increase thermal tolerance • Reduce product inhibition • Broaden substrate range 	<div style="border: 1px solid black; padding: 5px;"> <p style="text-align: center;">Consolidate Processing Steps</p> <p style="text-align: center;">Integrate hydrolysis and fermentation steps into a single microbe or mixed stable culture that</p> <ul style="list-style-type: none"> • Produces hydrolytic enzymes • Ferments sugars to ethanol • Is process tolerant • Has stable integrated traits </div> <p>DOE GTL program contributions needed to</p> <ul style="list-style-type: none"> • Eliminate solid-liquid separation step • Coferment 5- and 6-carbon sugars from biomass feedstocks • Increase process tolerance and resistance to inhibitors • Return minerals to soils

For More Information

Related Websites

- **Human Genome Project Information**
www.ornl.gov/hgmis/home.shtml
- **Medicine and the New Genetics**
www.ornl.gov/hgmis/medicine/
- **Ethical, Legal, and Social Issues**
www.ornl.gov/hgmis/elsi/
- **Genetics Privacy and Legislation**
www.ornl.gov/hgmis/elsi/legislat.shtml
- **Gene Gateway**
genomics.energy.gov/genegateway/
- **Image Gallery (downloadable)**
genomics.energy.gov/gallery/
- **Resources for Teachers**
www.ornl.gov/hgmis/education/
- **Resources for Students**
www.ornl.gov/hgmis/education/students.shtml
- **Careers in Genetics and the Biosciences**
www.ornl.gov/hgmis/education/careers.shtml
- **DOE Joint Genome Institute**
www.jgi.doe.gov
- **NIH National Human Genome Research Institute**
www.genome.gov
- **Genomes OnLine Database (GOLD)**
www.genomesonline.org
- **National Center for Biotechnology Information**
www.ncbi.nlm.nih.gov

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

This document was revised in June 2008 by the Genome Management Information System (GMIS) at Oak Ridge National Laboratory, Oak Ridge, Tennessee, for the Office of Biological and Environmental Research within the U.S. Department of Energy Office of Science.

Free Copies and Presentation Materials

- **Individual print copies or class sets**
Contact: genome@ornl.gov, 865/574-0597
- **Downloadable files**
A PDF version of this document and accompanying PowerPoint file are accessible via the web (www.ornl.gov/hgmis/publicat/primer).

- **DOE Genomics:GTL Program**
genomicsgtl.energy.gov
- **Focus on Biofuels**
genomicsgtl.energy.gov/biofuels/
- **National Geographic: The Genographic Project**
www.nationalgeographic.com/genographic/
- **International HapMap Project**
www.hapmap.org

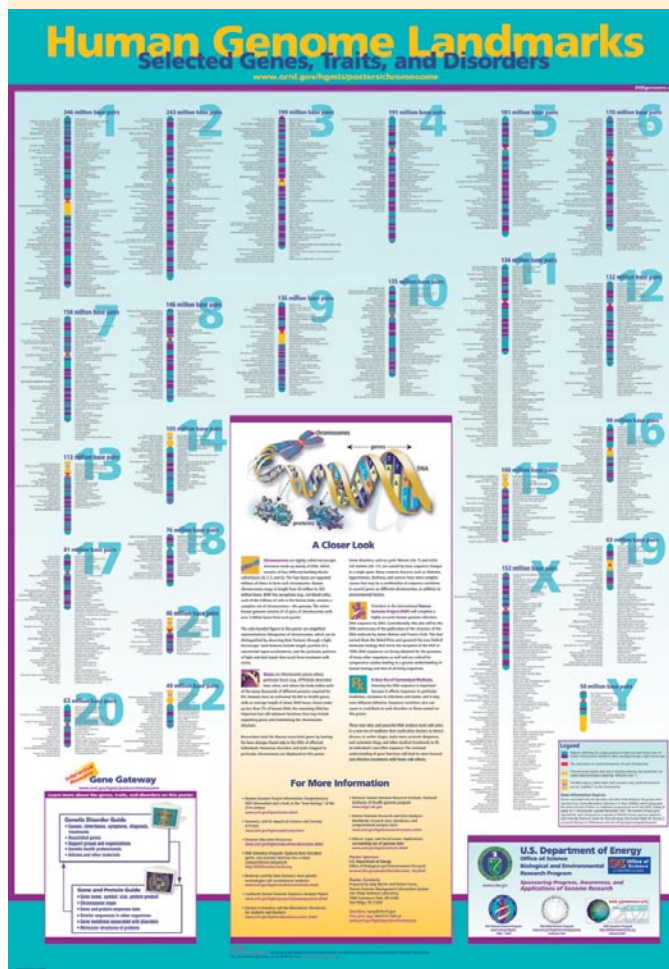
Free Wall Poster of Human Chromosomes and Genes

The poster also features sidebars explaining genetic terms, with URLs for finding more detailed information (see web companion, Gene Gateway, p. 4).

Order free copies via the web:

- genomics.energy.gov/posters/

Or see contact information below left.



June 2008

