Human Genome news

Linking Interdisciplinary Contributors and Users of Genome Data and Resources

U.S. Department of Energy Office of Biological and Environmental Research www.ornl.gov/hgmis/project/about.html

ISSN: 1050–6101, Issue number 46 Vol. 11, Nos. 1–2, November 2000

In This Issue

On the Shoulders of Giants: Private Sector Leverages HGP Successes

Data, Technologies Catalyze a New, High-Profile Life Sciences Industry

The deluge of data and related technologies generated by the Human Genome Project (HGP) and other genomic research presents a broad array of commercial opportunities. Seemingly limitless applications cross boundaries from medicine and food to energy and environmental resources, and predictions are that life sciences may become the largest sector in the U.S. economy.

Established companies are scrambling to retool, and many new ventures are seeking a role in the information revolution with DNA at its core. IBM, Compaq, DuPont, and major pharmaceutical companies are among those interested in the potential for targeting and applying genome data.

In the genomics corner alone, dozens of small companies have sprung up to sell information, technologies, and services to facilitate basic research into genes and their functions. These new entrepreneurs also offer an abundance of genomic services and applications, including additional databases with DNA sequences from humans, animals, plants, and microbes.

Other applications include gene fragments to use for drug development and target identification and evaluation, identification of candidate genes, and RNA expression information revealing gene activity. Products include protein profiles; particular genotypes associated with such specific medically important phenotypes as disease susceptibility and drug responsiveness; hardware, software, and reagents for DNA sequencing and other DNA-based tests; microarrays (DNA chips) containing tens of thousands of known DNA and RNA fragments for research or clinical use; and DNA analysis software.

Broader applications reaching into many areas of the economy include the following:

- **Clinical medicine.** Many more individualized diagnostics and prognostics, drugs, and other therapies.
- **Agriculture and livestock.** Hardier, more nutritious, and healthier crops and animals.
- **Industrial processes.** Cleaner and more efficient manufacturing in such sectors as chemicals, pulp and paper, textiles, food, fuels, metals, and minerals.
- **Environmental biotechnology.** Biodegradable products, new energy resources, environmental diagnostics, and less hazardous cleanup of mixed toxic-waste sites.
- **DNA fingerprinting.** Identification of humans and other animals, plants, and microbes; evolutionary and human anthropological studies; and detection of and resistance to harmful agents that might be used in biological warfare.

From the start, HGP planners anticipated and promoted the private sector's participation in developing and commercializing genomic resources and applications. The HGP's successes

p. 2 ➡

(from p. 1)

in establishing an infrastructure and funding high-throughput technology development are giving rise to commercially viable products and services, with the private sector now taking on more of the risk.

A Public Legacy

Substantial public-sector R&D investment often is needed in feasibility demonstrations before such start-up ventures as those by Celera Genomics, Incyte, and Human Genome Sciences can begin. In turn, these companies furnish valuable commercial services that the government cannot provide, and the taxes returned by their successes easily repay fundamental public investments. Following are a few key public R&D contributions that made some current genomics ventures commercially feasible. These examples describe DOE investments, but substantial commitments by NIH and the Wellcome Trust in the United Kingdom were equally important.

Scientific Infrastructure. The scientific foundation for a human genome initiative existed at the national laboratories before DOE established the first genome project in 1986. Besides expertise in a number of areas critical to genomic research, the laboratories had a long history of conducting large multidisciplinary projects.

Genomic Science and Pioneering Technology. GenBank, the world's DNA sequence repository, was developed at Los Alamos National Laboratory (LANL) and later transferred to the National Library of Medicine. Chromosome-sorting capabilities developed at LANL and Lawrence Livermore National Laboratory enabled the development of DNA clone libraries representing the individual chromosomes. These libraries were a crucial resource in genome sequencing.

Sequencing Strategies. When the HGP was initiated, vital automation tools and high-throughput sequencing technologies had to be developed or improved. The cost of sequencing a single DNA base was about $10 then; today, sequencing costs have fallen

HGP and the Private Sector: Rivals or Partners?

With the June 26 announcement by the publicly funded Human Genome Project (HGP) and Celera Genomics that the draft sequence of the human genome was essentially complete, the complementary aspects of the public and private sectors' sequencing projects were realized.

Since spring 1998, when Celera Genomics announced its sequencing goal, other private companies also have declared their intention to sequence or map genomic regions to varying degrees. Some people questioned whether the HGP and the private sector were duplicating work, and they wondered who would "win" the race to sequence the human genome. Although the HGP and private companies do have overlapping sequencing goals, their "finish lines" are different because their ultimate goals are not the same.

In a sense, through its policy of open data release, the HGP has all along facilitated the research of others. Additionally, the HGP funds projects at small companies to devise needed technologies. DOE, NIH, the National Institute for Standards and Technology, and other governmental funding sources also are supporting further application and commercialization of HGP-generated resources.

HGP products have spurred a boom in such spin-off programs as the NIH Cancer Genome Anatomy Project and the DOE Microbial Genome Program. Genomes of numerous animals, plants, and microbes are being sequenced, and the number of private endeavors is increasing. Technology transfer from developers to users and participation in collaborative, multidisciplinary projects closely unite researchers at academic, industrial, and governmental laboratories.

Scientific vs Commercial Goals

The HGP's commitment from the outset has been to create a scientific standard (an entire reference genome). Most private-sector human genome sequencing projects, however, focus on gathering just enough DNA to meet their customers' needs—probably in the 95% to 99% range for gene-rich, potentially lucrative regions. Such private data continue to be enriched greatly by accurate free public mapping (location) and sequence information.

Continued Support for Life Sciences Industry

A congressional hearing in April of this year presented testimony on the importance of both public and private sectors to future discoveries and the need for continued federal support of the burgeoning life sciences industry (see box, p. 3).

Celera's shotgun sequencing strategy, for example, creates millions of tiny fragments that must be ordered and oriented computationally using HGP research results. Most data at Celera, Incyte, and other genomics information–based companies are proprietary or available only for a fee. In addition, companies are filing numerous patent applications to stake early claims to genes and other potentially important DNA fragments (see p. 9).

More than the Reference Sequence

DNA sequencing will continue to be a major emphasis for the foreseeable future as gene sequences are surveyed across various populations. Both the DOE and NIH genome programs are continuing to support the development of fully integrated and innovative approaches to rapid, low-cost sequencing.

Other near-term HGP goals from the latest 5-year plan are to enhance bioinformatics (computational) resources to support future research and commercial applications. The HGP also aims to explore gene function through comparative mouse-human studies, train future scientists, study human variation, and address critical societal issues arising from the increased availability of human genome data and related analytical technologies. ♦

about 100-fold to $.10 to $.20 a base and still are dropping rapidly.

DOE-funded enhancements to sequencing protocols, chemical reagents, and enzymes contributed substantially to increasing efficiencies. The commercial marketing of these reagents has greatly benefitted basic R&D, genome-scale sequencing, and lower-cost commercial diagnostic services.

Sequencing Technologies and Biological Resources. Other major factors in cost and time reduction are greatly improved sequencing instruments and efficient biological resources such as the following:

- DOE-funded research on capillary-based DNA sequencing contributed to the development of the two major sequencing machines now in use. The core optical system concept of the Perkin-Elmer 3700 sequencing machine (used by Celera and others) was pioneered with DOE support. The instrumentation concepts that matured as the MegaBACE sequencer were pioneered by Richard Mathies (University of California, Berkeley). The DOE JGI chose this sequencing hardware platform after competitive trials.
- DNA sequencing originally was done with radiolabeled DNA fragments. Today, DOE improvements to fluorescent dyes decrease the amount of DNA needed and increase the accuracy of sequencing data.
- Bacterial artificial chromosome (BAC) clones, developed in the DOE program, became the preferred starting resource in sequencing procedures because of their superior stability and large size. A critical component of public- and private-sector sequencing, BACs were used to assemble both the draft and final human DNA reference sequences.
- Further extending the usefulness of BACs, the DOE HGP funded the production of sequence tag connectors (STCs) from BAC ends. This early information enabled the selection of optimal BACs for complete sequencing, thus saving time and money. STC use for the HGP was advocated by Craig Venter and Nobelist Hamilton Smith (both at Celera), and Leroy Hood (now at the Institute for Systems Biology).

A Successful Transformation

These successes transferred much of the repetitive labor from humans to automated machines. In addition, new software for data processing both alleviated and sped human decision making. Over the last decade, advances in instrumentation, automation, and computation have transformed the entire process. Further innovations, however, still are needed for completing many large sequences and increasing the effectiveness of sequencing. *[Denise Casey (HGMIS) and Marvin Stodolsky (DOE)]* ♦

In Memoriam

G. Christian Overton, Founding Director, Center for Bioinformatics, University of Pennsylvania, Philadelphia, was a pioneer in genomic research. His family, friends, and colleagues will miss his charm, wit, good nature, and academic brilliance.

Congressional Hearing Explores Controversies, Benefits of Genomics

In April the Subcommittee on Energy and Environment of the Committee on Science of the U.S. House of Representatives conducted hearings on the status and benefits of genome sequencing in the public and private sectors (*www.house.gov/science/106_hearing.htm#Energy_and_Environment*). Speakers included representatives of the U.S. HGP and Celera Genomics, members of Congress, and the director of the Office of Science and Technology Policy.

Robert Waterston, director of the HGP sequencing center at Washington University, St. Louis, pointed to fruitful data sharing by the HGP and the private sector. Examples include (1) collaborations led by the pharmaceutical company Merck to develop partial sequences identifying genes and (2) the fruit fly sequencing project by Celera and the HGP.

Examples of private-sector enrichment of public data include the SNP consortium, which is generating a publicly available map containing human DNA variations (see SNP articles, p. 10). In September, Celera Genomics announced a reference database with more than 2.8 million unique SNPs, including those screened from public-sector databases. In October a public-private consortium announced the joint sequencing of the laboratory mouse (see article, p. 12). Also, a Monsanto–University of Washington project recently generated a draft sequence of the rice plant genome to be released to the public. These efforts show the value of sharing data to increase knowledge and ensure future discoveries for mutual benefit.

Neal Lane (Assistant to the President for Science and Technology and Director of the Office of Science and Technology Policy) echoed the importance of partnerships between public and private sectors in his testimony to the House committee. His observations follow.

"Sequencing the genome. . . is only the beginning of genomics," he said. "It is the first step into a future of discoveries and innovations that genomics will enable, that the public and private sectors must pursue together. . . . An expanding, evolving partnership has made human genomic discoveries possible and is now poised to make those discoveries beneficial for everyone. . . . I believe that the policies we have pursued will help to strengthen this partnership, allowing genomic discoveries and innovations to move steadily forward for the benefit of our nation and for all humankind." ♦

Human Genome Project Milestones Celebrated at White House

Clinton Calls Working Draft "Starting Point for Even Greater Discoveries"

On June 26, Human Genome Project (HGP) leaders and representatives from the private company Celera Genomics joined President Bill Clinton at the White House to announce the completion of a working draft reference DNA sequence of the human genome. Clinton observed that the working draft is a "starting point for even greater discoveries."

This achievement provides scientists worldwide with a virtual road map to an estimated 95% of all genes. All HGP data are available on the Internet, and publication in *Science* and *Nature* is expected early in 2001.

Craig Venter (head of Celera Genomics), Ari Patrinos (director, DOE Human Genome Program and Biological and Environmental Research Program), and Francis Collins (director, NIH National Human Genome Research Institute).

The draft contains gaps and errors, but it provides a valuable scaffold for generating the *high-quality* reference genome sequence—the ultimate HGP goal expected to be achieved by 2003 or sooner. This knowledge will speed the understanding of how genetics influences disease development, aid scientists looking for genes associated with particular diseases, and contribute to the discovery of new treatments.

Ari Patrinos, head of the DOE Human Genome Program, led a series of meetings this year at his home that resulted in the joint announcement and agreement by the public- and private-sector projects to publish at the same time.

Speaking of the value of genome data and technologies, Patrinos said, "We are eager to offer a future to our children and grandchildren in which 'cancer' will be only a constellation in the sky."

CD-ROM, Video

A CD-ROM and educational video on the HGP, sponsored by DOE and NIH and a number of other organizations, will be released in 2001. Contact: HGMIS, p. 16, to order.

"Researchers in a few years will have trouble imagining how we studied human biology without genome sequence in front of us," said Francis Collins, head of the NIH genome program.

More than $3 billion has been spent worldwide on the Human Genome Project since its formal inception in 1990 (see box, p. 5, for U.S. costs since 1987).

Although 16 institutions participate in the HGP, most sequencing takes place at 5 locations. These are the DOE Joint Genome Institute, Washington University (St. Louis), Sanger Centre (U.K.), Baylor College of Medicine, and Whitehead Institute.

Bioinformatics teams at the Ensembl database project and the University of California, Santa Cruz, generated an ordered view of the 400,000 sequenced DNA fragments in the working draft.

In July, the Wellcome Trust (U.K.) announced a 5-year investment in Ensembl of more than $14 million (£8.8 million) for automatic annotation of human genome data, including identification of genes and other biologically important sequence features.

Human Genome Project FAQs

Working Draft vs Finished Sequence: What's the Difference?

In generating the draft sequence, scientists determined the order of base pairs in each chromosomal area at least 4 to 5 times (4× to 5×) to ensure data accuracy and to help with reassembling DNA fragments in their original order. This repeated sequencing is known as genome "depth of coverage." Draft sequence data are mostly in the form of 10,000 bp–sized fragments whose approximate chromosomal locations are known.

To generate finished high-quality sequence, additional sequencing is needed to close gaps, reduce ambiguities, and allow for only a single error every 10,000 bases, the agreed-upon standard for HGP finished sequence. Investigators believe that a high-quality sequence is critical for recognizing regulatory components of genes that are very important in understanding human biology and such disorders as heart disease, cancer, and diabetes. The finished version will provide an estimated 8× to 9× coverage of each chromosome. Thus far, finished sequences have been generated for only two human chromosomes–21 and 22 (see article, p. 7).

When is a Genome Completely Sequenced?

In December 1999, the 56-Mb sequence of human chromosome 22 was declared essentially complete, yet only 33.5 Mb were sequenced. In early spring of this year, the fruit fly *Drosophila's* 180-Mb genome also was announced as completed, although just 120 Mb were characterized. What's the deal?

Animal genomes have large DNA regions that currently cannot be cloned or assembled. In the human genome sequence, these regions include telomeres and centromeres (chromosome tips and centers), as well as many chromosomal areas packed with other types of sequence repeats.

Most unsequenceable areas contain heterochromatic DNA, which has few genes and many repeated regions that are difficult to maintain as clones for DNA sequencing. HGP scientists strive to sequence the entire euchromatic DNA, which generally is defined as gene-rich areas (including both exons and introns) that are translated into RNA during gene expression. In the case of human

Lowering Public, Private Costs

The project's early phase was characterized by efforts to generate the biological, instrumentational, and computational resources necessary for efficient production-scale DNA sequencing. Pilot studies on large-scale sequencing began in 1996, and successes led to a ramp up in 1998.

In 1999, international HGP leaders set the accelerated goal of completing a rough draft of all 24 human chromosomes a year ahead of schedule. This ever-increasing pace was facilitated by the commercialization of a new generation of automated capillary DNA sequencing machines and by BACs (DNA fragments) pioneered in DOE-sponsored projects. Researchers in both the public and private sectors use BACs to speed their sequencing procedures (see articles, pp. 1–3).

The extraordinary achievements of the HGP stand as a testimony to the successful collaborations among scientists intent on overcoming massive technological challenges to move toward the common goal of understanding life at its most basic level.

The situation today is well captured by the words of Winston Churchill, who said in November 1942, after 3 years of war, "Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning."

U.S. Human Genome Project Funding ($ Millions)

FY	DOE*	NIH	U.S. Total
1987	5.5	0	5.5
1988	10.7	17.2	27.9
1989	18.5	28.2	46.7
1990	27.2	59.5	86.7
1991	47.4	87.4	134.8
1992	59.4	104.8	164.2
1993	63.0	106.1	169.1
1994	63.3	127.0	190.3
1995	68.7	153.8	222.5
1996	73.9	169.3	243.2
1997	77.9	188.9	266.8
1998	85.5	218.3	303.8
1999	89.9	225.7	315.6
2000	88.9	271.7	360.6

*Note: These numbers do not include construction funds, which are a very small part of the budget.

And so it is for the new biology. [See "Post-Sequencing Research Challenges," p. 7.] ♦

HGP Data Sites

Sites with Assembled Human Genome (including Browsing Tools)

European Bioinformatics Institute
www.ensembl.org

National Center for Biotechnology Information
www.ncbi.nlm.nih.gov/genome/guide
(click on "Map Viewer")

University of California, Santa Cruz
http://genome.ucsc.edu

Other Sites

Baylor College of Medicine
www.hgsc.bcm.tmc.edu

Computational Biosciences, ORNL
http://compbio.ornl.gov/tools
http://genome.ornl.gov/jgi

DNA Data Bank of Japan
www.ddbj.nig.ac.jp

DOE Joint Genome Institute
www.jgi.doe.gov

European Bioinformatics Institute
www.ebi.ac.uk

Genome Database
http://gdbwww.gdb.org

GenBank
http://www.ncbi.nlm.nih.gov/genbank

Sanger Centre
www.sanger.ac.uk

Stanford Human Genome Center
www-shgc.stanford.edu

Washington University, St. Louis
http://genome.wustl.edu/gsc

Whitehead Institute
www-genome.wi.mit.edu

See Web site for answers to many more "Frequently Asked Questions": *www.ornl.gov/hgmis/faq/faqs1.html*

chromosome 22, the sequenced 60% represents 97% of euchromatic DNA. Similarly, nearly all the euchromatic regions were sequenced for *Drosophila*.

Although the HGP goal is to have complete strings of sequence for each chromosome from tip to tip, obtaining this high level of resolution presents a great challenge.

Whose Genomes Are Being Sequenced?

Diversity Represented

All humans share the same basic set of genes and genomic regulatory regions that control the development and maintenance of biological structures and processes. Therefore, the human reference sequence will not, and does not need to, represent an exact match for any one person's genome.

Investigators are using DNA from donors representing widely diverse populations. For example, HGP researchers collected samples of blood (female) or sperm (male) from a large number of people; only a few samples were processed, with source names protected so neither donors nor scientists know whose genomes are being sequenced. The private company Celera Genomics collected samples from five individuals who identified themselves as Hispanic, Asian, Caucasian, or African-American.

In addition to generating the reference sequence, another important HGP goal is to identify many of the small DNA regions that vary among individuals and could underlie disease susceptibility and drug responsiveness. The most common variations are called SNPs (single nucleotide polymorphisms). The DNA resources used for these studies came from 24 anonymous donors of European, African, American (north, central, south), and Asian ancestry.

Although the sequence information will come from the DNA of many persons, it will be applicable to everyone.

Why DOE?

DOE's role in the HGP arose from the historic congressional mandate of its predecessor agencies (the Atomic Energy Commission and the Energy Research and Development Administration) to study the genetic and health effects of radiation and chemical by-products of energy production. From this work the recognition grew that the best way to learn about these effects was to study DNA directly. ♦

DOE Hits Sequencing Goal

JGI Strategies Pay Off for Chromosomes 5, 16, 19

On April 13, the U.S. Secretary of Energy announced that researchers at the DOE Joint Genome Institute (JGI) had determined the draft sequence for human chromosomes 5, 16, and 19. The three contain more than 300 million bases or about 10% of the total human genome, with an estimated 10,000 to 15,000 genes (see box for associated disorders).

"These three chapters in the reference book of human life are nearly complete," said Energy Secretary Richardson. "Scientists already can mine this treasure trove of information for the advances it may bring in our basic understanding of life and in such applications as diagnosing, treating, and eventually preventing disease."

JGI, now headed by Trevor Hawkins,* was established by DOE at Walnut Creek, California, in 1997. It is one of the largest publicly funded human genome sequencing centers in the world.

JGI Sequencing Strategies

A critical part of JGI's strategy was to sequence paired-end plasmids instead of M13 subclones used in most other HGP sequencing facilities. Because of the forward and reverse links between them, however, plasmids provided excellent order and orientation value when the fragments were assembled into large contiguous stretches (contigs). The result was "virtual" megabase-sized contigs whose lengths facilitate gene discovery. This is immensely helpful to gene hunters, who are finding that data on order and orientation are not available for many contigs in the current human genome maps.

Computational Analysis of Draft Data

The Oak Ridge National Laboratory's (ORNL) Computational Biology Section enriched the draft sequence by maximizing fragment order and orientation, assembling contiguous sequence stretches, and finding genes. An IBM SP3 supercomputer, one of the world's most powerful, provided the massive computing capability for analyzing millions of DNA base pairs. Standard data-analysis methods first identified such genomic features as sequence tagged sites (STSs), BAC end sequence tag connectors (STCs), and expressed sequence tags (ESTs). Data were refined further by programs for gene identification such as GRAIL-Exp that use both EST and complete cDNA data to add greater confidence in gene prediction. These analyses not only allowed for gene identification but also provided some fragment- or clone-ordering information.

The Java-based Genome Channel browser developed at ORNL provides a view of genomic sequences, computational and experimental annotation, and related links. The HTML-based Genome Catalog includes genomic summary reports, gene and protein lists, homologies, and other Internet capabilities (*http://genome.ornl.gov*).

Finishing the Draft to High Quality

Some limitations of rough draft data include project-to-project contamination, floating contigs (sequence reads that don't seem to belong anywhere), and false joins and other assembly errors. Finding useful biological information, even with accompanying cDNA sequences, is extremely difficult with gaps, incomplete order and orientation, incorrect assemblies, and base-pair errors. Prefinishing steps at Stanford Human Genome Center involve reassembling and analyzing the sequence, with the goal of fixing low-quality regions and filling in gaps. Finishing includes performing computational analysis of the assembly and resolving discrepancies. Final finished data are submitted to GenBank when clones are completely contiguous (see p. 5 for data Web sites).

Bug Month

During October, JGI launched its first "Microbial Month," turning out high-quality draft sequences at a rate of more than one every 1.5 working days. JGI sequence data is sent to ORNL's "annotation pipeline," where it is analyzed rapidly for genes and other important biological features. In addition to the basic research value of the 15 selected bacterial genomes, many have immediate implications for the economy and the environment (data at *www.jgi.doe.gov/tempweb/JGI_microbial/html*). The next two bug months are scheduled for March and August 2001.

Future Directions

Sequencing has begun on mouse genomic regions that are similar to gene-containing regions in human chromosomes 5, 16, and 19. The extensive 9× coverage of chromosome 19 has enabled the rapid generation of sequence-ready mouse maps that are providing clones for the sequencing pipeline. These maps also furnish reagents for basic studies of genome evolution and analysis of mouse mutations. Furthermore, a collaborative project is in the works to sequence 30 to 50 Mb of mouse genomic clones generated by ORNL-developed knockout mice (those with deleted or inactivated genomic regions).

In October, JGI announced a collaboration to sequence the genome of *Fugu rubripes* (pufferfish). Joining JGI are the Institute for Molecular and Cell Biology (Chris Tan), U.K. HGMP Resource Centre (Greg Elgar), Molecular Sciences Institute (Sydney Brenner), and Institute for Systems Biology (Leroy Hood). Because of its strong similarity to the human genome in number of genes and control sequences, the *Fugu* genome is considered a powerful, compact tool for identifying these regions in the much larger human genome. Scientists expect to sequence more than 95% of *Fugu* by March 2001 (*www.jgi.doe.gov*). ♦

Some Disorders Linked to Genes on Chromosomes 5, 16, and 19

Chromosome 5 (est. 194 Mb, ~6% of human genome): Colorectal cancer, basal cell carcinoma, acute myelogenous leukemia, salt-resistant hypertension, and a type of dwarfism

Chromosome 16 (est. 98 Mb, ~3% of human genome): Breast and prostate cancers, Crohn's disease, and adult polycystic kidney disease

Chromosome 19 (60 Mb, ~2% of human genome): DNA damage repair, atherosclerosis, diabetes mellitus, and myotonic dystrophy

*On November 3, DOE announced Trevor Hawkins' appointment as JGI director. JGI's first director Elbert Branscomb will assume leadership in developing the new OBER program, Bringing the Genome to Life.

High-Quality Sequence of Human Chromosomes 21, 22 Achieved

Two international research consortia marked major milestones in the Human Genome Project (HGP) with the completion of the first high-quality DNA sequences for two human chromosomes. Chromosomes 22 and 21 sequences, respectively, were reported in the December 2, 1999, and May 18, 2000, issues of *Nature*. These two chromosomes, smallest in the human genome, account for 2% to 3% of the total 3 billion DNA bases. [For an explanation of when a chromosome is considered "finished," see sidebar, p. 4.]

Chromosome 22

Chromosome 22's euchromatic (gene-containing) portion is estimated to be a 33.5-Mb structure comprising at least 545 and possibly up to 1000 genes ranging in size from 1000 to 583,000 bases. Genes are pinpointed by their sequence similarities to those already identified in other organisms and by complex computer modeling of potential ("putative") genes that may be only partially accurate. Chromosome 22's sequenced DNA is of extremely high quality with an error rate of less than 1 in 50,000 bases.

Gene variants on chromosome 22 have been implicated in immune system function and in at least 27 disorders, including congenital heart disease, schizophrenia, mental retardation, birth defects, and leukemia and other cancers. Scientists reported that at least eight regions are present in duplicate, leading to speculation about this phenomenon's evolutionary importance. Duplication can be studied closely when comparable animal genome sequences become available.

Chromosome 21

Chromosome 21 revealed a relatively low gene density, estimated at about 225 active genes in the 33.8 Mb of DNA covering 99.7% of the chromosome's long arm. Scientists speculate that this gene scarcity could contribute to the viability of individuals possessing a third copy of the chromosome, resulting in trisomy 21 (Down syndrome). The sequence also includes a contig of 28.5 Mb, the longest continuous DNA sequence reported thus far. The entire sequence has only 3 gaps totaling about 100,000 bases, compared with 10 gaps (totaling about 1 Mb) for chromosome 22's long arm.

Analysis of chromosome 21 genes may permit a deeper understanding of Down syndrome and its complications, as well as a range of such other linked genetic disorders as Alzheimer's disease and some forms of cancer. ♦

Chromosomes 21 and 22 Papers in *Nature Online*
See "Library of Original Research Papers" at
- ***www.nature.com/genomics***

Post-Sequencing Research Challenges

The working draft DNA sequence and the more polished version planned for 2003 or sooner represent an enormous achievement, akin in scientific importance, some say, to developing the periodic table of elements. And, as in most major scientific advances, much work remains to realize the full potential of the accomplishment.

Early explorations into the human genome, now joined by projects on the genomes of dozens of other organisms, are generating data whose volume and complex analyses are unprecedented in biology. Genomic-scale technologies will be needed to study and compare entire genomes, sets of expressed RNAs or proteins, gene families from a large number of species, variation among individuals, and the classes of gene regulatory elements.

Deriving meaningful knowledge from DNA sequence will define biological research through the coming decades and require the expertise and creativity of teams of biologists, chemists, engineers, and computational scientists, among others. A sampling follows of some research challenges in genetics—what we still won't know, even with the full human sequence in hand.

- Gene number, exact locations, and functions
- Gene regulation
- DNA sequence organization
- Chromosomal structure and organization
- Noncoding DNA types, amount, distribution, information content, and functions
- Coordination of gene expression, protein synthesis, and post-translational events
- Interaction of proteins in complex molecular machines
- Predicted vs experimentally determined gene function
- Evolutionary conservation among organisms
- Protein conservation (structure and function)
- Proteomes (total protein content and function) in organisms
- Correlation of SNPs (single-base DNA variations among individuals) with health and disease
- Disease-susceptibility prediction based on gene sequence variation
- Genes involved in complex traits and multigene diseases
- Complex systems biology including microbial consortia useful for environmental restoration
- Developmental genetics, genomics ♦

Online Bioinformatics Newsletters

BioInformer (EMBL European Bioinformatics Institute): Quarterly. Bioinformatics research, developments, and services (*http://bioinformer.ebi.ac.uk*)

NCBI News (National Center for Biotechnology Information): Quarterly. Research activities, new databases, and software services (*www.ncbi.nlm.nih.gov/About/newsletter.html*)

What's New (DNA Data Bank of Japan): Updated as needed. News, upgrades, and release information (*www.ddbj.nig.ac.jp/whatsnew-e.html*) ♦

In the News

DOE and NIH Teams to Unlock Power of Proteins

Seven new grants, four of them awarded to scientists at DOE sites, are key components in the Structural Genome Initiative started by the NIH National Institute of General Medical Sciences (NIGMS). Over the next decade, the new study will determine the form and function of thousands of proteins.

"These awards demonstrate the continued importance of the physical sciences to life-science research and the strong role the national laboratories play in providing expertise and world-class facilities in our quest to understand the structure and function of genes," noted Dr. Mildred Dresselhaus, Director of the DOE Office of Science.

Proteins come in many sizes and shapes, and their functions often depend on tiny structural details. Obtaining the 3-D structure may help scientists understand how each protein functions normally and how faulty structures can cause or contribute to disease. "We expect this effort to yield major biological findings that will improve our understanding of health and disease," said NIGMS Director Marvin Cassman in announcing the grants. These data also can help in designing drugs that bind to the proteins and affect their activity.

The grants total around $4 million each for the first year. NIGMS plans to spend about $150 million on the seven grants over the next 5 years. The four DOE-involved projects are listed first below. Investigators at DOE national laboratories also are involved in some of the other projects.

Grant Recipients, Team Leaders, Specific Goals

- Structural Genomic Center (Sung-Hou Kim, Lawrence Berkeley National Laboratory): Speed up structure determination by X-ray crystallography; study proteins essential for independent life by focusing on two extremely small, closely related bacteria (*Mycoplasma genitalium* and *M. pneumoniae*) [*www.lbl.gov/Science-Articles/Archive/nigms-grant.html*].
- Tuberculosis Structural Genomics Consortium of 13 institutions in 6 countries (Tom Terwilliger, Los Alamos National Laboratory): Determine and analyze structures of about 400 proteins from *Mycobacterium tuberculosis* to facilitate new and improved drugs and vaccines for tuberculosis [*www.lanl.gov/worldview/news/releases/archive/00-127.html*].
- Midwest Center for Structural Genomics consortium of seven institutions (Andrzej Joachimiak, Argonne National Laboratory): Reduce the average cost of determining a protein structure from $100,000 to $20,000; select protein targets from all three kingdoms of life, with emphasis on previously unknown folds and on proteins from disease-causing organisms.
- New York Structural Genomics Research Consortium of five institutions (Stephen K. Burley, Rockefeller University): Develop techniques to streamline structural genomics and solve several hundred human and model-organism protein structures.
- Joint Center for Structural Genomics (Ian Wilson, Scripps Research Institute): Develop high-throughput methods for protein production, crystallization, and structure determination by initially focusing on novel structures from *Caenorhabditis elegans* and human proteins thought to be involved in cell signaling; determine structures of similar proteins from other organisms to include the greatest number of different protein folds [*www.stanford.edu/dept/news/report/news/september27/ssrl-927.html*].
- Northeast Structural Genomics Consortium (Gaetano Montelione, Rutgers University): Target proteins from various model organisms including the fruit fly, yeast, and roundworm and related human proteins; use both X-ray crystallography and nuclear magnetic resonance spectroscopy to determine protein structures.
- Southeast Collaboratory for Structural Genomics (Bi-Cheng Wang, University of Georgia): Analyze part of human genome and all of two representative organisms, *C. elegans* and *Pyrococcus furiosus*; emphasize technology development, especially for automated crystallography and nuclear magnetic resonance imaging techniques. ♦

Award information

- ***www.nigms.nih.gov/news/releases/sgpilots.html***

High-Resolution Image Reveals Structure of Protein Machine

Using a high-energy X-ray beam from the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory, researchers at Yale University and the Howard Hughes Medical Institute obtained the most detailed images ever seen of the ribosome (the protein-making structure inside all living cells). NSLS is a DOE Office of Biological and Environmental Research structural biology user facility.

In prokaryotes (bacteria and other simple organisms) as well as the more complex eukaryotes, ribosomes help translate gene-encoded information into a specific protein. Ribosomes consist of two unequally sized subunits containing RNA and proteins. The smaller component binds the messenger RNA (mRNA), which contains genetic instructions that specify the amino acids required to build a particular protein. The larger ribosomal subunit attaches one amino acid to the next in the growing protein chain.

In the August 11 issue of *Science*, investigators reported visualizing the atomic structure of the bacterium *Haloarcula marismortui*'s larger ribosomal subunit at an unprecedented resolution of 2.4 Å. Until this report was published, researchers did not know whether ribosomal

Proteins, p. 19 ➠

In the News

Gene Patenting Update: U.S. PTO Tightens Requirements

Worries Continue over "Patent Stacking" and Early, Broad Patents

Massive amounts of data flowing from the Human Genome Project and other genomics projects have stimulated an avalanche of applications to the U.S. Patent and Trademark Office (PTO) for patents on genes and gene fragments. Some 3 million ESTs (fragments that identify pieces of genes) and thousands of other partial and whole genes are included within pending patents. This situation has sparked controversy among scientists, many of whom have urged the PTO not to grant broad patents at this early stage to applicants who have neither characterized the genes nor determined their functions and specific uses.

Genes and other biological resources have been patentable since the landmark 1980 U.S. Supreme Court decision in *Diamond v Chakrabarty* that granted a patent for an oil-dissolving microbe. Patents give owners exclusive rights to their inventions or ideas for 20 years from the filing date. The rationale is to allow inventors time to recoup their investment costs in exchange for a public description of their knowledge, thereby revealing technical advances to competitors and the general public and avoiding duplicated efforts. Biological inventions are patentable if they meet the standard requirements for all patents: they must be novel, useful, not obvious, and described sufficiently for others to reproduce.

A single gene may be patented, in principle, by different scientists or companies. One concern is that such "patent stacking" may discourage product development because royalties are owed to all patent owners. Additionally, because applications remain secret, companies may work on developing a product, only to find that "submarine patents" already have been granted, leading to unexpected licensing costs and possible infringement penalties.

Some past controversies have centered around the "utility" requirement. Some fear the large-scale patenting of gene fragments by biotechnology companies who are unaware of their functions but would stake a claim to all future discoveries on those genes (sometimes called "reach-through patents").

In December 1999, the PTO published revised interim guidelines clarifying the utility requirement for patent claims on genomic and other biotechnological inventions. The new rules call for "specific and substantial utility that is credible," but some still feel the rules are not stringent enough. Public comments have been posted to the PTO Web site (*www.uspto.gov*; scroll to "Notices of Public Comments").

In the comments, the National Advisory Council for Human Genome Research observes that "a broad allowance of claims is unjustified and will strongly discourage the further research efforts necessary to translate gene discovery into medically important therapies."

More patenting information:

- ***www.ornl.gov/hgmis/elsi/patents.html***

Instead of patent protection for specific gene sequences, Bruce Alberts, President of the National Academy of Sciences, advocates patents for "the new treatments and drugs that will result from the research and development efforts of many different individuals and companies working from the basic information in the human genome sequence."

Final revised guidelines are expected from the PTO (see also box below). *[Denise Casey, HGMIS]* ♦

Witnesses Testify About Patenting Genes

On July 13, witnesses presented testimony during a House of Representatives hearing on "Gene Patents and Other Genomic Inventions" held by the Committee on the Judiciary, Subcommittee on Courts and Intellectual Property. A complete transcript is on the Web (*www.house.gov/judiciary/4.htm*).

At the hearing, Harold Varmus (Memorial Sloan-Kettering Cancer Center) stated that some of the issued patents appear to reward the obvious in DNA sequencing and diminish the innovative work required to determine gene function and utility. This new environment, he said, has led many academic institutions to establish expensive offices to protect intellectual property and regulate the exchange of biological materials that once would have been shared freely. The use of new scientific findings has been hampered, and the open exchange of ideas and materials has been inhibited, he continued.

Dennis Hopper (Genentech, Inc.) testified that his company invests about $400 million a year in the research and development of therapeutic products, focusing on identifying human proteins. He said patent protection and market exclusivity are very important considerations in making such investments.

Jon Merz (University of Pennsylvania, Philadelphia) expressed concern about exclusive licensing of disease-gene patents that claim a gene sequence and one or more mutations leading to disease. In addition to covering all uses of the chemical sequences, patents claim all methods of diagnosing disease in a specific patient through the identification of the disclosed genetic alleles, mutations, or polymorphisms. Merz stated that some licensees thus are exercising their patent rights to prevent physicians—in particular, molecular pathologists—from performing genetic testing on their patients.

Merz pointed out that most disease genes are found, at least in part, through federally funded research. Exclusive licensing is contrary to the longstanding policy that the public should not have to pay twice. Merz recommends reserving exclusive licensing for inventions that require substantial downstream investment. Other witnesses said that patents should be even more available to encourage the development of critically needed medical advances.

PTO Director Q. Todd Dickenson stated that both points of view are relevant and that his office is responsible for balancing them. To that end, he said, the PTO is finalizing guidelines to require the demonstration of "real-world" utility for gene-related patents, rather than just theoretical uses. ♦

SNP Consortium Collaborates with HGP, Publishes First Progress Reports

The Human Genome Project (HGP) and The SNP Consortium (TSC, *http://snp.cshl.org*) announced plans to generate a new set of human DNA sequence data that will contribute 125,000 to 250,000 validated and useful DNA markers known as SNPs. The DNA to be sequenced will come from 24 anonymous, unrelated donors with diverse geographic origins, and all data will be made publicly available. Researchers expect to complete the project by December.

A high-density map of SNPs (single base-pair variations that occur about once every 100 to 300 bp throughout human DNA) is expected to be a valuable research tool. It will help scientists pinpoint genetic differences that predispose some people to disease and underlie variable individual responses to treatment.

Three genome research centers are participating in the HGP-TSC collaboration: Whitehead Institute, Washington University School of Medicine (St. Louis), and the Sanger Centre (United Kingdom). These centers will isolate at least 2.5 million DNA fragments (each about 6000 bp long) from the human genome and determine the sequence of about 500 bp at both ends of the fragments, resulting in paired-end sequences a known distance from each other. The sequences then will be compared to those already in GenBank. The paired-end data will help span some gaps in the human genome working draft, thus making the draft more accurate.

"The collaboration between the HGP and TSC demonstrates that public-private cooperation can be an efficient means for developing basic research tools essential for the application of genetic information to the understanding and treatment of diseases," noted Arthur Holden, chairman and chief executive officer of TSC.

The nonprofit TSC foundation was established by the Wellcome Trust and a group of pharmaceutical and technological companies with the initial goal of identifying and locating up to 300,000 SNPs by the end of 2001. An exponential increase in the amount of HGP data, however, has since enabled TSC to proceed at a much faster pace. By September, it had identifed more than 350,000 SNPs and mapped almost 250,000 to the working draft sequence. With the HGP collaboration, the total number of useful SNPs mapped may exceed 750,000 by December.

First TSC DNA Variation Map

The first publications on the methodology and progress of TSC appeared

International SNP Meetings

Some 100 invited international researchers from industry and academia attended the Second International Meeting on Single Nucleotide Polymorphism (SNP) and Complex Genome Analysis held in Munich, Germany, in September 1999. Below is a summary of meeting highlights with updates from the September 2000 meeting in Taos, New Mexico.

The overall tone of the 1999 meeting showed that SNP research still lacks a solid consensus about how best to proceed—and this is at a time when vast sums of money are being spent on public and private SNP programs. This state of affairs is only partially improved today. SNP discovery alone will not determine how and whether SNPs should be used. Instead, this knowledge has to come from empirically determined guidelines and studies based upon best guesses and theory using the latest technologies.

SNP Discovery

The SNP Consortium (TSC) has established a well-structured program toward discovery and public cataloging of many hundreds of thousands of random genomic SNPs (at a 95% accuracy target) in 2 years, plus detailed mapping of over 170,000 of these. Original goals are being greatly surpassed. Other projects involve automated alignment and comparison of expressed sequence tag (EST) sequences for SNP discovery. Although limited by the depth and number of distinct EST contigs to perhaps a few tens of thousands of SNPs, the exercise is highly cost-effective. Furthermore, the intragenic location of these variants could make them individually more useful than TSC-discovered markers.

SNP Scoring and Detection

A plethora of competing SNP genotyping methods is being developed, but useful approaches must meet stringent requirements of both high throughput and accuracy. A fully ideal method still does not exist today. Extremely large sample materials may be required to achieve sufficient statistical power in association studies, and a genotyping error rate of as little as 1% could have a disastrous effect on statistical power. Claimed genotyping costs still range from a fraction of one dollar to many dollars per sample.

Two principal approaches to SNP scoring are in individual reactions and in a multiplexed fashion, usually achieved by using immobilized oligonucleotides on microarrays. The requirement for PCR before the detection reaction remains a major bottleneck. Simultaneous analysis of pooled DNA samples may be useful for increasing throughput and decreasing the cost of SNP scoring in association studies.

Diseases and Phenotypes

SNPs and association analysis are being used to (1) home in on disease-related mutations within large regions previously identified by linkage scans, (2) screen several variations surrounding a few or many prechosen candidate genes, or (3) follow extensive sequence studies of single candidate genes to determine associations between specific haplotypes and disease. These strategies can sometimes work—although how best to maximize the rate of success is not known.

If there is no initial linkage to guide the search, however, the current answer seems to be to make educated guesses about which genes are likely to be important. Presenters summarized attempts to do this for Alzheimer's disease, schizophrenia, dyslexia, and substance dependence. These studies began with a strong prior belief in the relevance of the tested candidate gene. Other presented

in the September 28 issue of *Nature*. Eric Lander and his team at the Whitehead Institute Center for Genome Research reported on a new method called reduced representation shotgun sequence (RSS) that increases the accuracy of SNP mapping. RSS scans subsets of markers from several people and compares the resulting sequences to identify DNA variations.

In a second *Nature* paper, a team of TSC researchers from the Sanger Centre presented a human chromosome 22 map featuring 2730 SNPs identifed by RSS and aligned to the human genomic sequence. Most of the SNPs are within 25 kb of a transcribed exon (protein-coding region), making them useful for association studies. Chromosome 22, the second smallest of the chromosomes, is linked to more than 35 diseases and syndromes including some cancers, schizophrenia, and heart disease.

More information on SNPs:

- ***www.ornl.gov/hgmis/faq/snps.html***

The goal of TSC is to build a map containing one SNP every 5 kb that is integrated with the human genome sequence and freely available to all researchers (600,000 SNPs evenly spaced throughout the human genome's 3 billion bases). In September, scientists reported the total number of SNPs now in the public domain to be more than 1.2 million, scattered across the genome.

The ultimate aim of SNP studies is to develop customized therapies to treat or prevent disease. In the September 13 issue of *Proceedings of the National Academy of Sciences*, researchers moved a step closer to that goal as they reported for the first time the ability to predict an individual's response to a drug based on that person's particular group of SNP markers. The study involved asthmatic volunteers and responses to albuterol, a drug commonly used to achieve rapid improvement in lung function.

As more clinical studies get under way and advances eventually become established in clinical practice, the urgency increases to find effective ways to protect the genetic privacy of individuals. In response to these projected needs, several members of TSC formed a new company, First Genetic Trust, Inc., to act as an independent intermediary between genetic information providers (individuals) and users (researchers and healthcare providers). The company aims to protect the rights of individuals in regard to the confidentiality and access of their genetic data. The first priority of the company will be to address the needs of the pharmaceutical industry as it conducts clinical trials. ♦

---

data showed how intricate estimations of haplotype configurations, regression, and cladistic analyses can lead toward the precise intragenic location of the pathogenic allele and genotype combinations. This evolutionary perspective was a key take-home message.

The emphasis of most SNP research on tools and genotyping (finding the link between marker and pathogenic allele) was contrasted with the relative lack of attention to careful study design and sample ascertainment (finding the link between pathogenic allele and disease). Many case-control association studies may be futile because the lack of even rare families segregating the disease could indicate too-high genetic complexity.

Population Genetics

Speakers presented work on characterizing extant haplotypes and developing maps describing typical distances up to which allelic variants can be expected to be in linkage disequilibrium. These efforts can benefit disease-gene and population-genetics studies. Identifying potential targets of selective fixation via SNP analyses may be useful in revealing footprints of adaptive evolution. Revealing such dominant beneficial alleles could provide important targets for study.

Databases and Bioinformatics

Future high-throughput detection will require efficient systems for collecting and integrating voluminous amounts of data in high-quality databases. Representatives from the European Molecular Biology Laboratory–European Biotechnology Institute presented software and database solutions to linking various locus-specific mutation databases with SNP databases allowing complex queries.

Many groups endeavor to mine SNPs from EST data, but measuring allele frequencies in silico is difficult, and many rare alleles might be among them. Prediction accuracy is estimated at only 60% to 80% in most cases, although this identifies gene candidates for further investigation. Others are attempting to map large numbers of SNPs onto human chromosomes and three-dimensional protein structures to understand phenotypic differences and human evolution using SNP data.

Intellectual Property, Commerce

A clear trend is toward granting patents on partial nucleic acid sequences or SNPs only when functions and commercial applications can be defined. Commercially, profits are expected to be generated in three areas: applying SNPs to pharmacogenomics by discovering functional implications, genotyping individuals for particular SNPs, and creating technology platforms for SNP discovery and use.

In the first situation, proprietary rights to the "important" SNPs are expected to generate profits via licensing. In the third instance, profits can be generated over shorter times by sales and licensing of patented technologies. But the greatest business success from SNP knowledge may be realized only if and when solid correlations between SNPs and gene function are determined.

A key question is whether the most obvious and rewarding SNPs (from a cost-benefit standpoint) already have been discovered and patented or if latecomers still have a good chance of finding valuable SNPs. *[Reported by Anthony Brookes; Center for Genomics Research; Karolinska Institute; Stockholm, Sweden (*anthony.brookes@cgr.ki.se*)]* ♦

Detailed 1999 Meeting Report

www.ornl.gov/meetings/brookes.html; Eur. J. Hum. Genet. **8**(2), 154–56 (2000)

September 2000 Meeting

www.cgr.ki.se/cgr/groups/brookes/snp2000/abstracts.htm

In the News

Public, Private Sectors Join in Mouse Consortium

Sequencing Results will Spur Discovery of Human Genes and Their Functions

In October, a collaboration was announced to speed up sequencing of the mouse genome and produce a draft map by spring 2001. The Mouse Sequencing Consortium (MSC) consists of six NIH institutes, the Wellcome Trust philanthropy, and three private companies. It provides another example of public and private sectors joining forces to support large-scale genomics research and generate freely available data crucial for basic biomedical research (see related articles, pp. 1–3).

MSC members and their contributions are SmithKline Beecham ($6.5 million), the Merck Genome Research Institute ($6.5 million), Affymetrix, Inc. ($3.5 million), Wellcome Trust ($7.75 million), and NIH ($34 million). Total funding of $58 million will support sequencing for 6 months at three centers: Whitehead Institute (Cambridge), Washington University (St. Louis), and the Sanger Centre in the United Kingdom. ➠

BERAC Report Endorses New Program

Bringing the Genome to Life: Energy-Related Biology in the Post-Genomic World, issued in June, recommends a new research program for the DOE Office of Biological and Environmental Research (BER), headed by Ari Patrinos. At the request of the DOE Office of Science, the report was produced by the BER Advisory Committee's (BERAC) Genome Subcommittee, chaired by Raymond Gesteland (University of Utah). It suggests that the new program's major challenge is to understand and predict the responses of single- and multicellular organisms to biological and environmental cues. Elbert Branscomb, first director of DOE's Joint Genome Institute, is in charge of developing the program. ♦

BERAC Report:
www.er.doe.gov/production/ober/berac/genome-to-life-rpt.html

¶ Microbial Genome Program Report

A color booklet on the DOE Microbial Genome Program (MGP), published in February 2000, is available in print and electronic formats. It includes program origins, targeted microbes and their potential uses, graphics, and abstracts of MGP research. The MGP, which DOE initiated in 1994 as a spinoff of its Human Genome Program, has been extremely successful. As of November, MGP researchers had completed the sequencing of 14 microbes, with around 35 more in progress. Future plans call for more detailed microbial analysis. At least 5 other governmental agencies have begun their own microbial programs, and about 170 microbial genomes are being sequenced and studied at present.

- **Print copies:** 865/574-7582,Fax: /574-9888, *yustln@ornl.gov*
- **Web**: *www.ornl.gov/hgmis/publicat/microbial*
- **PDF (printer-friendly) file**: *www.ornl.gov/hgmis/publicat/microbial/mgp.pdf* ♦

¶ *Intimate Strangers: Unseen Life on Earth*

The four-part documentary series *Intimate Strangers: Unseen Life on Earth*, which was broadcast on public television last November with partial support from DOE, is available on videotapes (800/532-7637 or 800/423-1212). A 225-page companion book of the same name can be purchased at 800/546-2416. The MicrobeWorld Web site, tailored to middle and high school students as well as professional and lay audiences, includes downloadable hands-on educational activities correlated with the television series, information and stories about microbes, and teacher resources (*www.microbeworld.org*). ♦

¶ Microbial Genomics

Interagency Report on the Federal Investment in Microbial Genomics discusses each relevant federal agency's activities, plans, and areas of interest in microbial genomics and summarizes opportunities for and limitations to microbial research. Prepared for the Executive Office of the President of the United States in 2000. [Print copies: HGMIS, 865/574-7582, *yustln@ornl.gov, http://whitehouse.gov/WH/EOP/OSTP/html/microbial/nsf00203_1.html*] ♦

¶ Federal Technology Funding Guide

The fifth edition of the 2001 *Federal Technology Funding Guide* may be downloaded free in PDF (printer-friendly) format from *www.larta.org/ecommerce/FTFG2001.htm*. Created by the Los Angeles Regional Technology Alliance (larta), the 152-page guide profiles nearly 100 regularly scheduled federal programs that support technology development and deployment. More than half of all U.S. research and development (over $77 billion) is funded by the federal government. [Larta contact: 213/743-4150, *www.larta.org*] ♦

¶ Genetics, Public Health

Genetics and Public Health in the 21st Century: Using Genetic Information to Improve Health and Prevent Disease explores the genetic revolution's impact on public health practices and delineates a framework for the integration of related advances and technologies. Editors are Muin Khoury (Centers for Disease Control and Prevention), Wylie Burke (University of Washington, Seattle), and Elizabeth Thomson (NIH National Human Genome Research Institute). Some 75 contributors represent a wide range of disciplines. Portions of the book are online at *www.cdc.gov/genetics/publications/21stcentury.htm* [639 pp., Oxford University Press, 2000]. ♦

¶ Genetic Testing

Enhancing the Oversight of Genetic Tests, by the Secretary's Advisory Committee on Genetic Testing, assesses current oversight and offers recommendations on ensuring public access to quality genetic tests (*http://www4.od.nih.gov/oba/sacgtfinal.pdf*). ♦

Why the Mouse?

With the working draft sequence of the human genome in hand, scientists in industry and academia now seek to interpret its meaning. The mouse genomic sequence is a powerful comparative tool because genes in the two organisms are very similar. Understanding gene function in the mouse will accelerate knowledge about comparable human genes and will aid in understanding human disease and in developing new treatments.

On average, protein-coding regions in the mouse and human genomes are 85% identical. These regions are evolutionarily conserved because they are required for biological functions shared by both organisms. In contrast, noncoding genomic regions are less than 50% identical. When comparing the same DNA regions from human and mouse, therefore, functional elements stand out clearly because of greater similarity.

"Data from this project will be invaluable to us in annotating the final draft of the human genome," observed researcher John McPherson (Washington University). "It is exciting that we are moving rapidly toward completion of both projects," he said.

More information on the MSC

- ***www.nhgri.nih.gov/news/mousegenes/mouse_release.html***

Trans-NIH Mouse Initiative

- ***www.nih.gov/science/models/mouse/index.html***

Rapid Data Release

The consortium project focuses on the "black six" (*C57Black/6*) mouse strain, which is different from the three strains being sequenced by Celera Genomics. Celera's data are available by paid subscription.

The MSC data-release policy calls for raw data (individual DNA sequence traces, about 500 bases long) taken directly from automated instruments to be deposited in two public databases. These are operated by the National Center for Biotechnology Information (*www.ncbi.nlm.nih.gov*) and the European Bioinformatics Institute (*www.ebi.ac.uk*). Individual sequences will be assembled into larger units as soon as a working draft is obtained.

Sequencing Strategy

MSC sequencing melds the best features of two strategies used to produce a working draft of the human genome: the map-based shotgun method used by the public Human Genome Project consortium and the whole-genome shotgun system used by Celera. The overall depth of coverage for the mouse genome will be 2.5× to 3×, a level of detail useful to researchers; a finished, highly accurate sequence is expected at a later date. ♦

Fast Forward to 2020: What to Expect in Molecular Medicine

This article was originally written for and will appear in the online magazine *TNTY Futures* (*www.tnty.com/newsletter).* Free subscriptions are available through the Web site. In the article, the authors speculate about possible future changes in medical practice resulting from genome research.

The first phase of the ambitious international effort to determine the entire sequence of the human chromosome set is virtually complete. Human Genome Project scientists plan to finish the human sequence by 2003, along with a database of the most common sequence variations that distinguish one person from another. This knowledge base, freely available to any interested person over the Internet, will revolutionize biology and medicine. But how? What will be different 20 years from now because the human genome was sequenced?

Only time will prove the accuracy of the following predictions, but here is a list of some effects we might expect in 2020.

More Effective Pharmaceuticals

A virtually complete list of human gene products will give us a vast repertoire of potential new drugs. From 500 or so drugs in 2000, at least six times this number will have been identified, tested, and commercialized in 2020. Most will be manufactured by recombinant DNA technology so they will be "reagent-grade pure," just as human insulin and growth hormone are today.

Your medical record will include your complete genome as well as a catalogue of single base-pair variations that can be used to accurately predict your responses to certain drugs and environmental substances. This will permit you to be treated as a biochemical and genetic individual, thus making medical interventions more specific, precise, and successful. In addition, the increased power of medicine to predict susceptibility to specific diseases will allow you to alter your lifestyle to reduce the likelihood of developing such diseases or to be treated with preventive or disease-delaying medicine.

Treatment failures occasionally happen today with drugs for hepatitis C infections, antihypertensives, and certain antidepressants (selective serotonin reuptake inhibitors like Prozac). In the next 15 to 20 years, more effective drugs will be developed, and doctors will test individual genetic profiles against panels of drugs available for a specific condition and choose the treatment with the greatest potential benefit for each patient.

Today, some 100,000 people die each year from adverse reactions to drugs, and millions of others must bear uncomfortable or even dangerous side effects. We see such current examples as heart-valve abnormalities from diet drugs, muscle damage from some hormone-regulating drugs, and nervous system effects with certain types of antidepressant

p. 14 ➠

Forward to 2020 (from p. 13)

medications. As genes and other DNA sequences that influence drug response are identified, we can expect the number of toxic responses to drop dramatically and most side effects to be eliminated.

Societal Implications

Another consequence of greater knowledge about individual variation is more disturbing, and we may face some unpleasant consequences unless society makes some hard choices. These considerations include the likelihood that your medical information will be available to others not in the medical profession—your insurer or employer, perhaps. Employers may have a strong motive to learn about your risks of developing certain conditions so they can avoid hiring you or restrict the kinds of work you may do.

Genetic Testing, Therapy

Although now plagued by technical difficulties, gene therapy for single-gene diseases will be routine and successful in 20 years. Certain aberrant disease-associated genes will be replaced with normally functioning versions, and several hundred diseases will be curable. Neonatal genetic testing for these treatable conditions will be routine.

Some of the mysteries of early embryonic development will be solved. We should know the timing of expression of most, perhaps all, of the human gene set. We may have learned how to direct differentiation so that a desired cell type or even relatively "simple" organs and parts of more complex organs can be grown for transplantation. In 2020, we will have made substantial progress towards true "cloning" of certain organs, but many difficult technical steps will remain before successful cloning of a heart or liver.

As genetic testing using DNA sequence becomes less expensive and more accurate, it will be used commonly and reliably in cases of mistaken identity, false or misattributed paternity, and the identification of missing persons. Misguided attempts to ascribe behavioral tendencies to a person's genes will cause many problems, especially for the courts that must resolve disputes when an individual's behavior and actions conflict with laws. Should society (via the courts) interpret behavior as a consequence of free will or as influenced by genetic constitution? At what point does society mitigate responsibility or punishment?

Understanding Life

On the brighter side, an inevitable consequence of the genome project will be a much greater understanding of fundamental biology. Already, more than three dozen organisms (mostly one-celled microbes) have been completely sequenced. The fruit fly, the latest organism to be sequenced, is being used to model the essential features of human disorders such as Parkinson's, making possible a powerful genetic approach to garnering knowledge about diseases as well as to developing more effective treatments. In 2020, perhaps 1000 complete genomes will be in hand. Besides furnishing insights into evolution, this vast repertoire of new genes and their products can be explored for their potential in solving challenging problems such as environmental cleanup.

We will fitfully and slowly gain some insights into biological complexity. In 2020, we will know how to build a functioning cell capable of free-living existence. We will understand certain pathways used by this simplest cell, but there still will be unanswered questions about it. We will be virtually no closer than we are today to the mysteries of such true "emergent" properties as intelligence in complex multicellular organisms.

Challenges

We speculate that the Human Genome Project will have vast and largely positive impacts on people living in 2020. Of the various predictions noted above, the last two are the most profound because the most powerful and momentous impacts come from fundamental knowledge, usually in unforeseen ways. As this astonishing treasure trove is introduced into society, we need to be alert to challenges and misuses of the knowledge about ourselves. Society as a whole, not just genome scientists, must address these considerations. It has to be all of us. *[Daniel Drell (DOE) and Anne Adamson (HGMIS)]* ♦

More information on Medicine and the New Genetics

- ***www.ornl.gov/hgmis/medicine/medicine.html***

Judging Molecular Biology of Murder, Addictive Disorders, and Dementia

Sixty learners stare at cross-section images of living brains projected on the screen at Airlie Conference Center, an hour out of Washington in Northern Virginia's rolling horse country. Drawn to a specialized 3-day training program on neurobehavioral genetics, they strain to find red and white spots indicating mental activity. As they search intently for dark stretches of turned-off brain segments, they might be mistaken for participants in a medical school's clinical seminar. But these are judges absorbed in the speaker's rapid, precise language as she points out image slices of brains from murderers, cocaine addicts, demented elderly persons, and gunshot victims.

Today's judges are beginning to view defendants and plaintiffs in a new light while presiding over some 15 million cases a year in the nation's state and federal courts, where they

p. 15 ➠

¶ Judges' Guidebook

The Biology of Conduct: Judges' Guidebook to Neuro- and Behavioral Genetics serves as an up-to-date archive, a case-related teaching tool for the next advanced conference, and a means for forging a community of discourse among the nation's adjudicators (*www.einshac.org*). ♦

are accustomed to observing behavior and its results. At the conference, they peer into genotypes and phenotypes linked to behavior they will find on their dockets upon their return home. For most, it is an opportunity to assess scientific progress in discovering causes of violence, addictions, and dementia. They are able to compare clues from neuroscience with those from genetic and behavioral development.

Among 22 science advisors at Airlie, the judges' guide is Nora Volkow, Brookhaven National Laboratory's brain-mapping expert. Sacrificing a holiday weekend in aid of the courts, she has joined her colleagues—basic and clinical scientists, psychiatrists, and psychologists—to paint a landscape of molecular and organic knowledge. In the past 3 years, more than 250 neutral and independent scientists have served as advisors for Genetics in the Courtroom conferences sponsored nationwide by the Ethical, Legal, and Social Implications component of DOE's Human Genome Program. These courses were conceived and are carried out by Franklin Zweig and his colleagues at the Einstein Institute for Science, Health, and the Courts (EINSHAC).

Each scientific expert has participated in one or more workshops to prepare judges for the testimony of nonneutral experts in cases involving genetics. To date, 1900 jurists have participated, and another thousand are expected next year. Subject matter ranges from basic molecular biology to advanced biotechnology to briefings about policy.

At this meeting on behavioral genetics, an advanced offering, the jurists' attention ratchets up as Volkow and others project photos of brains plagued by chronic alcoholism. All the justices have attended the basic course, so they have a grasp of DNA structure and function, the nature of genes, and the fundamental dynamics of protein production. They strive for something more—the link between genetic predisposition or susceptibility and the trouble-causing conduct they see every day. They monitor current scientific developments, hoping for the day when alcoholism's root causes are known and judges can direct defendants to effective treatments. At Airlie, the judges are exposed to the connections (and the disconnects) between underlying causation and behavior. The disconnects dominate in alcoholism's case, but jurists agree that breakthroughs are so rapid that the next one might be valid and could be introduced in one of their cases.

The judiciary appears to share a general belief that biological inheritance fuels generations of cases involving violence, alcohol, drugs, and family breakdown. They are urged by lawyers to use their authority to protect the mentally ill, and they intuit that the schizophrenia and bipolar illnesses frequently appearing in their courtrooms are, at least in part, forged by genetics gone wrong.

On the conference's final day, the discussion turns from science to its implications for judicial management. The judges are looking to the time when molecular medicine will make possible the substitution of effective treatment for long jail sentences. One speaker warns against blaming biology for behavior, but the judges are more concerned about having insufficient tools against a tidal wave of challenges. They also worry that their

p. 16 ➠

Genetics in the Courtroom

- ***http://www.ornl.gov/hgmis/courts.html***

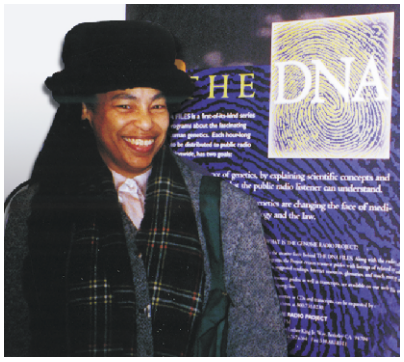
¶ *Genes and Justice*

The *Genes and Justice* symposium issue of *Judicature* **83**(3), the journal of the American Judicature Society, focuses on the growing societal impact of DNA technology and related issues that courts will confront in the near future (*www.ornl.gov/hgmis/publicat/judicature*). The DOE Human Genome Program's Ethical, Legal, and Social Issues component contributed to the publication of this issue and its distribution to judges nationwide. Denise Casey (HGMIS) served as editor. ♦

DOE Grantee Scott Wins Award

SoundVisions Productions, founded and directed by Barinetta Scott and Jude Thilman, won a Silver Baton award for *The DNA Files* in the 2000 Alfred I. Du Pont–Columbia University television and radio competition. Hosted by John Hockenberry, the series consisted of nine 1-hour programs broadcast on National Public Radio. It was funded in part by the Ethical, Legal, and Social Implications component of DOE's Human Genome Program.

Twelve winners were selected from 650 submissions aired between July 1, 1998, and June 30, 1999. Bill Moyers won a Gold Baton for *Facing the Truth*, a 2-hour documentary about the aftermath of apartheid in South Africa. Other Silver Baton winners included Diane Sawyer (*20/20: The Unwanted Children of Russia*); Youth Radio, Berkeley (*E-Mails from Kosovo*); and CBS News and Bob Simon (*60 Minutes II: The Shame of Srebrenica*). The Du Pont–Columbia awards, among the most prestigious in broadcast journalism, began in 1943.



Bari Scott
SoundVisions Productions

For this school year, Scott is a Knight Science Journalism fellow at the Massachusetts Institute of Technology.

[Orders for *The DNA Files* tapes and transcripts: *www.dnafiles.org/about/tapes.html* or 510/486-1185.] ♦

Airlie advisors are not typical of an expert-witness industry bent on made-for-litigation research, experts who have a biased stake in case outcomes.

Conference cochairmen summarize sentiments near the meeting's close. "We need at least a coherent state of the science merged with new developments in the law," Chief Judge Eugene N. Hamilton declares. Judge John Garrett Penn rejoins, "We need a legal guide to the scientific landscape facing federal judges." The meeting ends, but its work continues. *[Franklin M. Zweig, Einstein Institute for Science, Health, and the Courts, www.einshac.org]* ♦

Human Genome news

This newsletter is intended to facilitate communication and collaboration, help prevent duplication of research effort, and inform persons interested in genome research. Views expressed are not necessarily those of the Department of Energy Office of Biological and Environmental Research. Suggestions are invited.

Human Genome Management Information System (HGMIS)
Oak Ridge National Laboratory
1060 Commerce Park, MS 6480
Oak Ridge,TN 37830
865/576-6669, Fax: /574-9888
www.ornl.gov/hgmis

Managing Editor
Betty K. Mansfield
mansfieldbk@ornl.gov

Editors/Writers/ Designers
Anne E. Adamson
Denise K. Casey
Judy M. Wyrick

Production Assistants
Marissa D. Mills
Sheryl A. Martin
Laura N. Yust

U.S. Department of Energy Office of Biological and Environmental Research
Ari Patrinos, Associate Director
www.er.doe.gov/production/ober/ober_top.html

Life Sciences Division, OBER
Marvin E. Frazier, Director
www.er.doe.gov/production/ober/lsd_top.html

Contact: Daniel W. Drell, 301/903-6488, Fax: -8521
daniel.drell@science.doe.gov or *genome@science.doe.gov*

This newsletter is prepared at the request of the DOE Office of Biological and Environmental Research by the Life Sciences Division at Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, under contract AC05-00OR22725.

DOE ELSI Grants, FY 2001

FY 2001 sees the tenth cycle of applications, merit reviews, and awards in the Ethical, Legal, and Social Implications (ELSI) component of the DOE Human Genome Program (HGP). Of 26 applications received and reviewed by a peer panel, the DOE HGP funded 6 new, 2 renewed, and 3 small exploratory projects for FY 2001. These are summarized below. A more complete version and abstracts of previous DOE ELSI projects are on the Web (*www.ornl.gov/hgmis/research/elsi.html*).

DOE ELSI goals encompass research on the uses, impacts, implications, privacy, and protection of genetic information in the workplace and in databases; ELSI implications of complex or multigenic characteristics and conditions; gene-environment interactions that result in diseases or disease susceptibilities; and human polymorphisms. DOE also produces and distributes relevant ELSI educational materials for the public or specified groups, especially Institutional Review Boards and Ethics Boards.

NEW PROJECTS

Model Program for Public Libraries

Miriam Pollack, North Suburban Library System, Chicago

Focus: Use libraries for education project to encourage public "genetic literacy." Activities include organizing experts, gathering resources, and establishing programs about genetic science and ELSI issues.

Ethical Concepts in Laws Limiting Genetic Screening

Lynn Pasquerella and Lawrence Rothstein, University of Rhode Island, Kingston

Focus: Collect documents concerning laws and legislative proposals of ten eastern states and the federal government to determine how ethical concepts have influenced legislation on the use of genetic information in the workplace.

Genetics, Mental Illness, and Complex Disease

Joe McInerney, National Coalition for Health Professional Education in Genetics, Baltimore

Focus: Produce an interactive CD-ROM for instructing genetic counselors about mental disorders believed to have a genetic basis.

Economic Analysis of HGP Intellectual Property Rights

David Bjornstad, Oak Ridge National Laboratory, Oak Ridge, Tennessee

Focus: Explore the economic implications of different intellectual-property strategies for commercializing genomic information and products. The work is based in part on the research of Rebecca Eisenberg (University of Michigan Law School).

Tribal Ethical, Moral, Cultural, and Legal Issues Related to the HGP

Mervyn Tano, Institute for Indigenous Resource Management, Denver

Focus: Introduce Native American tribes to the HGP by identifying specific factors that influence perceptions of genetic research. Educate members in the basics of genetics and related research and inform DOE HGP managers about Native American perspectives.

Bioinformatics and the HGP

Mark Bloom, Biological Sciences Curriculum Study, Colorado Springs, Colorado

Focus: Produce and distribute educational module, *Bioinformatics and the Human Genome Project*, for high school audiences. The module will address ELSI issues surrounding genetic databases and will introduce teachers and students to bioinformatics.

Judicial Conference for Maine, New Hampshire, and Vermont

Elizabeth Hodges, Administrative Office of the Courts, Concord, New Hampshire

Focus: Plan and conduct a conference to introduce judges of three states to the basics of genomics and genetics. The approach is modeled on the EINSHAC conferences that have trained some 1900 judges (see article, p. 14).

RENEWAL PROJECTS

Science Literacy Workshops for Public Radio

Barinetta Scott, SoundVisions Productions, Berkeley, California

Focus: Enlarge the pool of skilled public-radio science reporters and producers and increase the number and accuracy of science reports. Four-day workshops will teach basic science, journalistic skills, and methods for crafting complex science stories.

ELSI Grants, p. 17 ➠

For Your Information

Poster Shows Disorders, Traits Mapped to Chromosomes

The revised and updated "Human Genome Landmarks: Selected Traits and Disorders Mapped to Chromosomes" is a colorful wall poster prepared by HGMIS, with printing and mailing costs supported by QIAGEN. Informative sidebars explain genetic terms and provide URLs for finding more detailed information on the Web. Expanded version: *www.ornl.gov/hgmis/posters*. A copy will be mailed to all *HGN* print subscribers. Others may request a copy from HGMIS (see p. 16) or visit the Web site. ♦

ELSI Grants *(from p. 16)*

Science Education on the Internet Conference

Ray Gesteland, University of Utah, Salt Lake City

Focus: Plan and organize the second conference for developers and Webmasters of science and biological science sites. Maximize these resources by making the sites more effective and increasing their educational value.

SMALL EXPLORATORY PROJECTS

"All Not Fit to Breed": Survivors of America's First Eugenics Movement

Mary Bishop, Virginia Polytechnic Institute and State University, Blacksburg

Focus: Collect and preserve the experiences of some ten people forcibly sterilized under Virginia's eugenics policies of the 1920s and conduct interviews with people in authority then.

Worker Perspectives on Complex Diseases

Laura Roberts, University of New Mexico, Albuquerque

Focus: Explore a possible connection between workplace exposures and complex genetic disorders by studying people with many different illnesses and examining occupational healthcare issues.

Genetics Conference for Clergy

Paul Sullins, Catholic University, Washington, D.C.

Focus: Hold a conference for the clergy, modeled on the EINSHAC series for judges (see article, p. 14). The clergy often are counsels in cases involving moral and ethical issues, such as genetic test results and reproductive choices. ♦

HGMIS Resources Available

New Version of Web Site

To address the evolving needs of a public facing unfamiliar genomic issues and challenges, HGMIS has redesigned the *Human Genome Project Information* Web site (*www.ornl.gov/hgmis*). Nearly a dozen pages have been added, and the site has been rearranged into suites.

Pharmacogenomics, Gene Therapy, and Genetic Counseling pages are new in the Medicine and the New Genetics suite, and the Gene Testing and Home pages have been revised.

The Ethical, Legal, and Social Issues suite now incorporates pages on Privacy and Legislation, Patenting, Forensics, Behavioral Genetics, and Genetics in the Courtroom, along with updated Home and ELSI Research Information pages. An online edition of the DOE ELSI Bibliography (1995) has been added to the site. The Audio/Video Online Webcasts page has been revamped as part of the Educational Resources suite.

The News Sources suite combines the traditional What's New with a weekly research digest and the new Genome Headlines and Genetics News Sources. In addition to up-to-date background text on HGP history and science, the Media Guide includes graphics that would be valuable to any interested persons as well as to those preparing news articles and presentations. An alphabetical index complements the site's existing subject-style index and search engine.

Calling All Teachers!

HGMIS would like to know how our Human Genome Project Information Web site is being used in teaching biology, genetics, and other subjects at all levels (*www.ornl.gov/hgmis*). Responses including lesson plans may be posted with the author's permission on the Education page (*www.ornl.gov/hgmis/education/education.html*). [Contact: Anne Adamson, 865/574-9851, *adamsonae@ornl.gov*]

HGP Handouts

HGMIS will send multiple copies of *HGN* and other genomics-related materials to relevant meetings on request and without charge (contact: see above). ♦

Resources

European Commission's socioeconomic research in the life sciences (*www.biosociety.dms.it*)

Nature's Genome Gateway: Research papers and news service from *Nature* and *Nature Genetics*, postgenomics section, and links; free access (*www.nature.com/genomics*)

Report on DOE's Microbial Cell Project *(http://microbialcellproject.org)*

International database on legal, social, and ethical aspects of human genetics *(www.humgen.umontreal.ca)*

GeneLetter: Originally funded by DOE, now published online monthly by GeneSage; issues since July 1996 also accessible (*www.genesage.com*)

Genetics & Your Practice: Set of professional education services for physicians and other health and social workers, sponsored by the March of Dimes (*www.modimes.org/programs2/profed/gyp.htm*)

HGBASE (Human Genic Bi-Allelic Sequences): Summaries of all known human genome sequence variations (*http://hgbase.cgr.ki.se*)

International Journal for Parasitology: Genome issue **30**(4), April 2000 (*www.elsevier.com*)

Genetic Secrets: Protecting Privacy and Confidentiality in the Genetic Era: Now in paperback, Yale University Press (203/432-0960 or *www.yale.edu/yup*)

Biotechnology: Law, Business, and Regulation: Aspen Law & Business; 1100 pp., one volume (800/638-8437, *www.aspenpublishers.com*)

Joint degree: University of Minnesota, combines law with graduate degree in health or life sciences (612/625-0055, *vanpe005@tc.umn.edu*, *www.jointdegree.org*)

Mapping Public Policy for Genetic Technologies: A Legislator's Resource Guide: Legislators and staff, free; (National Conference of State Legislatures; 303/830-2200, ext. 157; *rita.thaemert@ncsl.org*; *www.ncsl.org/programs/employ/Genetics/BOOK*)

Genes and Society: Impact of New Technologies on Law, Medicine, and Policy: Audios of speakers at the May meeting (*www.aslme.org/news/index.html*); Free CD-ROM (*cervini@wi.mit.edu*)

"The Human Genome Project: What a Legal Assistant Needs to Know": Article by Dan Drell (DOE) in *Facts and Findings*, Aug. 2000 (*www.ornl.gov/hgmis/publicat/miscpubs/legalasst.html*) ♦

Calendar of Genome and Biotechnology Meetings*

More comprehensive lists of genome-related meetings and organizations offering training are available on the Web (*www.ornl.gov/hgmis*) from HGMIS (see p. 16 for contact information).

January 2001

3–7. Pacific Symp. on Biocomputing; Honolulu [K. Lauderdale, 650/725-0659, Fax: -7944; *http://psb.stanford.edu*]

6–11. Gene Therapy 2001: Gene Odyssey; Snowbird, UT [Keystone, 970/262-1230, Fax: -1525; info@keystonesymposia.org; *www.symposia.com*]

9–14. Cell Cycle 2001; Taos, NM [see contact: Jan. 6–11]

13–17. Plant and Animal Genome IX; San Diego [D. Scherago, 212/643-1750, ext. 20, Fax: -1758; pag@scherago.com; *www.ag-microbial.org*]

17–19. Agricultural Microbes Genome II; San Diego [see contact: Jan. 13–17]

22–24. Lab on a Chip and Microarrays for Biomedical and Biotechnical Applications; Zurich, Switzerland [CHI, 617/630-1300, Fax: -1325; chi@healthtech.com; *www.healthtech.com*]

24–26. 3rd Annu. Integrated Bioinformatics: High-Throughput Interpretation of Pathways and Biol.; Zurich, Switzerland [see contact: Jan. 22–24]

25–27. Oncogenomics: Dissecting Cancer through Genomics Research; Tucson, AZ [AACR, 215/440-9300, Fax: -9313; meetings@aacr.org; *www.aacr.org*]

28–31. TIGR/ASM Microbial Genomes; Monterey, CA [TIGR, 301/610-5959, Fax: /838-0229; www@tigr.org; *www.tigr.org/cet*]

February 2001

1–2. NCHPEG 2001 Annu. Meeting; Bethesda, MD [NCHPEG, 410/337-6728; *www.nchpeg.org*]

2–7. Impact of Genomics on Drug Discovery and Development; Santa Fe, NM [see contact: Jan. 6–11]

3–6. Advances in Genome Biol. and Technol. in conj. with Automation in DNA Mapping and Sequencing; Marco Island, FL [G. Corp., 877/343-8895, Fax: 781/466-9988; ritad@gcorp.org; *www.gcorp.org*]

4–8. 26th Annu. Lorne Conf. on Protein Structure and Function; Lorne, Australia [Secretariat, +613/9662-7284, Fax: -7101; lorne.proteins@hsn.csiro.au; *http://grimwade.biochem.unimelb.edu.au/lorne.htm*]

7–13. Bacterial Chromosomes; Santa Fe, NM [see contact: Jan. 6–11]

8–11. 13th Annu. Lorne Conf. on Cancer; Lorne, Australia [J. Laird, +603/9496-3548; Fax: -3577; jacqui@austin.unimelb.edu.au; *www.ludwig.edu.au/lorne*]

12–13. NIH NHGRI Advisory Council; Bethesda, MD [K. Malone, 301/402-2205, Fax: -0837; kimberly@od.nhgri.nih.gov]

11–15. 22nd Lorne Genome Conf; Lorne, Australia [Secretariat; +612/9351-2233; Fax: -4726; M.Crossley@biochem.usyd.edu.au; *http://lorne-genome.angis.org.au*]

15–20. AAAS Annu. Meeting; San Francisco [AAAS, 202/326-6450, Fax: /289-4021; aaasmeeting@aaas.org; *www.aaas.org*]

March 2001

3–4. 4th Annu. Genomic Partnering: Emerging and Early Stage Companies; San Francisco [see contact: Jan. 22–24]

5–7. 8th Annu. Human Genome Discovery: Commercial Implications; San Francisco [see contact: Jan. 22–24]

5–11. Membrane Protein Structure/Function Relationships; Tahoe City, CA [see contact: Jan. 6–11]

6–11. Microbe Interactions with Their Environments: Genome Perspectives; Taos, NM [see contact: Jan. 6–11]

8–9. 5th Annu. Gene Functional Analysis; San Francisco [see contact: Jan. 22–24]

13–18. 21st Fungal Genetics Conf.; Pacific Grove, CA [K. McCluskey, 913/588-7044, Fax: -7295; fgsc@kuhub.cc.ukans.edu; *www.fgsc.net/fungalgenetics2001*]

21–25. 42nd Annu. Drosophila Conf.; Washington, DC [M. Ryan, 301/571-1825; mryan@genetics.faseb.org; *http://flybase.bio.indiana.edu:82/docs/news/announcements/meetings/42_Dros_Conf.html*]

31–April 4. ASBMB in conj. with Experimental Biol.; Orlando, FL [Secretariat, 301/530-7010 Fax: -7014; eb@faseb.org; *www.faseb.org/meetings/eb2001*]

31–April 6. Human Genetics and Genomics; Breckenridge, CO [see contact: Jan. 6–11]

April 2001

2–4. Proteome Project; Washington, DC [see contact: Jan. 22–24]

5–6. Protein Expression; Washington, DC [see contact: Jan. 22–24]

18–22. 4th European Symp. Protein Society; Paris [Secretariat, 301/530-7010, Fax: -7014; rroth@faseb.org; *www.faseb.org/meetings/ep01*]

19–22. HGM 2001; Edinburgh [HUGO, +44-20/7935-8085, Fax: -8341; hugo@hugo-international.org; *www.gene.ucl.ac.uk/hugo*]

21–24. RECOMB 01: Fifth Annu. Intl. Conf. on Computational Molecular Biol.; Montreal [T. Langauer, +49-2241/142-777, Fax: -656; recomb2001@gmd.de; *http://recomb2001.gmd.de*]

May 2001

15–19. 10th Intl. Congress of Human Genetics 2001; Vienna [Secretariat, +431-405/138322; office@ichg2001.org; *www.ichg2001.org*]

20–24. Toxicology: Environment Meets Genetics in the Genomics Era; Victoria, BC, Canada [G. Heaton, +44-1865/373-625, Fax: /375-855; toxicology@heaton-connexion.co.uk]

20–24. ASM 101st General Meeting; Orlando, FL [ASM, 202/942-9356, Fax: -9340; meetingsinfo@asmusa.org; *www.asmusa.org/mtgsrc/mtgs.htm*]

30–June 3. ASGT 4th Annu. Meeting; Seattle [J. Geiger, 414/278-1341, Fax: /276-3349; jgeiger@asgt.org; *www.asgt.org*]

June 2001

1–2. 22nd Annu. Health Law Teachers Conf.; Boston [S. Black, 617/262-4990, Fax: /437-7596; sblack@aslme.org; *www.aslme.org/conferences*]

24–27. BIO 2001; San Diego [BIO, 202/857-0244, Fax: -0357; *www.bio.org/events/2001/2001main.html*]

30–July 5. 11th Intl. Congress on Genes, Gene Families, and Isozymes; Stockholm [Secretariat, +46-8/459-6600, Fax: /661-9125; gene@congrex.se; *www.gene2001.org*]

July 2001

8–11. Behavior Genetics Assoc. 2001 Meeting; Cambridge, UK [T. Eley, +44-20/7848-0863, Fax: -0866; t.eley@iop.kcl.ac.uk; *www.bga.org/meetings*]

30–Aug. 3. 4th Intl. Conf. on Biological Physics; Kyoto, Japan [Secretariat, icbp2001@kokusai.phys.nagoya-u.ac.jp; *http://kokusai.phys.nagoya-u.ac.jp*] ♦

Training Calendar

January 2001

8–12. Advanced Linkage Course; New York [K. Montague, 212/327-7979, Fax: -7996; montagk@rockefeller.edu; *http://linkage.rockefeller.edu*]

February 2001

19–20. System-Based Modeling in Bioinformatics; Piscataway, NJ [M. Liebman, michael.liebman@roche.com; *http://dimacs.rutgers.edu/Workshops/index-compmolecbiol.html*]

March 2001

8–9. Protein Structure and Structural Genomics: Prediction, Determination, Technol., and Algorithms; Piscataway, NJ [S. Istrail, sorin.istrail@celera.com; *http://dimacs.rutgers.edu/Workshops/index-compmolecbiol.html*]

April 2001

19–20. DNA Sequence and Topology; Piscataway, NJ [W. Olson, olson@rutchem.rutgers.edu; *http://dimacs.rutgers.edu/Workshops/index-compmolecbiol.html*]

May 2001

6–9. Genetic Analysis of Complex Human Diseases; Durham, NC [V. Scales, 919/684-2458, Fax: -2275; vscales@chg.mc.duke.edu; *http://phg.mc.vanderbilt.edu/gachd.htm*]

14–26. Bioinformatics; Ottawa [F. Ouellette, course_info@cmmt.ubc.ca; *www.bioinformatics.ca*]

June 2001

11–16. Proteomics; Edmonton, AB, Canada [see contact: May 14–26]

21–22. Integration of Diverse Biological Data; Piscataway, NJ [A. Califano, acal@bellatlantic.net [see Web site: Apr. 19–20]

July 2001

16–21. Genomics; Fredericton, NB, Canada [see contact: May 14–26]

August 2001

7–Sept. 1. Developing Tools; Ottawa [see contact: May 14–26] ♦

*Dates and meeting status may change; courses may also be offered at other times and places; check with contact person. Attendance may be either limited or restricted.

For Your Information

NIH Centers of Excellences in Genomic Science (CEGS)

PAR-00-101

http://grants.nih.gov/grants/guide/pa-files/PAR-00-101.html

Topic: NHGRI will establish academic centers (P50 grants) for advanced genome research and support multi-investigator interdisciplinary teams to develop innovative approaches. CEGS grant (P20) offered to facilitate planning. Inquiries strongly encouraged (301/496-7531, *jeff_schloss@nih.gov*). [Check Web site for due dates.] ♦

¶ Bridging the Gap Between Life Insurer and Consumer

"Bridging the Gap Between Life Insurer and Consumer in the Genetic Testing Era: The RF Proposal," by Christopher Keefer (Barrett Law), is on the Web (*www.ornl.gov/hgmis/resource/keefer.html*). Published in the *Indiana Law Journal* in 1999, the article deals with stresses between life insurers and consumers in light of newly available genetic information that can reveal predispositions and presymptomatic conditions. Consumers fear unfair discrimination based on their genetic profiles, and insurers think applicants may conceal such data to attain higher coverage or a more favorable premium rate. Keefer discusses government response to this dilemma and proposes a solution based on risk factors.♦

¶ Computers and Biotechnology

The *Computers and Biotechnology* issue of *Your World: Biotechnology and You* [**9**(1), Fall 1999] includes molecular puzzles and stories on gene hunting, SNPs and chips, viewing molecules and proteins, and searching for drugs. In simple language for all ages, *Your World* is published by the Biotechnology Institute, which is supported by ten leading biotechnology and pharmaceutical companies. Price, subscription, ordering information, and a sample teacher's guide in PDF and HTML formats are on the Web, as are back issues and *Biotechnology and AIDS*, published this year (800/796-5806 or 814/238-4083, Fax: 814/238-4081, *www.biotechinstitute.org*).♦

¶ Workers as Human Subjects

Creating an Ethical Framework for Studies that Involve the Worker Community was produced by the Human Subjects Research Program of the DOE Office of Biological and Environmental Research. The book contains chapters on the need for worker protection, foundations of an ethical framework, challenges in using genetic data, protection of privacy, stakeholder concerns and responsibilities, and planning and implementing worker studies. [Print copies: 301/903-4731, *www.science.doe.gov/ober/humsubj/wsguidebk.html*] ♦

¶ Imaging Workshop Report

Imaging Gene Expression In Vivo: Extending Nuclear Medicine for the Next Millennium is the final report of a June 1999 workshop on imaging gene expression and the technical and scientific obstacles to this methodology. With this workshop, the DOE Office of Biological and Environmental Research took the first step toward applying fruits of the Human Genome Project to nuclear medicine. [Print copy: Sharon Betson (see Medical Sciences contact at right); PDF (printer-friendly) file: *www.er.doe.gov/production/ober/73_reports/lajolla_report.pdf*] ♦

❁Computational Molecular Biology

The Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) at Rutgers University is sponsoring the 2000–2003 Special Focus on Computational Molecular Biology (*http://dimacs.rutgers.edu/SpecialYears/2000_2003*).♦

Proteins *(from p. 8)*

RNA or protein was responsible for catalyzing link formation. The new images revealed the proteins deeply embedded in the RNA and their essential role in its folding. The images also showed where binding takes place, proving RNA's ability to break or form bonds. Further studies should show the orientation of mRNA and the growing protein's components in the ribosome's active catalytic site.

In this study, researchers used the 2.5-billion–electron volt beam to perform crystallography on painstakingly grown crystals of the 50S subunits. Additional data were gathered using the Advanced Photon Source at Argonne National Laboratory.

During X-ray crystallography, intense beams pass through and bounce off atoms in the crystal, where they leave a diffraction pattern that can be analyzed to determine the protein's 3-D shape. Researchers resolved the atomic structures of all 100,000 or so crystal atoms in the RNA 50S subunit. This involved carefully growing larger, more complete ribosome crystals and solving their structures at progressively higher resolutions. Each level provided information that helped scientists understand the final high-resolution map.♦

U.S. Genome-Related Research Funding

Investigators wishing to apply for funding are urged to discuss projects with agency staff before submitting proposals.

DOE Office of Biological and Environmental Research Human and Microbial Genome Programs

- Funding opportunities: *www.sc.doe.gov/production/grants/grants.html*
- Life Sciences Division: *genome@science.doe.gov* or 301/903-6488
- Medical Sciences Division: 301/903-3213, *sharon.betson@science.doe.gov*

Alexander Hollaender Distinguished Postdoctoral Fellowships

Research opportunities in energy-related life, biomedical, and environmental sciences, including human and microbial genomes, global change, and supporting disciplines.

- **Note: No awards will be made in 2001** (*www.orau.gov/ober/hollaend.htm*)

Computational Molecular Biology Postdoctoral Fellowships

Support career transitions into computational molecular biology from other scientific fields. Funded by DOE and the Alfred P. Sloan Foundation.

- Next deadline: February 1, 2001
- Contact: Pat Stanley, Sloan Foundation; 212/649-1628, *stanley@sloan.org*, *www.sloan.org/main.htm*

NIH National Human Genome Research Institute

- NHGRI program: 301/496-7531, *www.nhgri.nih.gov/About_NHGRI*
- Funding opportunities: *www.nhgri.nih.gov/Grant_info*
- ELSI: 301/402-4997

Small Business Innovation Research Grants

DOE and NIH invite small business firms (under 500 employees) to submit grant applications addressing the human genome topic. The two agencies also support the Small Business Technology Transfer (STTR) program to foster transfers between research institutions and small businesses.

Contacts:

- DOE SBIR/STTR Office: 301/903-1414 or -0569, Fax: -5488, *sbir-sttr@science.doe.gov*, *http://sbir.er.doe.gov/sbir*; DOE SBIR and STRR due February 20, 2001.
- Bettie Graham (see ELSI contact, NHGRI). NIH SBIR and STTR due April 1, August 1, and December 1.
- National resources, calendar: *www.zyn.com/sbir*
- National SBIR/STTR conferences: 360/683-5742, Fax: -5391, *sbir@zyn.com*.
- Alerting service: *http://lyris.pnl.gov/cgi-bin/?enter=sbir-alert* ♦