Human Genome news

Linking Interdisciplinary Contributors and Users

Sponsored by the U.S. Department of Energy www.ornl.gov/hgmis

ISSN: 1050–6101, Issue No. 44 Vol. 10, Nos. 1–2, February 1999

In This Issue

U.S. HGP on Fast Track for Early Completion

In September 1998, advisory committees at DOE and NIH approved new 5-year goals aimed at completing the Human Genome Project (HGP) 2 years earlier than originally planned in 1990. The target date of 2003 also will mark the 50th anniversary of Watson and Crick's description of DNA's fundamental structure.

The new plan was published in the October 23, 1998, issue of *Science*, which also cited the contributions of international partners. These partners include the Sanger Centre in the United Kingdom and research centers in Germany, Japan, and France.

The U.S. HGP began officially in 1990 as a $3-billion, 15-year program to find the estimated 80,000 human genes and determine the sequence of the 3 billion DNA building blocks that underlie all of human biology and its diversity. The early phase of the HGP was characterized by efforts to create the biological, instrumentation, and computing resources necessary for efficient production-scale DNA sequencing. The first 5-year plan was revised in 1993 due to remarkable technological progress, and the second plan projected goals

(see Five-Year Plan, p. 2)

DOE Joint Genome Institute Exceeds DNA Sequencing Goal

The DOE Joint Genome Institute (JGI) surpassed its sequencing goal of 20 Mb of human DNA for FY 1998, marking almost a tenfold increase in production over the previous year.

"With this milestone, JGI rises to third position worldwide in terms of its total contribution of human DNA sequence to public databases and signals great promise for completion of the entire [Human Genome] project in 5 years," noted Martha Krebs, Director of the DOE Office of Science.

Further dramatic increases are expected as JGI's main sequencing efforts move to its new facility in Walnut Creek, California, half of which has just been occupied. When the second half is completed in March 2000, about 200 staff members will maintain around-the-clock operations.

JGI's sequencing goal for the current fiscal year is 70 Mb, including 30 million "finished" bases and 40 million "draft" bases. ["Finished" sequence has been checked for accuracy, with gaps filled in to form a continuous stretch of DNA across a chromosomal region.] The JGI sequencing effort is targeting chromosomes 5, 16, and 19.

"We are seeking to break the 100-Mb barrier in the year 2000," said JGI Director Elbert Branscomb. With 1998 worldwide sequencing capacity at about 200 Mb per year, all major sequencing laboratories are ramping up production. At least 600 Mb of sequence is expected for 1999.

JGI, established at the end of 1996, is a consortium of scientists, engineers, and support staff from the Lawrence Berkeley, Lawrence Livermore, and Los Alamos national laboratories [*HGN* **8**(2), 1; *www.ornl.gov/hgmis/publicat/hgn/v8n2/01doe.html*; JGI sequencing goals and progress: *www.jgi.doe.gov*].◊

Five-Year Plan *(from p. 1)*

through FY 1998. The latest plan was developed during a series of individual and joint DOE and NIH workshops held over the past 2 years (see box, p. 3).

Observers have predicted that the 21th century will be the "biology century." The analytical power arising from the reference DNA sequences of several entire genomes and other genomic resources is anticipated to help jump start the new millennium.

Human DNA Sequencing

The HGP's continued emphasis is on obtaining a complete and highly accurate reference sequence (1 error in 10,000 bases) that is largely continuous across each human chromosome. Scientists believe that knowing this sequence is critically important for understanding human biology and for applications to other fields.

The plan calls for generating a "working draft" of the human genome DNA sequence by 2001. The working draft will comprise shotgun sequence data from mapped clones, with gaps and ambiguities unresolved. If these data sets can be merged with those from the private sector, they may increase the depth of the mapped draft, which scientists expect will contain about half the genes. Draft sequence will provide a foundation for obtaining the high-quality finished sequence and also will be a valuable tool for researchers hunting disease genes.

Resource Success Story

A Critical Resource in Public, Private Sequencing

For sequencing the human genome, scientists prefer the larger and more stable BAC clones first developed with DOE support.

BACs will be a critical component of the much-publicized private-sector sequencing efforts of such companies as Celera Genomics and Incyte [see p. 7 and *HGN* **9**(1–2), 1 (*www.ornl.gov/hgmis/publicat/hgn/v9n3/01venter.html*)]. As with all HGP resources and data, BACs are freely available to the entire research community. For more details on BACs, see p. 4.◊

According to Ari Patrinos, DOE Associate Director for Biological and Environmental Research, "Although we have as our primary goal the finished 'Book of Life' by the end of 2003, we also want the working draft to be as useful as possible."

NIH and DOE sequencing centers expect their facilities to generate about 60% to 70% of the human DNA sequence, which will be made available broadly and rapidly via the Web to stimulate further research.

Sequencing Technology

Although current sequencing capacity is far greater than at the inception of the HGP, achieving the new sequencing goals will require a two- to threefold improvement. Further incremental advances in sequencing technologies, efficiency, and cost will be needed. For future sequencing applications, planners emphasize the importance of supporting novel technologies that may be 5 to 10 years in development.

Sequence Variation

A new goal focuses on identifying individual variations in the human genome. Although more than 99% of human DNA sequences are the same across the population, variations in DNA sequence can have a major impact on how humans respond to disease; environmental insults such as bacteria, viruses, toxins, and chemicals; and drugs and other therapies.

Methods are being developed to detect different types of variation, particularly the most common type called single-nucleotide polymorphisms (SNPs), which occur about once every 100 to 300 bases. Scientists believe SNP maps will help them identify the multiple genes associated with such complex diseases as cancer, diabetes, vascular disease, and some forms of mental illness. These associations are difficult to establish with conventional gene-hunting methods because a single altered gene may make only a small contribution to disease risk.

Functional Genomics

Efficient interpretation of the functions of human genes and other DNA sequences requires that resources and strategies be developed to enable large-scale investigations across whole genomes. A technically challenging first priority is to generate complete

To find out "Who's Sequencing the Human Genome," see p. 6.

sets of full-length cDNA clones and sequences for human and model-organism genes. Other functional-genomics goals include studies into gene expression and control, creation of mutations that cause loss or alteration of function in nonhuman organisms, and development of experimental and computational methods for protein analyses.

Comparative Genomics

The functions of human genes and other DNA regions often are revealed by studying their parallels in nonhumans. To enable such comparisons, HGP researchers have obtained complete genomic sequences for the bacterium *Escherichia coli,* the yeast *Saccharomyces cerevisiae*, and the roundworm *Caenorhabditis elegans*. Sequencing continues on *Drosophila melanogaster* and the laboratory mouse. The availability of complete genome sequences generated both inside and outside the HGP is driving a major breakthrough in fundamental biology as scientists compare entire genomes to gain new insights into evolutionary, biochemical, genetic, metabolic, and physiological pathways. HGP planners stress the need for a sustainable sequencing capacity to facilitate future comparisons.

Ethical, Legal, and Social Implications (ELSI)

Rapid advances in the science of genetics and its applications present new and complex ethical and policy issues for individuals and society. ELSI programs that identify and address these implications have been an integral part of the U.S. HGP since its inception. These programs have resulted in a body of work that promotes education and helps guide the conduct of genetic research and the development of related medical and public policies.

A continuing challenge is to safeguard the privacy of individuals and groups who contribute DNA samples for large-scale sequence-variation studies. Other concerns are to anticipate how the resulting data may affect concepts of race and ethnicity; identify potential uses (or misuses) of genetic data in workplaces, schools, and courts; identify commercial

uses; and foresee impacts of genetic advances on the concepts of humanity and personal responsibility.

Bioinformatics and Computational Biology

Continued investment in current and new databases and analytical tools is critical to the success of the HGP and to the future usefulness of the data it produces. Databases must adapt to the evolving needs of the scientific community and must allow queries to be answered easily. Planners suggest developing a human genome database, analogous to model organism databases, that will link to phenotypic information. Also needed are databases and analytical tools for studying the expanding body of gene-expression and functional data, for modeling complex biological networks and interactions, and for collecting and analyzing sequence-variation data.

Training

Planners note that future genomic scientists will require training in interdisciplinary areas that include biology, computer science, engineering, mathematics, physics, and chemistry. Additionally, scientists with management skills will be needed for leading large data-production efforts.

The HGP already has revolutionized biology by providing tools and resources for basic research and has catalyzed the growth of the life sciences industry. Current and potential applications of genome research address national needs in molecular medicine, waste control and environmental cleanup, agriculture and animal husbandry, biotechnology, energy sources, and risk assessment.◊

Launchpad to Human Chromosomes

A new Web site, designed by HGMIS as a launchpad to information about all the human chromosomes, is online (*www.ornl.gov/hgmis/launchpad*). The page for each chromosome contains links to gene maps, sequences, associated genetic disorders, nonhuman genetic models, identified genes, and research efforts and laboratories. Suggestions for additions and corrections are welcome (*martinsa@ornl.gov*).◊

Five-Year Research Goals of the U.S. Human Genome Project

October 1, 1998, to September 30, 2003 (www.ornl.gov/hgmis/hg5yp)

Human DNA Sequence

- Achieve coverage of at least 90% of the genome in a working draft based on mapped clones by the end of 2001.
- Finish one-third of the human DNA sequence by the end of 2001.
- Finish the complete human DNA sequence by the end of 2003.
- Make the sequence totally and freely accessible.

Sequencing Technology

- Continue to increase the throughput and reduce the cost of current sequencing technology.
- Support research on novel technologies that can lead to significant improvements in sequencing technology.
- Develop effective methods for the advanced development of sequencing technologies and the introduction of new approaches.

Human Genome Sequence Variation

- Develop technologies for rapid, large-scale identification and scoring of single-nucleotide polymorphisms (SNPs) and other DNA sequence variants.
- Identify common variants in the coding regions of the majority of identified genes.
- Create a SNP map of at least 100,000 markers.
- Develop the intellectual foundations for studies of sequence variation.
- Create public resources of DNA samples and cell lines.

Functional Genomics Technology

- Generate sets of full-length cDNA clones and sequences that represent human genes and model organisms.
- Support research on methods for studying functions of nonprotein-coding sequences.
- Develop technology for comprehensive analysis of gene expression.
- Improve methods for genome-wide mutagenesis.
- Develop technology for large-scale protein analyses.

Comparative Genomics

- Complete the sequence of the roundworm *Caenorhabditis elegans* genome by 1998.
- Complete the sequence of the fruit fly *Drosophila* genome by 2002.
- Develop an integrated physical and genetic map for the mouse, generate additional mouse cDNA resources, and complete the sequence of the mouse genome by 2008.
- Identify other useful model organisms and support appropriate genomic studies.

Ethical, Legal, and Social Issues

- Examine issues surrounding the completion of the human DNA sequence and the study of human genetic variation.
- Examine issues raised by the integration of genetic technologies and information into healthcare and public-health activities.
- Examine issues raised by the integration of knowledge about genomics and gene-environment interactions in nonclinical settings.
- Explore how new genetic knowledge may interact with a variety of philosophical, theological, and ethical perspectives.
- Explore how racial, ethnic, and socioeconomic factors affect the use, understanding, and interpretation of genetic information; the use of genetic services; and the development of policy.

Bioinformatics and Computational Biology

- Improve content and usefulness of databases.
- Develop better tools for data generation, capture, and annotation.
- Develop and improve tools and databases for comprehensive functional studies.
- Develop and improve tools for representing and analyzing sequence similarity and variation.
- Create mechanisms to support effective approaches for producing robust, exportable software that can be shared widely.

Training and Manpower

- Nurture the training of scientists skilled in genomic research.
- Encourage the establishment of academic career paths for genomic scientists.
- Increase the number of scholars who are knowledgeable both in genomic and genetic sciences and in ethics, law, or the social sciences.◊

U.S. HGP Timeline: *www.ornl.gov/hgmis/project/timeline.html*

Genome Project

BAC End Sequencing Speeds Large and Small Projects

➤ Acronyms

A list of acronyms is printed on the back page of this newsletter.

Ultimate goals of the Human Genome Project (HGP) are to determine the sequence of the 3 billion DNA bases that make up the human genome and to increase understanding of gene function. In search of the best route to these ends, researchers have generated several different types of useful chromosomal maps. Eventually, the human genome will be represented by DNA chromosome sequences with various levels of annotation.

Interim maps have proven useful for biomedical research, but the most valuable map resources for production DNA sequencing are megabase-scale assemblies of overlapping DNA clones (contigs). Building long contigs, however, has proven a difficult task. Although contig maps of chromosomes 16 and 19 (developed at Los Alamos and Lawrence Livermore national laboratories, respectively) were largely complete in 1995, comparable contig maps of other chromosomes are less ready to support high-throughput sequencing. To help alleviate this impending bottleneck, in 1998 DOE sponsored projects to enrich the BAC clone resources preferred for high-throughput sequencing systems.

➤ Mapping with STCs and STSs

STCs

An STC is a short stretch of sequence read from one end of the human DNA insert in a clone. BAC clone STCs can be useful in a number of ways. First, STCs help researchers expand contigs, as outlined in the article. Second, when the insert length is determined, the STC spacing helps verify the contiguous sequence created by assembly software. Third, BACs with STCs serve to physically define and thus "capture" gaps that occur when sequencing biochemistry is stalled by occasional difficult-to-read stretches of DNA sequence. Finally, STC reads can be used for the design of STSs.

STSs

An STS is a DNA segment that can be copied repeatedly by PCR without amplifying unwanted DNA regions from the source genome. STS markers have been used by members of the International RH Mapping Consortium to construct the RH maps that complement contig building. STSs generated from BAC STC reads are helping to enrich RH maps. Conversely, a mapped STS can be used to isolate a BAC (or any DNA clone type) from a library of clones representing a genome.◊

BACs and STCs

BACs, which typically contain 100- to 200-kb inserts of human DNA, were designed as larger, more stable recombinant DNA clones that would represent the human genome more uniformly than previous systems. BAC development was pioneered with DOE support by Melvin Simon's team at the California Institute of Technology, with Pieter de Jong of Roswell Park Cancer Institute contributing to subsequent improvements.

Recent DOE-sponsored projects are producing sequence tag connectors (STCs) on BACs to help extend the human chromosome sequence already acquired (see figure, p. 5). STCs are DNA sequence reads at both ends of the BACs. In 1995 and 1996 investigators began to advocate that the STC concept, which had proven useful in smaller-scale sequencing projects, be applied to large-scale human genome sequencing (Venter et al., *Nature* **381**, 364–66, *orcas.htsc.washington.edu/Papers/STC_Papers_NatureCommentary.html*). DOE accepted related applications in 1996 and implemented a fast-track, special review process involving a panel composed of international experts in human and mouse genetics, mapping, sequencing, informatics, and management. Following the panel's recommendations, in September 1996 DOE initiated pilot projects at six laboratories to refine protocols and clarify cost and quality factors.

Several months later in 1997, a workshop and review was held to assess progress. Attendees recommended that DOE maintain its level of support at about $5 million a year. They also suggested concentrating STC production at sites that achieve the highest-quality sequence reads to allow the design of valuable STSs (see "Mapping with STCs and STSs," above).

High-throughput STC production is now being carried out at The Institute for Genomic Research (TIGR) under Bill Nierman and at the University of Washington Department of Molecular Biology (UWMB) by Gregory Mahairas of Leroy Hood's team. These sequencing projects are slated for completion in late 1999, with STC data sets on some 450,000 BACs. As of February 1999, more than 378,000 STCs had been acquired at the two sites (see BAC Projects in box at left).

➤ Web Sites with Related Information

***Arabidopsis* Genome Initiative**
genome-www.stanford.edu/Arabidopsis/agi.html

BAC Projects (progress, articles, resources, and related Web sites)
www.ornl.gov/bac

CalTech BAC Projects and Protocols
www.tree.caltech.edu

Genome Systems Inc.
www.genomesystems.com

German Resource Centre
www.rzpd.de

NCBI Resources and Databases: UniGene, dbEST, dbGSS
www.ncbi.nlm.nih.gov

Research Genetics
www.resgen.com

RH Consortium
www.ncbi.nlm.nih.gov/genemap

Roswell Park Cancer Institute BacPac Resource Center
bacpac.med.buffalo.edu

TIGR Human BAC Ends
www.tigr.org/tdb/humgen/bac_end_search/bac_end_intro.html

U.K. Human Genome Mapping Project Resource Centre
www.hgmp.mrc.ac.uk

University of Washington, Seattle, Human STC Project and Databases
orcas.htsc.washington.edu

STC data will provide researchers with an STC marker spaced an average of every 3000 to 4000 bases across the entire human genome, a 100-fold improvement over other current human genome maps.

The availability of STC data sets encourages more participation by smaller laboratories. Their contig building has been hindered previously by the prohibitive cost of maintaining and processing libraries on the human genome scale. With the number of STC data sets now expanding, BACs to extend chromosomal sequence can be screened computationally over the Internet. Scientists need to order only those BACs identified as candidates for contig extension (see box at right).

Enriching STC Data

Teams at UWMB and CalTech are generating additional enrichments to core BAC-STC data sets. Restriction fingerprints, which are useful for validating candidate contig extensions, will be available from UWMB for most BACs processed there. At CalTech, a team led by Ung-Jin Kim is correlating BACs with cDNAs from the NCBI UniGene* listing. These correlations will allow concurrent sequencing of chromosomal regions and their derivative cDNAs, thus promoting the interpretion of sequence function. If cDNAs already have been mapped via expressed sequence tags (ESTs) to particular chromosomal regions, their correlated BACs also will be assigned candidate positions on the chromosomes. In addition, some STCs are being used to design STSs that are useful in other mapping methods (see "Mapping with STCs and STSs," p. 4).

STCs Useful in Non-HGP Efforts on Human, Other Genomes

Several smaller-scale mapping and sequencing projects have adapted STC concepts since the HGP began. STC data sets either are in use or planned for genome projects in other species, including those for the flowering plant *Arabidopsis thaliana* and the laboratory mouse. Other examples include microbial genome sequencing strategies using STCs developed at TIGR with DOE support. The private company Celera Genomics is planning to use a similar strategy to sequence the human genome [*HGN* **9**(3), 1; *www.ornl.gov/hgmis/publicat/hgn/v9n3/01venter.html*].◊

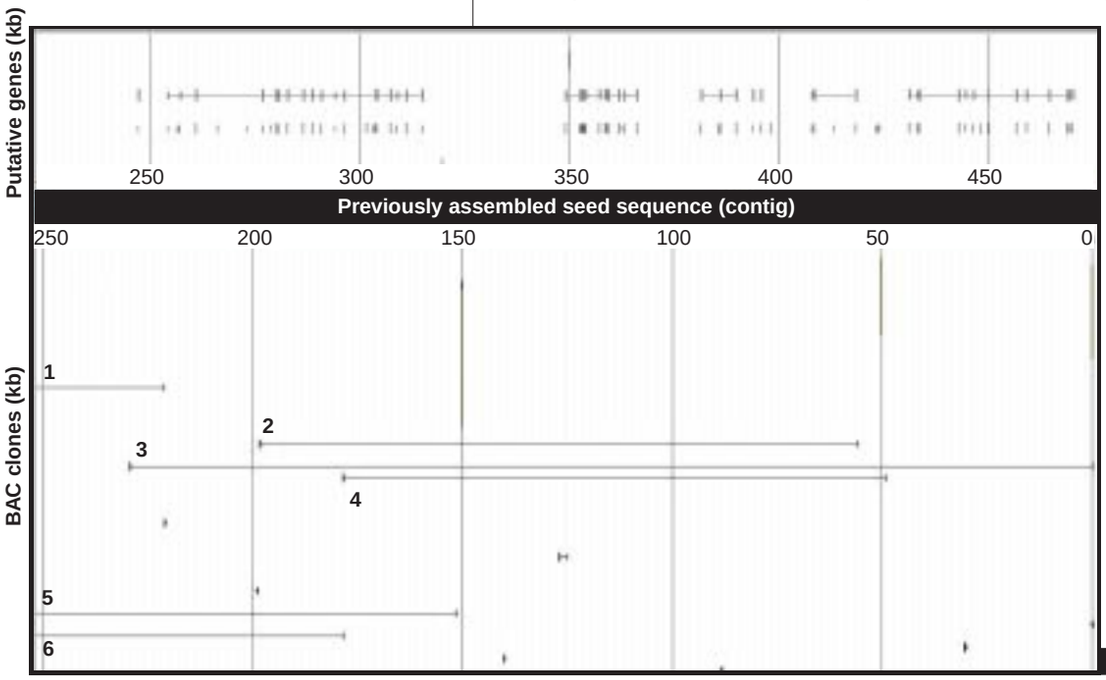
☛ Availability of BAC Clones and STC Data

Major sequencing centers may request BAC libraries directly from CalTech and Roswell Park Cancer Institute. Facilities requiring fewer BACs can obtain them through commercial suppliers after identifying needed BACs by searching the STC database against their own seed sequences. In the United States, Genome Systems Inc. and Research Genetics distribute clonal resources and provide screening services. In Europe, similar services are provided by the Sanger and German resource centers.

STC data sets are available at the NCBI database dbGSS, with more detailed information and protocols on the TIGR and UWMB Web sites. Web sites for these and other resources are listed in the box on p. 4.◊

*UniGene lists entries for nonredundant EST sequences read from the ends of cDNA clones generated and arrayed for wide distribution by the international I.M.A.G.E. Consortium [*HGN* **6**(6), 3; *www.ornl.gov/hgmis/publicat/hgn/v6n6/3image.html*)]. I.M.A.G.E. clone libraries are an outgrowth of a 1991 DOE initiative to enrich the developing human genome physical maps with gene loci and open broad access to the resulting data and resources.

BAC End Sequencing Extends Contigs. Software tools are helping to position STCs. One tool, provided by the Genome Channel (*compbio.ornl.gov/tools/index.html*), allows investigators to view the contig positions of more than 15,000 BAC end sequences and their relationships to other clones and predicted genes and exons (gene-coding regions). In the figure, the black bar represents 250 kb of a much longer contig. Below the bar, the long horizontal lines denote BAC clones, of which 1, 5, and 6 are candidates for extending the seed contig to the left. Above the bar, vertical tick marks indicate exons as predicted by GRAIL software. Exons connected by short horizontal lines represent putative gene models for the contig's forward DNA strand. [Figure contributed by Richard Mural, Morey Parang, and Manesh Shah (ORNL)]



Genome Project

Scientists Hunt SNPs to Uncover Variation, Disease

Why does one man live to celebrate his hundredth birthday with a glass of wine in one hand and a cigar in the other while another succumbs in midlife to cancer or heart disease? And why may one woman's breast cancer be effectively eradicated while another's shows no significant response to the same treatment?

The explanations may reside in the cumulative effect of a small number of differences in DNA base sequence called single-nucleotide polymorphisms (SNPs), which underlie individual responses to environment, disease, and medical treatments.

SNPs are the most common type of sequence variation. Other variations include the number of base insertions and deletions and sequence repeats (called mini- and microsatellites). Some disease-causing mutations are SNPs, for example, the single base change in the gene associated with sickle cell anemia. SNPs occur inside and outside genes, about once every 100 to 300 bases throughout the human genome.

DNA variations are important in understanding the genetic basis for disease and individual responses to environmental factors, as well as for such normal variations in biological processes as development and aging. For this reason, scientists in the public and private sectors are beginning to focus their attention on methodically searching for SNPs throughout the human genome. [See articles on new HGP goals (p. 1) and private human genome sequencing projects (p. 7).]

In 1997 the NIH National Cancer Institute launched a Genetic Annotation Initiative to gather SNPs in regions of thousands of cancer-associated genes (*www.ncbi.nlm.nih.gov/ncicgap*). In another NIH program, a 1998 RFA involves 18 institutes interested in developing genomic-scale technologies or in implementing projects to catalogue and detect SNPs in different DNA samples (*www.nhgri.nih.gov/Grant_info/Funding/RFA/rfa-hg-98-001.html*).

SNPs generated in these public projects will be freely available from dbSNP, a new database at the NIH National Center for Biotechnology Information, which serves as a central repository for SNPs and for short-deletion and insertion polymorphisms (*www.ncbi.nlm.nih.gov/SNP*). *[Denise Casey, HGMIS,* caseydk@ornl.gov*]* ◊

Who's Sequencing the Human Genome?

Listed below are the major large-scale sequencing facilities in the U.S. Human Genome Project* as of February 1999. Washington University and the DOE Joint Genome Institute led in total human DNA sequence contributed to public databases in 1998. To access the Web sites of the centers listed below, see *www.ornl.gov/hgmis/CENTERS.HTML*.

DOE-Funded

- Joint Genome Institute
- University of Washington, Seattle (BAC end sequencing, p. 4)
- The Institute for Genomic Research (BAC end sequencing, p. 4)

NIH-Funded

- Washington University, St. Louis
- Whitehead Institute/Massachusetts Institute of Technology
- Baylor College of Medicine
- University of Washington, Seattle**
- University of Texas Southwestern Medical Center
- Stanford University
- University of Oklahoma

*In addition to sequencers in the U.S. project, centers in the United Kingdom, Germany, and Japan are making major contributions toward sequencing the human genome. See URL above.

**DOE and NIH co-funded.◊

1999 DOE Human Genome Program Workshop Proceedings

Proceedings of the 1999 DOE Human Genome Program Contractor-Grantee Workshop, held January 12–16 in Oakland, California, are on the Web (*www.ornl.gov/hgmis/publicat/99santa/index.html*). The searchable abstracts, which represent DOE's latest human and microbial genome research, are categorized under Sequencing; Sequencing Technologies; Mapping; Informatics; Functional Genomics; Microbial Genome Program; Ethical, Legal, and Social Issues; and Infrastructure. An author index is included.◊

¶ *Nature Genetics* Supplement

A special supplement to *Nature Genetics* [**21**(1)] is devoted to nucleic acid microarrays in various formats. Published in January 1999 with the sponsorship of NIH NHGRI, the issue is also available on the Web (*genetics.nature.com*).◊

Science Highlights Progress in Genomics

The annual genome issue of *Science* (October 23, 1998) highlights progress in genomics, including the analysis and use of genomic data from a variety of organisms. Articles report on new plant genome initiatives, provide an overview of 10 years of plant comparative genetics, and assess the conceptual organization and approaches of some current genome-related databases. Other features include the latest plan for the U.S. Human Genome Project (p. 1) and a report on the newest physical map of human gene–based markers.

In the "Report" and "Perspective" sections, papers on the complete sequence of *Chlamydia trachomatis* summarize major findings of the sequencing project for this bacterium, which is an agent of trachoma. Trachoma is a major cause of blindness in Asia and Africa and the most common sexually transmitted bacterial disease in the United States (p. 11).

A fold-out chart of *Arabidopsis thaliana*'s genome illustrates advances in characterizing the flowering plant, a popular model for studying plant biology. Genome data generated by this project hold the potential for improved crops and plant factories that generate products such as biodegradable plastics. Another potential research outcome, which has relevance to human health, is an increased understanding of basic cellular processes extending across species. Researchers expect to finish this 120-Mb genome's DNA sequence by 2000.◊

Genome Project

GeneMap '98 Doubles Density of 1996 Map

In the special genome issue of *Science* (October 23, 1998), researchers reported the release of GeneMap '98, an updated human gene map that provides an early look at some of the most important regions of the human genome. Two to three times more detailed than the 1996 version, the new map contains some 30,000 human gene–based markers. It doubles the gene density of the previous release and represents perhaps half of all human genes.

The map highlights important chromosomal landmarks that (1) provide a valuable resource for studying complex (polygenic) genetic traits and (2) offer a framework and focus for constructing complete physical maps of chromosomes for genome sequencing. An important tool for aiding design and construction of large-scale gene-expression arrays, the map also can be used to study comparative analysis of mammalian chromosome structure and evolution. GeneMap '98 is available on a redesigned Web site that includes mapping information and associated data and annotations (*www.ncbi.nlm.nih.gov/genemap*).◊

Second Private Human Genome Sequencing Project Under Way

In August 1998, Incyte Pharmaceuticals Inc. of Palo Alto, California, announced plans to spend $200 million over the next 2 years to sequence human genes in its new unit, Incyte Genetics. Incyte also stated that it would acquire Hexagen Inc. (Cambridge, U.K.), which has developed a proprietary technique for identifying genetic variations in mice.

Incyte Genetics will concentrate on cataloguing SNPs (p. 6). Using this knowledge to design drugs—an application of genetic data known as pharmacogenetics—may help companies produce more effective therapeutics. The pace of development may be accelerated by genetically prescreening for appropriate participants in clinical trials.

The announcement came 3 months after another private company was formed for human genome sequencing. Celera Genomics, established by researcher J. Craig Venter (formerly president of The Institute for Genome Research) and Perkin-Elmer's Applied Biosystems Division, also is expected to focus on genes and their sequence variations [*HGN* **9**(3), 1; *www.ornl.gov/hgmis/publicat/hgn/v9n3/01venter.html*].

In contrast to the emphasis on identifying genes by these private companies, the sequence produced by the government-backed Human Genome Project (HGP) will reflect the entire, 3-billion–base human genome. Obtaining the complete reference human genome sequence will enable scientists to begin exploring the function of genes as well as important extragenic regions and their roles in human health and disease. The HGP also is funding the creation of clone resources for mapping and sequencing, bioinformatics and comparative genomics infrastructure, and next-generation sequencing technologies (see box, p. 3, for new goals).

All HGP data is freely accessible over the Internet and released daily for immediate use by scientists throughout the world. Celera plans to release data freely to the research community on a quarterly basis, and Incyte data will be available for a fee.◊

In the News

EMSL User Facility Promotes Remote Access to Instrumentation

On October 1, 1998, the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), DOE's newest National Scientific User Facility, celebrated the first anniversary of its opening in Richland, Washington (*www.emsl.pnl.gov*). The mission of a user facility is to provide unique research resources to scientists from DOE and government laboratories, universities, and industry. Operated by Pacific Northwest National Laboratory, EMSL's goals are to (1) attain a molecular-level understanding of the physical, chemical, and biological processes needed to solve DOE's most critical environmental problems and (2) advance molecular science in support of DOE's long-term environmental missions.

EMSL is recognized as a leader in using collaboratories to facilitate the remote use of nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry [see *Analytical Chemistry* (November 1, 1998), *pubs.acs.org/hotartcl/ac/98/nov/long.html*].

Through the Internet, EMSL is making these very expensive, cutting-edge technologies available to researchers and students who might otherwise find the instruments difficult or impossible to access (*www.emsl.pnl.gov:2080/docs/collab*).

NMR Spectrometry

Over a year ago, the EMSL Collaboratory team, in cooperation with researchers in the Macromolecular Structure and Dynamics directorate, began to develop the Virtual Nuclear Magnetic Resonance Facility (VNMRF). VNMRF is now "open for business" to allow NMR spectrometer users to conduct videoconferences with EMSL researchers, run the spectrometers remotely, collaboratively analyze data, and share their notes in a Web-based electronic notebook. A modern computer, video camera, microphone, and Internet connection are all that is needed.

During the last year, researchers from several universities and a national laboratory have saved considerable time and money by using VNMRF to participate in experiments remotely and to virtually extend or enhance visits to the facility. NMR spectrometry users desiring to run all or part of their experiments remotely can get started by simply noting their plans in their applications to EMSL.

Education

Internet access to instrumentation and researchers is bringing cutting-edge technology to the classroom. For example, undergraduate chemistry students at a small college in Washington state recently used EMSL's technology to remotely control a mass spectrometer, run spectra on unknown samples, and calculate distribution of isotopes and fragmentation patterns. Members of a sixth-grade science class, still in their Illinois classroom, manipulated a high-powered Argonne electron microscope for a close-up look at computer chips.

Other User Facilities

Other DOE user facilities are at Stanford University and at Argonne, Brookhaven, Lawrence Berkeley, Lawrence Livermore, Los Alamos, and Oak Ridge national laboratories (*www.ornl.gov/hgmis/publicat/97pr/06g_usma.html*).◊

In the News

International Team Delivers *C. elegans* Sequence

Major HGP Milestone Offers First Whole-Genome View of a Multicellular Animal

For the first time, scientists have the nearly complete genetic instructions for an animal that, like humans, has a nervous system, digests food, and reproduces sexually. The 97-million–base genome of the tiny roundworm *Caenorhabditis elegans* was deciphered by an international team led by Robert Waterston (Washington University School of Medicine, St. Louis) and John Sulston (Sanger Centre, Cambridge, England). The work was reported in a special issue of the journal *Science* (December 11, 1998) that featured six articles describing the history and significance of the accomplishment and some early sequence-analysis results.

Although sequencing has been almost completed, investigators pointed out that analysis and annotation will continue for years, facilitated by more information and better technologies. "We have provided biologists with a powerful new tool to experiment with and learn how genomes function," said Waterston. Obtaining genomic sequence, they noted, is more a beginning than an end.

C. elegans and the 12-Mb genome of the budding yeast *Saccharomyces cerevisiae* (completed in 1996) represent the only eukaryotes completely sequenced thus far. The two genomes are being compared in an attempt to identify elements essential for eukaryotic life and the genetic requirements for progression from a unicellular to multicellular existence. Eukaryotes, which include plants and animals, are the most complex of the three major branches of life on earth. The other branches are the least complex prokaryotes (bacteria) and the moderately complex Archaea, which share features with both other branches.

During its 2- to 3-week life span in the dirt of temperate regions, the benign *C. elegans* carries out many of the same processes as humans. Unlike the much smaller microbes sequenced so far, it begins life as a single fertilized cell that undergoes a series of divisions as it grows into an adult animal, forming complex tissues and organ systems. Researchers have found it particularly useful for studying early development, neurobiology, and aging—processes that have parallels in human biology.

Why Sequence Entire Genomes? A Worm's-Eye View

In the *Science* special issue on *C. elegans*, international researchers explain the rationale for sequencing farther than the protein-coding (gene) regions of a genome. They note that whole-genome data provide a basis for discovery of every gene, show long-range relationships among genes, provide structural and control elements for each gene, and offer a complete archive of genetic information. The data also provide a set of tools for future research into an organism's biology from fertilization to death, most of which is not yet understood.◊

C. elegans Data

Notes associated with the *Science* paper and links to data resources are on the sites below:

- *genome.wustl.edu/gsc/index.shtml*
- *www.sanger.ac.uk*.

The 9-year sequencing project required 2 million individual "reads" performed on DNA sequencing instrumentation to spell out the worm DNA sequence, 500 bases at a time. It began with the development of a clone-based physical map to facilitate gene analysis and grew into a collaboration among *C. elegans* Sequencing Consortium members and the entire international community of *C. elegans* researchers. In addition to the nuclear genome-sequencing effort, other researchers sequenced its 15-kb mitochondrial genome and carried out extensive cDNA analyses that facilitated gene identification. Free data exchange and immediate data release have been hallmarks of the project, which has been a model for cooperation and sharing among Human Genome Project researchers.

The first of a two-part sequencing process used to parse the *C. elegans* genome was the "shotgun" sequencing of randomly chosen subclones (each only a small piece of a much larger cloned DNA molecule). The finishing phase used a more ordered (directed) sequencing strategy to close specific remaining gaps and resolve ambiguities. Members of the Sequencing Consortium noted that, were they to begin the project again today, they would use the same combination strategy but with larger bacterial clones such as BACs. This is the strategy currently being used for large-scale human genome sequencing in the HGP (p. 4). Although tools for both sequencing phases have improved greatly over the years, finishing remains labor intensive.

The magnitude of this effort underscores the challenge of sequencing the human genome, which is some 30 times larger than that of *C. elegans*. Methods and data from the work are helping researchers sequence and interpret the human genome. In fact, a significant amount of production sequencing occurs at Washington University and the Sanger Centre.

Early analysis highlights the importance of sequencing entire genomes for finding all genes and understanding the function of nonprotein-coding DNA regions in the genomes of such eukaryotic organisms as humans and roundworms (see sidebar, "Why Sequence Entire Genomes?"). The *C. elegans* genome is packaged into 6 chromosomes containing about 19,000 genes, several times the number originally predicted by classical genetics experiments. About 40% of identified genes match those of other organisms, including humans. Like the human genome, *C. elegans* contains large amounts of repeated DNA that does not encode proteins but probably plays a role in chromosome function, gene organization, or regulation of gene activity. The *C. elegans* project was funded by NIH and the Medical Research Council (U.K.). *[Denise Casey, HGMIS]* ◊

DOE Biological and Environmental Research Helps Fuel "Biology Century"

Taking advantage of the wealth of information generated by the "new biology" of the Human Genome Project, DOE's Life Sciences Division is funding $16 million in projects that focus on high-throughput approaches to solving complex biological problems related to DOE's diverse missions. The research, which is taking place at 5 DOE national laboratories and 13 universities and research institutions, will address unresolved issues in the following 4 major areas.

Biochemical Potential of Microbes *($2 million).* Researchers seek to develop methods to decode the complete genomes of microbes more rapidly, identify potentially useful microbes, and explore their potential for energy production and use and for environmental cleanup.

Health Risks from Low-Level Exposures to Radiation and Other Energy-Related By-Products *($6 million).* New information will be useful in the ongoing development of federal health-risk policies that protect workers and the public from radiation and environmental pollutants, including those at DOE sites.

"Engineered" Biomolecules for Use in Energy Production, Environmental Cleanup, Drug Design, and Industry *($2 million).* Developing methods to rapidly determine the structure of large numbers of proteins will contribute to capabilities for designing biomolecules such as enzymes, antibodies, and other proteins for these applications.

New Genetic Information from Mice, Yeast, and Fruit Flies for Understanding Human Gene Functions More Quickly *($6 million).* This research will contribute to more accurate disease prediction and diagnosis and the design of drug therapies tailored to an individual's genetic makeup.

DOE's life sciences research program began more than 50 years ago to study the health effects of radiation, initially focusing on epidemiological studies of exposed people and genetic studies in animals. Nearly 15 years ago, DOE started planning its Human Genome Program to obtain DNA sequencing and analysis technologies and information at the genetic level regarding the effects of radiation and energy production on biological systems. In seeking to translate genomics for applications in diverse fields, DOE is helping to usher in what has been called the "biology century."

The research projects were funded following extensive peer review of proposals. [List of principal investigators and projects: *www.er.doe.gov/production/ober/projlist.html*; OBER: *www.er.doe.gov/production/ober/ober_top.html*] ◊

Hollaender Fellows Named

DOE announced the award of five FY 1998 Alexander Hollaender Distinguished Postdoctoral Fellowships for up to 2 years of research at DOE laboratories having substantial programs supportive of the Office of Biological and Environmental Research's mission. The mission is to understand health and environmental effects associated with energy technologies and to develop and sustain research programs in life, biomedical, and environmental science.

Fellowship winners were chosen from a field of 40 applicants who received their doctoral degrees after April 30, 1996. Listed below are each fellow's name, university and subject of doctoral degree, host laboratory and research mentor, and proposed research topic.

- **David Boisvert** (Yale University, Genetics): University of California at Berkeley, Sung-Hou Kim. Structural approaches to understanding ribosome biogenesis and rRNA methylation at extreme temperatures.
- **Carl Friddle** (Stanford University, Genetics): Lawrence Berkeley National Laboratory, Edward Rubin. High-throughput functional analysis of expressed sequences in the mouse.
- **Thomas Kirchstetter** (University of California at Berkeley, Environmental Engineering): Lawrence Berkeley National Laboratory, Tica Novakov. Hygroscopic growth and optical properties of carbonaceous aerosols.
- **Timothy Onasch** (University of Colorado, Chemistry): Brookhaven National Laboratory, Dan Imre. Studies of cloud particle formation mechanisms.
- **James Randerson** (Stanford University, Biology): Lawrence Berkeley National Laboratory, Inez Fung. Impact of disturbance in high-latitude terrestrial ecosystems on atmospheric measurements of CO_2, $^{13}CO_2$, and $^{14}CO_2$.

Past winners are listed on the Web site (*www.orau.gov/ober/proglist.htm*). A complete description of the program, including history and application forms, is at *www.orau.gov/ober/hollaend.htm*. See contact information on p. 23 for the Hollaender Fellowships.◊

SBIR 1998 Human Genome Awards Announced

The DOE Office of Biological and Environmental Research has announced four Phase I and three Phase II awards for 1998 in human genome topics of the Small Business Innovation Research (SBIR) program. The highly competitive SBIR awards are designed to stimulate commercialization of federally funded research and development for the benefit of both the private and public sectors. SBIR emphasizes cutting-edge, high-risk research with potential for high payoff in hundreds of areas, including human genome research (contacts: box, p. 23).

SBIR Awards in Genome, Structural Biology, Related Technologies

Searchable abstracts: *www.ornl.gov/hgmis/publicat/99santa*

Phase I

Atom Sciences, Inc. (Oak Ridge, Tennessee): A Quantitative Analytical Tool for Producing DNA-Based Diagnostic Arrays

Fidelity Systems, Inc. (Gaithersburg, Maryland): D-Strap DNA Sequencing Chemistry

MacConnell Research Corporation (San Diego, California): Automated Purification of Blood or Bacterial Genomic DNA

TPL, Inc. (Albuquerque, New Mexico): Micromachined Silicon Sensor for DNA Sequencing by Hybridization

Phase II

Cimarron Software, Inc. (Salt Lake City, Utah): (1) A Simulation Extension of a Workflow-Based LIMS and (2) A Workflow-Based LIMS for High-Throughput Sequencing, Genotyping, and Genetic Diagnostic Environments

SpectruMedix Corp. (State College, Pennsylvania): A Fully Automated 384-Capillary Array DNA Sequencer ◊

In the News

Mouse Resources Critical to Understanding Human Genome

➤ More Information

- **March meeting:** *www.nih.gov/welcome/director/reports/mgenome.htm*
- **NIH action plan:** *genetics.nature.com*
- **October meeting:** *www.nih.gov/welcome/director/reports/mgenom3.htm*
- **Mouse Genome Sequencing Network RFA:** p. 23. ◊

Some 60 scientists met for 3 days in March 1998 in Bethesda, Maryland, to define priorities for producing resources to make the mouse a more valuable tool for understanding mammalian biology. Convened by NIH Director Harold Varmus, the Mouse Genomics and Genetics Resources Working Group's recommendations, as summarized by cochairs William Dove (University of Wisconsin) and David Cox (Stanford University), are outlined below. Total direct costs for the first year are estimated at $49.3 million.

The first follow-up meeting was held in October 1998 to discuss implementation of the March recommendations. Representatives from DOE and the U.K.'s Medical Research Council were present to develop a coordinated strategy and share expertise in this international effort.

Recommendations

Recommendations for structural analysis, functional analysis, and resources include the following:

Structural Analysis

- Generate an additional 60,000 new markers, identified as crucial for scientists who are cloning genes.
- Genotype inbred mouse strains and generate a low-resolution (5-cM) single-nucleotide polymorphism map to determine its value for mouse research.
- Sequence and map 3′ ends of partial cDNAs and improve methods for isolating missing and full-length cDNAs.
- Generate 12 Mb of sequence for the first year and ramp up to 400 Mb within 5 years, obtaining a completed reference mouse genomic sequence by 2008.

Functional Analysis

- Develop standardized genome-wide mutagenesis protocols and improved tools and assays for characterizing phenotypes within new, specialized centers using the supermutagen ENU (ethyl nitrosourea, developed at Oak Ridge National Laboratory by William Russell).
- Develop phenotyping protocols in ENU centers and by individual investigators.
- Set up targeted mutagenesis programs to validate embryonic stem lines from different mouse strains for specialized uses.
- Couple molecular genotyping with the construction of congenic mouse strains.

Resources

- Develop cryopreservation methods and facilities for maintaining mutant mouse sperm and ovaries, thus reducing the cost of maintaining live animals.
- Build a new repository for live mouse strains.
- Evaluate and expand some existing databases.
- Train researchers in cryopreservation technology and animal pathology.◊

Mouse Consortium for Functional Genomics

Six Tennessee research organizations located from Memphis to Knoxville signed a Memorandum of Cooperation on December 4, 1998, to form the Tennessee Mouse Consortium for Functional Genomics. The consortium's purpose is to induce gene mutations in mice as models for human genetic diseases and as subjects for studying gene function. Consortium members are Oak Ridge National Laboratory (ORNL), University of Tennessee (Knoxville and Memphis), Vanderbilt University Medical Center, Meharry Medical College, and St. Jude Children's Research Hospital. The collaboration will combine ORNL's experience in mouse genetics and functional genomics with the other institutions' biological and clinical expertise. Vanderbilt, for example, will contribute proficiency in behavioral neurosciences, while Meharry is especially interested in mutations in the sensory systems. Each institution will play a crucial role in screening mutagenized mice for induced changes in behavior, physiology, biochemistry, and morphology and will choose mutations of interest for detailed study.

The ORNL Laboratory for Comparative and Functional Genomics, with its large collection of mutant mouse stocks and large-scale mutagenesis and phenotype screening program, is the center facility of the consortium. The six sites will be linked by the Internet, and the consortium will be managed by Darla Miller (ORNL, *bpw@ornl.gov*).◊

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
423/576-6669, Fax: /574-9888
www.ornl.gov/hgmis

Managing Editor
Betty K. Mansfield
bkq@ornl.gov

Production Assistants
Marissa D. Mills
Laura N. Yust

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

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/hug_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 Toxicology and Risk Analysis Section of the Life Sciences Division at Oak Ridge National Laboratory, which is managed by Lockheed Martin Energy Research Corp. for the U.S. Department of Energy, under Contract DE-AC05-96OR22464.◊

Chlamydia Genome Offers Surprises

Stimulates New Research into Treatment of Major STD, Prevention of Blindness

Analysis of the 1-Mb genome of *Chlamydia trachomatis* has revealed some unexpected biology for the tiny organism. *C. trachomatis* is responsible for causing the most common bacterial sexually transmitted disease (STD) in the United States as well as trachoma, a major cause of blindness in Asia and Africa. A collaboration among scientists at the University of California at Berkeley and Stanford University, the study was reported in the genome issue of *Science* (October 23, 1998).

Of 18 fully sequenced bacterial genomes, *Chlamydia* is the only obligate intracellular parasite, growing exclusively within eukaryotic cells and requiring host enzymes and cellular machinery for several necessary functions. Researchers were surprised to learn that it harbors genes that could allow it to generate its own energy-storage molecule, ATP (adenosine triphosphate).

Another new finding explained why *Chlamydia* is vulnerable to penicillin. Although *Chlamydia* was thought to lack peptidoglycan, a vital bacterial cell-wall component and the antibiotic's major target, scientists have identified the genes for synthesizing this molecule. Other genes found for new surface proteins may be important for future vaccine development, possibly by using the gene sequence itself instead of the protein to stimulate an immune response. Data are available on the *Chlamydia* Genome Project Web site (*chlamydia-www.berkeley.edu:4231*).

The project is focusing now on sequencing the genome of the organism *C. pneumoniae*, which causes a mild pneumonia and also may contribute to the development of atherosclerotic lesions. This project is funded by the genome data company Incyte Pharmaceuticals (Palo Alto, California), which also is sequencing human genes (p. 7). *[Denise Casey, HGMIS,* caseydk@ornl.gov*]* ◊

Embnet.news on Web

The latest issue of *embnet.news*, the newsletter of EMBnet, is available in html format on the Web (*www.ie.embnet.org/embnet.news*) and in printable Postscript and Adobe Acrobat formats via ftp (*ftp.ie.embnet.org/pub/embnet.news*). The newsletter contains information, articles, reviews, comments, and announcements of interest to the European and global bioinformatics communities.◊

European Biotech Program

The European Union's Biotechnology Program and funded projects are at *www.cordis.lu/biotech/home.html*. Areas of research include cell factories, genome analysis, plant and animal biotechnology, cell communication, immunology, and structural biology.◊

HUGO Consolidates Offices, Web Sites

The Human Genome Organisation (HUGO), whose purpose is to promote international collaboration within the Human Genome Project, has merged its HUGO Americas office with the London entity. The London Web site lists regional HUGO contacts and links to the HUGO Pacific office, publications and reports, and information on HGM '99 and other genome meetings (*www.gene.ucl.ac.uk/hugo*; E-mail, *hugo@hugo-international.org;* Pacific Office: Web, *hugo-pacific.genome.ad.jp*; E-mail, *tito@ims.u-tokyo.ac.jp* or *shobu@ims.utokyo.ac.jp*).◊

☛ System Identifies Polymorphisms

POMPOUS is a computational system for predicting polymorphic loci directly and efficiently from human genomic sequence (*pompous.swmed.edu*). The suite of programs detects tandem repeats ranging from dinucleotides to 250-mers, scores them according to predicted level of polymorphism, and designs appropriate flanking primers for PCR amplification.

In the verification process, the computer accurately predicted markers in genomic sequence 67% of the time. According to senior author Harold Garner (University of Texas Southwestern Medical Center), the system significantly enhances the discovery of gene function and reduces the cost of finding markers.

PANORAMA, a genetic features computation and visualization server, aims to give researchers maximum information about sequences of interest by revealing all their properties at a glance (*pompous.swmed.edu/panorama.htm*). Detailed output is provided as five major files: Features, EST Hits, Non-EST Hits, GenScan, and POMPOUS Results.◊

☛ SmithKline Licenses Software

SmithKline Beecham (SB) licensed Gene Logic's bioinformatics system and software tools based on the Object Protocol Model (OPM). OPM was developed by Victor Markowitz and his team while funded by the DOE HGP at Lawrence Berkeley National Laboratory. The program enables the rapid development of relational databases, integration of relational and flat-file databases, and building of cross-database query systems.

Gene Logic and SB also will use OPM to develop a series of customized databases and servers for integrating a wide range of public and proprietary genomic and biological data sources into SB's data-mining process. Under the agreement, Gene Logic will receive software licensing fees and funding while retaining the right to license software and products developed under the collaborative program to third-party customers.

Michael J. Brennan, president and chief executive officer of Gene Logic, said, "This relationship with SB is a validation of the power of the OPM system to manage and integrate large volumes of genomic and biological data from disparate sources into a seamless data-mining process." ◊

Microbial Genomics

Superbug Survives Radiation, Eats Wastes

"Conan the Bacterium"

A can of spoiled meat and nuclear waste may appear to have little in common, but the microbe *Deinococcus radiodurans* finds both environments rather cozy. Scientists hope this organism's ability to withstand massive doses of radiation will make it a useful tool for toxic-site remediation.

Although scientists now find it in many different soil and water sites around the world, *D. radiodurans* was not identified until 1956. It was isolated from a can of ground beef that had been radiation sterilized but had spoiled nonetheless. Perhaps because it can efficiently repair radiation breakage of its own DNA, *D. radiodurans* can endure 1.5 million rads of radiation, a dose 3000 times higher than would kill organisms from microbes to humans. Scientists are unsure how this resistance evolved, although they suspect it may be a side effect of the microbe's ability to survive periods of severe dehydration, which also fragments DNA.

Recognition of *D. radiodurans*' resistance to radiation led DOE Microbial Genome Program (MGP) managers to believe the microbe could be useful in cleaning up mixed-waste sites contaminated with toxic chemicals as well as radiation. They began to fund projects to decipher the microbe's genome and alter it to detoxify the most common chemical contaminants at these sites. Such detoxification functions might include concentrating heavy metals and breaking down organic solvents such as trichlorethylene.

Some results are reported below.

Complete Genome Sequence

The complete sequence of the 3-Mb *D. radiodurans* genome is now in hand, and researchers led by Owen White at The Institute for Genomic Research (TIGR) in Rockville, Maryland, expect to publish their findings shortly (*www.tigr.org*). The genome consists of three chromosomes and a single extrachromosomal plasmid, with repeats highly abundant on each chromosome. Circularization of chromosomal regions, occurring across repeats distributed at least every 50 kb, may be part of the homologous recombination system that is the major form of repair for DNA double-strand breaks. Researchers have not yet determined if circularization occurs more frequently after irradiation. No evidence, however, exists for a causal link between circularization and radiation resistance; the bacterium *Escherichia coli*'s genome, in fact, also circularizes and yet is radiation sensitive. Plausible explanations for the extraordinary DNA-repair capability of *D. radiodurans* remain elusive in the early analyses of DNA repair genes.

In the sequencing effort, assembly problems were encountered in repeated regions over 500 bases long and more than 95% identical. To help verify the assemblies, TIGR scientists turned to a special type of "optical" chromosome map of *D. radiodurans* constructed by David Schwartz and colleagues [New York University (NYU)].

To create this type of map, the NYU team uses optical light microscopy to directly image individual DNA molecules bound to specially coated surfaces, which are then cut with restriction enzymes. When a cut is made, the linear DNA contracts and reveals a break. Scientists create a landmark map of the DNA sequence by determining where the cut sites lie and then measuring the distances between them. This type of high-resolution restriction enzyme map provides a useful scaffold for aligning and verifying the maps predicted by standard shotgun-sequencing procedures.

Optical mapping of *D. radiodurans*, which is providing insight into this organism's biology with a picture of the entire genome's basic organization, also may help scientists understand aspects of the microbe's radiation-resistant nature.

Genetic Enhancements

Cleanup of toxic sites created by improper disposal of nuclear wastes presents a massive global challenge requiring innovative remediation approaches. In *Nature Biotechnology* (Vol. 16, October 1998), DOE grantees Michael Daly and Kenneth Minton (Uniformed Services University for the Health Sciences in Bethesda, Maryland) described a first step toward enhancing the *D. radiodurans* genome to make it valuable for toxic-site cleanup. The work also was featured in a four-page "Conan the Bacterium" article in the July–August, 1998, issue of *The Sciences*, the magazine of the New York Academy of Sciences.

In the *Nature Biotechnology* article, Daly and Minton reported sucessfully altering the microbe's genome. This was accomplished by first fusing a gene encoding toluene dioxygenase (an enzyme that degrades the organic contaminant toluene) to a *D. radiodurans* promoter (a site that activates the gene). This DNA was then inserted into one of the bacterium's chromosomes. The resulting recombinant bacterium is capable of degrading toluene and other organic compounds in a high-radiation environment. It also is tolerant of toluene and trichloroethylene's solvent effects at levels exceeding those of many radioactive waste sites. *[Denise Casey, HGMIS,* caseydk@ornl.gov*]* ◊

➤ More Information

The DOE Microbial Genome Program report, in preparation, will include information on this and other microbes. Abstracts of microbial research presented at the 1999 DOE Contractor-Grantee Workshop are on the Web (*www.ornl.gov/hgmis/publicat/99santa/microb.html*).

Unfinished Microbial Genomes Searchable

The National Center for Biotechnology Information (NCBI) Web site links to sequences from unfinished microbial genomes for BLAST searching (*www.ncbi.nlm.nih.gov/BLAST/unfin_databases.html*). These unfinished sequences, which are not yet in GenBank nor accessible via Entrez, also can be retrieved from their associated sequencing centers by ftp or Web. The 18 finished microbial genomes are searchable by Entrez via the NCBI site (*www.ncbi.nlm.nih.gov/Entrez/Genome/org.html*).◊

(see Microbials, p. 13)

Cambridge Symposium: *The Human Genome Project: Science, Law, and Social Change in the 21st Century*

The highly successful symposium, "The Human Genome Project: Science, Law, and Social Change in the 21st Century," was held in Cambridge, Massachusetts, on April 23–24, 1998. It was sponsored by the Whitehead Institute of Biomedical Research and the American Society of Law, Medicine, and Ethics and supported in part by the Ethical, Legal, and Social Issues component of the DOE Human Genome Program. This largest ELSI meeting ever was attended by more than 840 lawyers, judges, physicians, state legislators, journalists, educators, students, consumer advocates, and religious leaders. Topics at plenary sessions and breakout groups included genetic privacy, DNA databanks, genetic discrimination, doctor-patient relationships, gene therapy, newborn screening, and gene alteration.

Highlights of several selected plenary talks are given below. Eric Lander set the stage, describing the science behind the Human Genome Project. Mark Rothstein spoke on protecting genetic privacy, which is increasingly important as genetic tests become available. The final two speakers, James Wilson and LeRoy Walters, discussed gene therapy, a class of disease prevention or treatment expected to become more available as technologies unravel the genetic factors involved in disease.

☛ Free CD-ROM

Meeting syllabus, plenary talk transcripts, and Web site links available from Gus Cervini (*cervini@wi.mit.edu*)

"Genetics in the 21st Century"

Eric Lander (Whitehead Institute)

According to Eric Lander, "People today are now living through the most stunning information revolution, unlike anything before in the history of science." He compared its importance to the chemist Mendeleev's critical observation around 1869 that all the elements of matter could be organized in a very simple table. With this discovery, Mendeleev laid the foundations for the chemical industry and for much of chemistry in the 20th century. The biological sciences and industry are now experiencing the same thing, Lander stated. Instead of a periodic table, the 100,000 human genes constitute a finite list that will be complete in the near future. This list will help biologists and scientists understand the tremendous diversity of the human race and determine the causes of disease.

People are variable, Lander said, and every possible DNA sequence and DNA change that can exist probably does exist somewhere in the world. On the other hand, he continued, there are only two or three common variants of most human genes. If two people were selected at random from the audience and a particular gene were sequenced from each, the odds are one in two or one in three that the two sequences of the coding regions would be identical. This reflects the fact that the human race descended from a small population in Africa only 10,000 generations ago or about 200,000 years. Small populations have relatively few variants, and the mutation rate of one in a billion bases is so low that 95% of all the genes in the audience have not undergone a single mutation in all those years.

Even though any two human chromosomes are nearly identical, the little differences in DNA sequence can be used to trace the inheritance pattern of chromosomes and localize particular genes to particular subregions. Finding genes in this manner requires good genetic, physical, and sequence maps. The Human Genome Project has been making very good progress in these three tasks, Lander said; the genetic maps are essentially finished, and more than 97% of the genome is well covered in physical maps that can be used to isolate disease genes. Sequencing is heating up, with about 10% of the sequence expected to be finished by the end of 1998.

The process of producing 3 billion letters of information (the DNA base sequences) requires extraordinary automation and cooperation around the world. Bizarre machines are being built, Lander said as he showed a picture of a machine at Whitehead nicknamed the Genomatron, which can set up 100,000 PCR reactions in an hour. This reflects a 1000- to 10,000-fold increase in capabilities over only 4 or 5 years ago, when a student might set up 10 to 100 reactions in an hour.

What are we making of this information revolution? he asked. How far have we come toward understanding the remarkable differences among humans, the basis of different traits? Finding gene associations for rare Mendelian disorders like cystic fibrosis or Huntington's disease is a piece of cake these days, Lander stated. Over 1000 relatively rare disorders already have been mapped to specific chromosomal regions—almost all of them within the last 10 years, and all within the last 14 years. About 140 have been specifically isolated and cloned.

For common diseases, the challenge has been to tease apart the contributions of multiple genes associated

(see Lander, p. 14)

Microbials *(from p. 12)*

TIGR Releases *Chlorobium tepidum* Sequence

In September 1998, The Institute for Genomic Research (TIGR) announced the release of more than 1.9 Mb of genome sequence from *Chlorobium tepidum*, a photosynthetic gram-negative bacterium (*www.tigr.org/cgi-bin/BlastSearch/blast.cgi?*). The TIGR program, supported by DOE, has reached 3× coverage in the random-sequencing phase. The photosynthetic *C. tepidum* may play an important role in the earth's overall cycle of carbon use.◊

ELSI: Cambridge Symposium

"Protecting Genetic Privacy: Why It is So Hard to Do"

Mark A. Rothstein (Health Law and Policy Institute, University of Houston Law Center)

Mark Rothstein began his presentation by assuring the audience, "Although it will be more complicated than most people imagine, protecting genetic privacy and confidentiality is a worthy goal." Steps taken toward this goal so far, however, he characterized as misguided and simplistic. Before explaining this position further, he gave the audience useful background information on relevant issues.

Rothstein defined "privacy" as the limited access to a person, the right to be let alone, and the right to keep certain information from disclosure to other individuals. "Confidentiality," he said, is the right of an individual to prevent the redisclosure of certain sensitive information that was originally disclosed in the confines of a confidential relationship. Protecting confidentiality can be difficult because others think they should have the right to see an individual's information.

Rothstein listed eight nonmedical uses of genetic information: insurance, employment, criminal law, personal-injury litigation, domestic relations, forensics, education, and commerce. Data also are being used for identification and in such contexts as immigration, paternity, settlement of estates, kinship, and schools.

In criminal law, defendants already are attempting to use as a defense unproved theories of genetic predisposition to violent behavior. When such a defense fails, defendants invoke similar claims as a method of mitigating their punishment at the sentencing stage.

Should defendants in personal-injury cases be allowed to compel victims to undergo genetic testing to determine what their life expectancy might have been before the accident? In child-custody cases, should the risk of an inherited disease keep either parent from gaining custody? How much genetic testing should be authorized before children are placed for adoption? Should a mortgage company be allowed to require a genetic test to assess an applicant's life expectancy?

Because of the financial incentives involved, Rothstein said, confidentiality is particularly difficult to maintain in health insurance and employment. Laws have been enacted in 16 states to prohibit insurance companies from using genetic information to deny coverage or raise health insurance rates. When these laws were passed, Rothstein pointed out, people thought they were wonderful. Now, however, it has become clear that they protect only individuals who are asymptomatic. Once symptoms become apparent, the laws don't apply. Rothstein suggested that a comprehensive law would need to say that no insurance company may deny coverage or raise rates based on an individual's past, present, or predicted health status.

On the federal level, Rothstein cited the Health Insurance Portability and Accountability Act (HIPAA), which applies to employer-based and commercially issued group health insurance. Although HIPAA is a step forward, he said, it does not apply to the unemployed, and employers are not required to provide any health insurance or specific benefits. Rothstein said, "I think we are living in a rather hopeful—or naive—world, where we may temporarily have been able to contain double-digit increases in medical costs and where we've been able to put our finger in the dike on the issue of the uninsured. But there are problems lurking, and I don't know that we are doing enough to address those issues."

Moving to employment discrimination, he cited data showing that 85% of people surveyed said they should be protected from having employers obtain their health records. Some employers, on the other hand, feel that they can save a great deal of money by eliminating prospective employees and

(see Rothstein, p. 15)

Lander *(from p. 13)*

with complex conditions. The most progress has been made by looking for rare Mendelian subtypes, but there are as yet no good published subtypes for asthma, schizophrenia, and bipolar disease, for example.

Human genetics eventually may come down to just one very large table of variants or traits common to the population. People already are talking about collecting all the roughly 300,000 variants (3 for each of the 100,000 genes) and genotyping everybody. This is what genetics may look like in the 21st century, Lander continued.

He showed some examples of extreme claims, particularly those in supermarket tabloids, regarding genes and how they determine what kind of work a person may do, whom he will marry, or how much money she will earn. As the audience laughed, Lander pointed out that if the subject were Alzheimer's disease or thrill seeking, it's not clear where the public would draw the line regarding behavior or other traits that might be explained by genes.

"We have to make the advantages of this genetic revolution available for biomedical research and yet still fight what I think is the danger of a naive biological determinism and the consequences that could have for society. We need a different model. The right model, for me, is captured on a poster [showing two people] I'm very fond of from the Musée de L'Homme in Paris, from an exhibit they had some years ago: '*Tous parents, tous differents.*' It can be translated two ways: '*All the same, all different,*' or '*all related, all different.*' "

Genetic variations influence our lives, he concluded, but they don't constrain us, nor do they shape us in the choices we can make as a society. What has happened so far in the information revolution will seem like nothing when compared to what will flow from the sluice gates of human genetics projects around the world over the next decade or so. We must explore "how to manage the information," Lander said, "and the choices and consequences of what science has to offer."◊

(lander@wi.mit.edu)

"Human Gene Therapy: Present and Future"

James M. Wilson (Institute for Human Gene Therapy, University of Pennsylvania)

In his presentation at the 1998 Cambridge meeting, James Wilson characterized gene therapy as a novel approach in its very early stages. Its purpose, he said, is to change the expression of some genes in an attempt to treat, cure, or ultimately prevent disease. Current gene therapy is primarily experiment based, with a few early human clinical trials under way.

Theoretically, he continued, gene therapy can be targeted to somatic (body) or germ (egg and sperm) cells. In somatic gene therapy the recipient's genome is changed, but the change is not passed along to the next generation. This form of gene therapy is contrasted with germline gene therapy, in which a goal is to pass the change on to offspring. Germline gene therapy is not being actively investigated, at least in larger animals and humans, although a lot of discussion is being conducted about its value and desirability.

Gene therapy should not be confused with cloning, which has been in the news so much in the past year, Wilson continued. Cloning, which is creating another individual with essentially the same genetic makeup, is very different from gene therapy.

Listing three scientific hurdles in gene therapy, Wilson emphasized the concept of vehicles called vectors (gene carriers) to deliver therapeutic genes to the patients' cells. Once the gene is in the cell, it needs to operate correctly. Patients' bodies may reject treatments, and, finally, there is the need to regulate gene expression. Wilson expressed optimism that many groups are making headway and cooperating to overcome all these obstacles.

Viruses have evolved a way of encapsulating and delivering their genes to human cells in a pathogenic manner. Scientists have tried to take advantage of the virus's biology and manipulate its genome to remove the disease-causing genes and insert therapeutic genes. These gene-delivery vehicles will make this field a reality, he said.

In the mid-1980s, the focus of gene therapy was entirely on treating diseases caused by such single-gene defects as hemophilia, Duchenne's muscular dystrophy, and sickle cell anemia. In the late 1980s and early 1990s, the concept of gene therapy expanded into a number of acquired diseases. When human testing of first-generation vectors began in 1990, scientists learned that the vectors didn't transfer genes efficiently and that they were not sufficiently weakened. Expression and use of the therapeutic genes did not last very long.

In 1995, Wilson continued, a public debate led to the consensus that gene therapy has value although many unanswered questions require continued basic research. As the field has matured over the last decade, it has caught the attention of the biopharmaceutical industry, which has begun to sort out its own role in gene therapy. This is critical because ultimately this industry will bring gene therapies to large patient populations.

Wilson reviewed several specific gene-therapy cases involving high cholesterol, hemophilia, and cystic fibrosis. He emphasized that the response to any therapy in a heterogeneous patient population will be quite variable.

He asked the audience to think about gene therapy, not necessarily to treat

(see Wilson, p. 16)

Rothstein *(from p. 14)*

dependents whose medical expenses are likely to be high.

In March 1995, the Equal Employment Opportunity Commission issued an interpretation that is helpful but not the final word, Rothstein continued. Basically, it says covered entities that discriminate on the basis of genetic predisposition are regarding the individuals as having impairments, and such individuals are covered by the Americans with Disabilities Act. The problem here, Rothstein said, is that this interpretation is not binding on the courts and does not apply to unaffected carriers of recessive and X-linked disorders. It also does not prohibit employers from requiring access to employees' clinical records, which could include genetic information. The consequences of this interpretation, Rothstein said, are that it permits disclosure of sensitive information within companies and discourages at-risk people from being tested.

Some 13 states have enacted laws that prohibit employers from requiring genetic testing or from using genetic test results to discriminate in employment. Unfortunately, Rothstein said, these laws are either too narrow or too broad. They don't protect genetic information in medical records or prevent employers from gaining access through health-insurance claims.

Rothstein then raised the questions: Is genetic information unique? Should it be protected separately from other forms of information? He listed six arguments for considering genetic information different from other kinds of medical data: It reveals the health of family members; it reveals parentage, reproductive options, and future health risks; it goes to the essence of who and what an individual is; and it's regarded as unique by individuals and third parties, who often overuse it. Rothstein said that, even if we are satisfied that genetic information is unique, it should not necessarily be protected separately. First, people don't know exactly what genetic information is. Second, it probably is impossible to segregate it from other information in a clinical record; and third, enacting genetic-specific legislation may be self defeating because it further stigmatizes people with genetic conditions.

The problem of genetic discrimination cannot be solved by a single law, Rothstein concluded, and resolution of the issue raises fundamental concerns of equality in the system. Due to the complexity and difficulty of the challenge, we should start to address these problems in depth.◊

(mrothstein@uh.edu)

ELSI: Cambridge Symposium

"Ethical Issues in Human Gene Therapy"

LeRoy Walters (Kennedy Institute of Ethics, Georgetown University)

LeRoy Walters provided a valuable perspective on some of the lessons learned by scientists and ethicists over the 18 years since the first human gene therapy protocol was approved. He also offered his predictions for future gene-therapy interventions and discussed some associated ethical dilemmas that society may be facing.

Walters began his talk with two case studies. The first was about David, known as "the boy in the bubble." He was born in 1971 with X-linked severe combined immune deficiency and died 12 years later after receiving a bone marrow transplant that, unknown to doctors, carried a silent Epstein-Barr virus.

In contrast to David's story, Walters continued, is the story of Ashanti, who was born in 1986 with an autosomal recessive form of severe combined immune deficiency. In Ashanti's early years, every environmental microbe attacked her body and made her sick. She was treated with a synthetic enzyme called PEG-ADA, which gradually decreased in efficacy, and in 1990 she became the first patient to receive gene therapy in an approved protocol. She is now almost 13 years old and living a normal life.

In reviewing the history of gene therapy in the United States, Walters referred to a document prepared by an interdisciplinary group in 1984 and 1985. Called "The Points to Consider," it contained 110 questions that investigators were asked to answer as they thought about performing gene therapy on human patients. The questions covered such topics as gene therapy's potential benefits and harms, fairness in selection of recipients, procedures to be followed, recipients' privacy and confidentiality, and possible alternative therapies. The same questions could constitute a checklist for gene therapy today, Walters said.

The review process in the early days was transparent and public, a fact that was important to gene therapy's acceptance. Policymakers knew exactly what was happening, and any member of the public could attend a meeting, see the investigators, hear the questions, and have access to a public list of approved gene therapy protocols.

Walters stated that as of February 1998, 200 therapeutic protocols had been formally reviewed: 23 dealing with HIV infection or AIDS; 33 with single-gene diseases, especially cystic fibrosis; 138 with cancer; and 6 with other diseases. Reviewing what has been learned from the past 18 years, he listed the following points:

- Somatic cell gene therapy has been successfully distinguished from more ambitious plans for human genetic engineering. ➨

Wilson *(from p. 15)*

genetic disease but as an alternative way to deliver proteins. Protein therapeutics currently are manufactured by placing genes in laboratory-cultured organisms that produce the proteins coded by those genes. Examples of such manufactured proteins include insulin, growth hormone, and erythropoietin, all of which must be injected frequently into the patient.

Recent gene therapy approaches promise to avoid these repeated injections, which can be painful, impractical, and extremely expensive. One method uses a new vector called adeno-associated virus, an organism that causes no known disease and doesn't trigger patient immune response. The vector takes up residence in the cells, which then express the corrected gene to manufacture the protein. In hemophilia treatments, for example, a gene-carrying vector could be injected into a muscle, prompting the muscle cells to produce Factor IX and thus prevent bleeding. This method would end the need for injections of Factor IX—a derivative of pooled blood products and a potential source of HIV and hepatitis infection. In studies by Wilson and Kathy High (University of Pennsylvania), patients have not needed Factor IX injections for more than a year.

In gene therapies such as those described above, the introduced gene is always "on" so the protein is always being expressed, possibly even in instances when it isn't needed. Wilson described a newer permutation in which the vector contains both the protein-producing gene and a type of molecular rheostat that would react to a pill to regulate gene expression. This may prove to be one of gene therapy's most useful applications as scientists begin to consider it in many other contexts, he said. Wilson's group is conducting experiments with ARIAD Pharmaceuticals to study the modulation of gene expression.

Wilson stated that only so much can be done in academia and that the biopharmaceutical industry has to embrace gene therapy and handle issues of patents, regulatory affairs, and the optimum business model. An example of a dilemma that society may be facing can be seen in the treatment of hemophilia. Infusing a patient with the replacement protein, which stops bleeding episodes but doesn't prevent them, currently costs about $80,000 a year. Why would a vector to prevent bleeding for 5 to 10 years be commercialized when it would displace such a lucrative treatment, and how would this gene therapy be delivered to the public?

Wilson concluded his presentation by outlining future milestones in the field: proof of concept in the next few years in model inherited diseases, followed by cancer and cardiovascular diseases; continued explosive activity in technological development; development of regulatory policy (with the Food and Drug Administration); and commercial development.◊

(wilsonjm@mail.med.upenn.edu)

➤ To find out more about Gene Therapy, see site below (select "Disease Intervention")

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

- The more neutral term "human gene transfer" might have been used, rather than "human gene therapy." "Therapy" seems to promise benefits to the patient; "gene transfer" covers even the Phase I studies that test a product's toxicity and are unlikely to be therapeutic to the subjects.
- The success of human gene therapy has been quite modest in the first 8 years; unfortunately, some researchers and companies have overstated the early results.
- An optimum location will be needed for a national public review body to examine new biomedical technologies.

Looking to the future, Walters said he thinks we will see prenatal interventions to prevent severe and irreversible damage to fetuses and gene transfer to prevent or treat neurological disease. In studies affecting the brain, the question of what is enhancement and what is cure, treatment, or prevention of disease will arise in an acute form, he said. For example, is it remediation or enhancement to intervene so that a child would have an IQ of 100 instead of 60 or 70?

Walters predicted that, in the next 18 years, proposals will emerge for germline genetic intervention, which will require a great deal of preliminary technical work. Instead of the current technologies of adding genes, something analogous to the "search and replace" function on a word processor will be needed to find the malfunctioning gene, splice it out, and replace it with the properly functioning gene.

He pointed out that there are some good moral arguments in favor of germline genetic intervention, whose goal is to prevent or alleviate disease or disability. Such intervention is more efficient than repeating gene therapy generation after generation, and even in utero gene therapy is too late for some diseases. The one case that could justify nuclear transfer in the early embryonic stage, Walters thought, is that in which a woman is likely to pass on a mitochondrial disease to her offspring. In such a situation, he said, after in vitro fertilization it would be justified at perhaps the four-cell stage to remove all the cells' nuclei and fuse them with enucleated egg cells from a donor. Because mitochondria are in the cytoplasm and would be derived from the donor, the resulting embryos would be free from mitochondrial disease. This type of case would involve simultaneous germline intervention and cloning in the technical sense.

Walters ended with a warning against repeating mistakes made in the time of the eugenics movement and the Third Reich. "We can applaud the war on disease that genetic research is waging. It will be a great day when a child is definitively cured of cystic fibrosis or when a particular family line is liberated from the burden of fragile X syndrome. But we will be humane warriors only if, in the midst of the battle, we also show respect for those who courageously cope with disability and for those who cannot yet be cured." ◊

(waltersl@gunet.georgetown.edu)

ELSI News

On Radio: *The DNA Files*

On November 2, 1998, an interactive Web site was launched for *The DNA Files*, a series of nine 1-hour documentaries hosted by John Hockenberry and distributed by National Public Radio (*www.dnafiles.org*). Supported in part by DOE, the series covers such topics as DNA and behavior, prenatal and predictive genetic testing, gene therapy, genetics of human evolution, genetics and biotechnology, and genetics and the law. The Web site, which lists radio stations that will broadcast *The DNA Files* around the country, provides information about each program, additional resources, and an opportunity for listeners to interact about some ethical issues introduced in the series. [Contact: *bariscot@aol.com* or *jthilman@aol.com*] ◊

Innovative Biotechnology Curriculum

An innovative curriculum to boost student enthusiasm and interest in biotechnology has been launched through a partnership involving the National 4-H Council, Monsanto Corporation, and Pioneer Hi-Bred International. Called *Fields of Genes: Making Sense of Biotechnology,* the curriculum is designed to help teachers provide students in grades 4–12 with a basic understanding of scientific principles that form the foundation of biotechnology. Curriculum activities for elementary students focus on understanding the living and nonliving parts of their world. Middle school students continue to explore and understand genetics, biotechnology, and genetic engineering, while high schoolers are encouraged to plan environmental stewardship activities. [Order leaders' guide (96 pages, Product No. ES0046) from National 4-H Supply Service: 301/961-2934, Fax: -2937] ◊

Microbial TV Series

Intimate Strangers: Unseen Life on Earth, a four-part series for prime-time public television, will be shown by PBS and distributed for international broadcast this fall. Funded in part by the DOE Human Genome, Microbial Genome, and Natural and Accelerated Bioremediation Research programs, the series is designed to increase public science literacy by using lessons from the microbial world to teach about more complex systems of life.

The series, produced by independent filmmakers Baker & Simon Associates with the American Society for Microbiology, is an initiative of the Microbial Literacy Collaborative (MLC). An organizational partnership headed by Cynthia Needham, MLC seeks to emphasize how basic research advances society's well-being, to improve decision-making on microbial issues, and to create more effective curricula for science teachers and students. In addition to *Intimate Strangers*, other MLC products include an interactive Web site (operational in May: *www.microbeworld.org*), a set of hands-on activities designed to introduce youth to the microbial world, and week-long leadership programs targeted to young people from challenging environments. *Unseen Life on Earth*, a 12-part telecourse for undergraduates, will support distance learning and provide teaching resources to college and precollege teachers.◊

Short Courses for Biology Teachers

Outreach to K–12 teachers and students is an aim of the new molecular biology teaching laboratory at Pennsylvania State University's Biotechnology Institute. Short courses including lectures and laboratory experience can be scheduled for area teachers on the principles and techniques used in genetic and molecular biology research, especially as they relate to the Human Genome Project. High school biology teachers are particularly encouraged to take advantage of this opportunity. [Contact: Loida Escote-Carlson (814/863-5751, *lje6@psu.edu*)] ◊

Proteomics

From Sequence to Systems: Looking at Proteins to Understand Genome Expression

The availability of entire genomic sequences for some 18 microbes (and many more to come) now offers investigators the opportunity to perform comparative analysis from an evolutionary perspective, identify conserved genes and metabolic capabilities based on protein sequence homology, and predict protein structures. Understanding how gene products—proteins—work together to create and maintain complex biological systems, however, requires data about the entire spectrum of protein production in the complex ecosystem of a cell.

In the account below, DOE Microbial Genome Program grantee Carol Giometti of Argonne National Laboratory (ANL) describes such studies on two microbial genomes, the heat-loving Methanococcus jannaschii *and* Pyrococcus furiosus, *both subjects of the DOE program. [Introduction by Dan Drell, DOE Microbial and Human Genome programs]*

In 1995, V. Wasinger and coworkers (University of Sydney, Australia) coined the term "proteome" to describe all the proteins encoded within a genome. Proteomics is the study of protein expression by biological systems, including relative abundance, post-translational modifications, stability within the cell, and fluctuations as a response to environment and altered cellular needs.

In contrast to genomic sequence, which captures DNA information that is stable throughout the lifetime of an organism, proteomics summarizes protein-expression patterns of a biological system at different times. Biochemical pathways and regulatory mechanisms can be deduced by manipulating the cellular environment or DNA sequence and observing coregulation of specific proteins or sets of proteins. Proteomics tools include high-resolution protein separation, detection, and quantitation methods and techniques for linking proteins to their corresponding gene sequences. These tools can be used to further annotate and validate completed genomes, reveal biochemical pathways and regulatory networks, and define targets for protein-structure determination.

In the context of the DOE Microbial Genome Program, analyzing the proteomes of organisms for which complete genomes are available offers the potential for rapid identification of the organisms' major gene products.

Although *M. jannaschii*'s complete genome sequence is publicly available and annotated according to sequence homology with other known proteins, the actual proteins synthesized by *M. jannaschii* and regulation of their synthesis have not been studied until now. Correlation of protein abundance, shifts in abundance in response to environmental changes, and post-translational modifications with the genome sequence will provide new information regarding gene expression and regulation in this member of the Archaea. In addition, proteome studies will serve to confirm or refute protein identifications based on sequence homologies alone.

The genome sequence of *P. furiosus* is virtually complete, and numerous *P. furiosus* enzyme activities have been well characterized. The regulation of specific gene expression (e.g., inducibility of enzyme activities of interest) is not characterized in *P. furiosus*, however, nor has the influence of post-translational modification been explored. Characterization of the *P. furiosus* proteome will bridge the gap between gene sequence and protein function by providing data on the regulation of protein synthesis. In addition, studies are in progress to determine the subcellular localization (soluble vs membrane fractions) of each *P. furiosus* protein.

Strategies rooted in 2-DGE (see sidebar) are being developed to link the proteome information with existing genome sequence databases for these two Archaea. Evolving approaches to characterizing small-genome proteomes and linking proteome and genome databases will be the foundation for developing protocols for similar investigations of large mammalian proteomes. *[Carol S. Giometti, ANL, csgiometti@anl.gov]* ◊

2-DGE: A Technique for Visualizing Protein Expression and Modification

One current proteomic tool for visualizing and quantitating all proteins expressed in a biological system at a given time is two-dimensional gel electrophoresis (2-DGE). As originally described by Patrick O'Farrell for analyzing *Escherichia coli* proteins in 1975, 2-DGE combines the electrophoretic separation of denatured proteins by isoelectric point charge differences in the first dimension with separation based on molecular size differences in the second dimension. The proteins, which can be detected using protein-specific stains, appear as constellations of spots in the 2-D space of the gel. Over the 20-year history of 2-DGE, numerous algorithms have been developed for comparing 2-DGE patterns and quantitatively analyzing protein abundance.

The recent addition of mass spectrometry to methods available for identifying proteins detected by 2-DGE has provided the needed capability for rapid identification. Proteins can be digested in the 2-DGE gel using a specific protease (e.g., trypsin or amino acid–specific endopeptidases), the resulting peptides eluted, and the masses of proteins determined using mass spectrometry. [Matrix-assisted laser desorption ionization (MALDI-MS) and electrospray currently are the preferred methods.] The peptide masses are then used to search protein and DNA sequence databases for the identity of predicted amino acid sequences to produce the same peptide masses when cleaved with the same protease. When a complete genome sequence is available and the peptide mass search is limited to just that sequence database, the protein identification process is highly reliable and efficient.

The work of John Yates's group at the University of Washington, in which 260 *Haemophilus influenzae* proteins separated by 2-DGE were identified in about a month with MALDI-MS, demonstrates this approach's potential for identifying the hundreds of proteins revealed in the 2-DGE patterns of cell lysates. The protein-expression information can then be compared to the cell's proteome and other proteomes to provide a better understanding of cell function.◊

Proteomics News

Tool for Protein Analysis

PEDANT is a software system for completely automatic and exhaustive analysis of protein sequence sets, from individual sequences to complete genomes (*pedant.mips.biochem.mpg.de*). This server now contains 20 complete genomic sequences and 1 plasmid, as well as 21 experimental and unfinished genomic sequences.

Entries for completed genomes include three sections:

- General Information such as genome summary, open reading frames, links, and search mechanism;
- Protein Function such as closest homologues, functional categories, Protein Information Resource keywords and superfamilies, and PROSITE patterns; and
- Protein Structure such as known 3-D, transmembrane, signal-peptide, low-complexity, coiled-coil, and structural classes.◊

TREMBL Release 6

Release 6 of TREMBL, a protein sequence database that supplements SWISS-PROT, has been announced. TREMBL contains the translations of all EMBL Nucleotide Sequence Database coding sequences not yet integrated into SWISS-PROT. Weekly TREMBL updates are available by anonymous ftp (*ftp.ebi.ac.uk/pub/databases/trembl*) and from the Sequence Retrieval System server of the European Bioinformatics Institute (*srs.ebi.ac.uk*). [Contact: *apweiler@ebi.ac.uk*] ◊

R&D 100 Award to LANL's SOLVE

One of the four R&D 100 awards won by Los Alamos National Laboratory in 1998 was for SOLVE, a system that produces 3-D pictures of protein structure. SOLVE automatically carries out all the steps necessary to fill in missing information in X-ray crystallography, a process that uses X rays to determine the structure of atoms, ions, or molecules in chemical substances. SOLVE's speed and ease of operation make it suitable for the rapid analysis of protein molecule shapes, and accurate protein pictures can be produced in hours rather than days. In addition, the automated system can evaluate hundreds of solutions and can be operated by a novice. SOLVE shows promise in helping researchers design new and improved drugs, enzymes for rapidly breaking down toxic waste or synthesizing useful chemicals, and heat-tolerant enzymes useful in chemical manufacturing processes.

Technologies funded by DOE accounted for 34 of the 100 R&D awards in 1998. [SOLVE Contact: Thomas Terwilliger (505/667-0072, *terwilliger@lanl.gov*)] ◊

NIH Proteomics Grant to Axys

Axys Pharmaceuticals Inc. of South San Francisco, California, has been awarded a Phase I Small Business Innovation Research grant from the NIH National Institute of General Medical Sciences to conduct a 6-month research study of proteomics. Proteomics is the global search for and identification and prediction of protein function. The Axys goal is to build the ProteomeBank, a software system and proprietary database of protein families for high-throughput, accurate prediction of protein function.◊

Completing the *E. coli* Proteome

A database of genes characterized since completion of the *Escherichia coli* genome sequence lists new and old gene names, SWISS-PROT entry, gene location, genetic structure, and identified function (*sun1.bham.ac.uk/bcm4ght6/genome.html*).◊

Genetics in Medicine

Organization for Rare Disorders

The National Organization for Rare Disorders (NORD) is a federation of more than 140 nonprofit voluntary health organizations dedicated to helping people with rare "orphan" diseases and to assisting the groups that serve them.

➤ NORD Publications

The third edition of the 675-page *NORD Resource Guide* lists more than 900 organizations that can benefit individuals with rare disorders and their families. The 1000-page *Physicians' Guide to Rare Diseases* contains information on over 900 such disorders, including symptoms and visual diagnostic signs. [Orders: *www.rarediseases.org*, click on "Services/Products." The physicians' guide may also be ordered from Dowden Publishing Co. (800/707-7040 or 201/391-9100, Fax: -2778).]

Orphan diseases, most of which are genetic in origin, are those affecting fewer than 200,000 people in the United States. More than 5000 rare disorders affect about 20 million Americans.

Responding to over 1 million inquiries each year, NORD attempts to educate the public and the medical community by distributing understandable information through its newsletters, publications, and databases; providing referrals to additional resources; and maintaining an extensive Web site (*www.rarediseases.org*). Through the Web, users can access NORD's Rare Disease Database (RDB), containing more than 1100 abstracts, as well as the Organization Database of support groups and the Orphan Drug Database. Complete RDB entries are available online at low cost, and printouts can be ordered from the NORD office.

NORD also maintains confidential patient networking for individuals and families. Since 1987, NORD has administered medication assistance programs for pharmaceutical companies, providing free prescription drugs from nine companies to thousands of uninsured, needy patients. In addition, the NORD grant program provides financial support to academic scientists for clinical research. [NORD; P.O. Box 8923; New Fairfield, CT 06812-8923 (800/999-6673 or 203/746-6518, Fax: -6481)] ◊

Cancer Web Site

Northwestern University researchers have developed a Web site to teach health professionals and the public about the genetic basis of cancer and new discoveries in the field of cancer genetics. Designed as a comprehensive educational program, the site provides a fundamental understanding of genetics, genetic testing and diagnosis, genetic counseling, and cancer risk assessment (*www.cancergenetics.org*).◊

(More on Genetics in Medicine, p. 20)

Informatics

☛ *Software Progams Provide Useful Resources*

BioToolKit

BioToolKit now provides 750 annotated links to Web tools for the study of nucleic acid, genome, and protein structure (*www.biosupplynet.com/cfdocs/btk/btk.cfm*).◊

Gene-Finding Programs at Sanger

Updated versions of gene-finding programs (including FGENES, FGENESH, and FGENES-m variant for mammalian sequences) are available for use through the Sanger Web site (*genomic.sanger.ac.uk*). Also, the Gapped BLASTP program from the National Center for Biotechnology Information allows users to check a gene's protein structure in the INFOGENEP database of finished and unfinished human sequences and receive the clone's name and sequence (*genomic.sanger.ac.uk/db.html*). See the Web site for more information.◊

New Sequin Version

The National Center for Biotechnology Information has released Version 2.80 of Sequin, the sequence-submission and editing tool, for all platforms (*www.ncbi.nlm.nih.gov/Sequin*). This version is expected to be particularly useful for genome centers that annotate large records.◊

Tandem Repeat Tool

Gary Benson (Mount Sinai School of Medicine) has developed a program to find tandem repeats in DNA sequence data without prior knowledge of pattern repeat, pattern size, or number of copies (*c3.biomath.mssm.edu/trf.html*). The current version finds pattern repeats ranging from 1 to 500 bases.

Users submitting a sequence (up to 2 Mb) in FASTA format will receive a summary table of repeats, including their location, size, number of copies, and nucleotide content. Clicking on an entry shows alignment against a consensus pattern, allowing the user to see the repeat pattern and mutation location.◊

Sequence Viewer

Sequence Viewer, a free public software tool for viewing and analyzing DNA sequences, is available on the National Center for Genome Resources (NCGR) Web site (*www.ncgr.org/gsdb/sv*). The NCGR tool was developed to fill the need for graphical representations of nucleotide sequences in the Genome Sequence DataBase and for detailed descriptions of sequence annotation. Sequence Viewer allows users to quickly find a sequence region that integrates with a gene rather than searching through a lengthy, complex flat-file report. It also can be used as a quality-control tool for readily locating mistakes in feature position.◊

Genetics in Medicine *(from p. 19)*

New HGMIS Site: Translation of Genetics to Medicine

At the request of medical professionals eager for translation of genomics to medical practice, the Human Genome Project Information suite of Web sites has added a new page called "Medicine and the New Genetics"(*www.ornl.gov/hgmis/resource/medicine.html*). This site covers topics of specific interest to physicians, nurses, genetic counselors, and allied health professionals. It contains information and links about disease prevention, diagnosis, and intervention; genetic-disease databases and support groups; gene testing; gene therapy; pharmacogenomics; genetic counseling; ethical, legal, and social issues associated with genetics; continuing medical education courses in genetics; publications; multimedia; professional societies; and other resources.

Medical professionals are asked to review the site and send comments and suggestions to HGMIS.◊

HuGEM Web Site

The Human Genome Education Model Project (HuGEM) offers education in the new genetics to specific groups of health professionals who provide services for individuals and families with genetic conditions (*www.dml.georgetown.edu/hugem/elsi.htm*).◊

¶ Calculation of Genetic Risks

The Calculation of Genetic Risks: Worked Examples in DNA Diagnostics (second edition) by Peter Bridge (Alberta Children's Hospital, Canada) explains how to calculate an individual's genetic risk based on information from genetic testing and family pedigrees. Worked examples are included. 272 pp., 1997. Order through bookstores or from Johns Hopkins University Press (800/537-5487, Fax: 410/516-6998).◊

¶ Genetics Manual

Genetics Manual: Current Theory, Concepts, Terms by George P. Redei (University of Missouri, Columbia) explains over 18,000 life science terms and concepts arranged alphabetically. Cross-references connect to a network within the book for comprehensive information on any covered topic. Further sources are given for many entries, and most biometrical procedures have worked examples. 1152 pp., 1998. [Orders: World Scientific Publishing Co. (800/227-7562, Fax: 888/977-2665, *sales@wspc.com*, *www.wspc.com*] ◊

Mutation Journal

Devoted to the union between genomics and mutation research, the fourth issue of *Mutation Research Genomics Online*—a section of *Mutation Research Online*—is at *www1.elsevier.com/journals/genomics/menu.htm*. Although full online access is restricted to subscribers, informative snapshots are available for selected articles in each issue.◊

☛ DNA Polymorphism Discovery Resource

A resource for detecting DNA sequence polymorphisms has been developed by the NIH National Human Genome Research Institute in collaboration with the NIH National Institute of General Medical Sciences and its Human Genetic Mutant Cell Repository. Designed to reflect the diversity of the human population, the resource is composed of cell lines and DNA samples from 450 unrelated individuals, both male and female. In addition to the complete set, predefined nested subsets with 8, 24, 44, and 90 samples will encompass the same range of diversity. Individuals sampled include Americans of European, African, Mexican, and Asian extraction as well as Native Americans [F.S. Collins et al., *Genome Research* **8**(12), 1229–31, 1998]. (Orders: 800/752-3805 or 609/757-4848; *ccr@arginine.umdnj.edu*; *locus.umdnj.edu/nigms*) ◊

Databases

GDB Mapping Database Operations Restored

Canadian Institution Takes Over Collection, Curation

The Bioinformatics Centre at the Hospital for Sick Children (HSC) in Toronto, Canada, has received funding from an anonymous source to continue data acquisition and curation activities of the Genome Database (GDB). Work is under way to obtain additional support for future software development.

GDB, which provides human gene mapping data to genetics researchers, was based at Johns Hopkins University (JHU) School of Medicine in Baltimore until July 1998. At that time, DOE withdrew major funding to focus its informatics resources on the sequencing phase of the Human Genome Project.

The new income will enable HSC to send GDB data from the central editable node to international nodes. The U.S. node will be maintained by the Computational Biosciences Section at Oak Ridge National Laboratory (ORNL) in conjunction with the Genome Annotation Consortium (GAC). GAC is a multi-institutional group established to help build a shared infrastructure for integrating diverse biological information [*HGN* **9**(3), 13; *www.ornl.gov/hgmis/publicat/hgn/v9n3/13anno.html*].

The ORNL node has been established (*genome.ornl.gov*), and the primary node is being transferred to HSC. To retain key members of the GDB staff and a presence in the United States, HSC is supporting a curatorial center at JHU. In addition to continuing GDB operations and access, researchers at participating institutions are exploring further collaborations in acquiring, analyzing, and exchanging data to benefit the genome community.

[Contacts: HSC, Jamie Cuticchia (*jamie@genet.sickkids.on.ca*); JHU, Christopher Porter (*cporter@gdb.org*) and Conover Talbot, Jr. (*cct@gdb.org*); ORNL and GAC, Edward Uberbacher (*ube@ornl.gov*) and Jay Snoddy (*snoddyj@ornl.gov*)] ◊

Publications

¶ Bioinformatics Journal

In Silico Biology, a peer-reviewed online journal, attempts to bridge the gap between experimental scientists and computational biologists by focusing on biologically significant computational methods and results (*www.bioinfo.de/isb*). A print version is also available. [Subscribe via Web site or by e-mail: *admin@bioinfo.de*] ◊

¶ Computational Methods

Computational Methods in Molecular Biology, edited by Steven Salzberg (The Institute for Genomic Research), David Searls (SmithKline Beecham Pharmaceuticals), and Simon Kasif (University of Illinois at Chicago), was published by Elsevier Science in 1998 (*www.cs.jhu.edu/~salzberg/compbio-book.html*). Leading researchers from the computational biology community are included among the authors.

Biologists who rely on computers are the primary audience, with a secondary audience of computer scientists who are developing techniques with biology applications. A list of Web resources in the book will be kept updated on the Web (*www.cs.jhu.edu/~salzberg/appendixa.html*). 398 pp., hardbound. [Orders: Web, *www.elsevier.com*, search on "Salzberg"; 888/437-4636 or 212/633-3730; *usinfof@elsevier.com*] ◊

(see Informatics, p. 23)

Influenza Database at LANL

Los Alamos National Laboratory (LANL) introduced its annotated Influenza Sequence Database in July 1998 (*www-flu.lanl.gov*). The database currently holds all the influenza sequences published in GenBank and, after verification and annotation, will add unpublished sequences collected around the world. LANL is working with the University of California and the Centers for Disease Control and Prevention to expand the database.◊

TRANSFAC Database

The TRANSFAC database compiles data about gene regulatory DNA sequences and protein factors binding to them (*transfac.gbf.de*). Programs help identify putative promoter or enhancer structures and suggest their features.

TRANSFAC consists of six cross-linked tables: SITE, CELL, FACTOR, CLASS, MATRIX, and GENE. FACTOR entries also are cross-linked with a proposed classification system for transcription factors (*transfac.gbf.de/TRANSFAC/cl/cl.html*). TRANSFAC is linked to a number of other databases. Among the most recent additions are enhanced internal hyperlinking between individual tables, improved linking of references to PubMed, and insertion of most training sequence sets used for matrix construction, including the corresponding site-matrix links.

The TRANSFAC server also provides access to such sequence-analysis tools as PatternSearch, which uses sequence information contained in the SITE table for analysis of submitted sequences; and MatInspector, which uses a library of matrices selected from the TRANSFAC MATRIX table. Another sequence-analysis program, FastM, developed by the group of Thomas Werner (National Research Centre for Environment and Health, Neuherberg), is included on the TRANSFAC server. Using the MatInspector algorithm, FastM analyzes sequences for user-defined combinations of transcription factor–binding sites. The Structural Analysis with Genetic Algorithms (SaGa) program can identify structural characteristics in the environment of aligned functional sites.

TRANSFAC tools are freely accessible for users from noncommercial organizations. Users from profit-oriented organizations are requested to obtain licensing from BIOBASE Ltd. (*info@biobase.de*).

[TRANSFAC contact: Edgar Wingender (+49-531/6181-427, Fax: -266, *ewi@gbf.de*)] ◊

p53 Mutation Database

The p53 mutation database contains information on all p53 missense mutations and small deletions in human tumors and cell lines as reported in peer-reviewed literature (*www.iarc.fr/p53/whatsdb1.htm*).◊

TBASE at Jackson Laboratory

TBASE, the database of transgenic animals and targeted mutations, is at the Jackson Laboratory in Bar Harbor, Maine (*tbase@jax.org*; *www.jax.org/tbase*).◊

Intein Database on Web

The Intein Database Web site contains a registry of all submitted experimental and theoretical inteins (inframe protein introns) as well as information on protein splicing and intein structure (*www.neb.com/neb/inteins.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*) or from HGMIS (see p. 10 for contact information).

April 1999

8–9. Functional Proteomics:Integrating Technologies for Target Discovery and Disease Therapy; Boston [IBC, 508/481-6400, Fax: -7911; reg@ibcusa.com; *www.ibcusa.com*]

8–10. Human Genetics: Principles and Applications in Medical Practice; Monterrey, Mexico [A. Morales, +11-528/348-6982; al603471@academ07.mty.itesm.mx]

11–14. RECOMB '99; Lyon, France [INRIA, +33-139/635-053, Fax: -638; symposia@inria.fr; *www.inria.fr/RECOMB99*]

14. Genes Synapses and Long-Term Memory. TIGR/NRC/DOE Distinguished Speaker Series: Eric Kandel (Columbia Univ. College of Physicians and Surgeons); Washington, DC [D. Hawkins, 301/838-3501, Fax: -0209; dhawkins@tigr.org; *www.tigr.org*]

15. Using the Yeast Genome Sequence to Learn About Biol. NHGRI Lecture Series: Robert Waterson (CSHL); Washington, DC [L. Brooks, 301/496-7531; lisa_brooks@nih.gov]

17–21. Experimental Biol. '99; Washington, DC [Meeting office, 301/530-7010, Fax: -7014; eb@faseb.org; *www.faseb.org/meetings/eb99/index.htm*]

19–20. Protein Expression; Washington, DC [CHI, 617/630-1300, Fax: -1325; chi@healthtech.com; *www.healthtech.com*]

23–24. Protein Sequence Structure Function; San Francisco [K. Clarke, 415/476-1913, Fax: /502-4690; kristina@cgl.ucsf.edu; *mdi.ucsf.edu/PSSF_Mtg.html*]

27–28. 2nd Intl. Workshop on Advanced Genomics: Genomics and Drug Discovery; Tokyo [Secretariat, +81-3/5563-4342, Fax: -4887; genome@lbs.co.jp]

May 1999

8. Clinical Cancer Genetics: A Practical Approach; Baltimore [Johns Hopkins Univ. School of Medicine CME Office, 410/955-2959, Fax: -0807; ckowarski@jhmi.edu]

13–14. Gene Quantification-Europe; Munich, Germany [see contact: April 19–20]

16–20. BIO '99; Seattle, WA [Meetings Dept., 202/857-0244, Fax: /331-8132; *www.bio.org/meetings*]

16–20. New World Science for the Next Millennium. 1999 ASBMB Joint Meeting; San Francisco [ASBMB, 301/530-7010, Fax: -7014; asbmb@asbmb.faseb.org; *www.faseb.org/meetings/asbmb/asbmb99*]

17. Genomic Partnering-Europe; Munich, Germany [see contact: April 19–20]

17–18. NIH NHGRI Advisory Council Meeting; Bethesda, MD [K. Malone, 301/402-2205, Fax: -0837; kimberly@od.nhgri.nih.gov]

18–19. 3rd Annu. Human Genome-Europe; Munich, Germany [see contact: April 19–20]

19–23. Genome Sequencing and Biol.; Cold Spring Harbor, NY [CSHL, 516/367-8346, Fax: -8845; meetings@cshl.org; *www.cshl.org*]

20. Involving Children in Genetic Susceptibility Research: Implications for Informed Consent. NHGRI Lecture Series: Gail Geller (Johns Hopkins University); Washington, DC [see contact: April 15]

25–30. 19th Intl. Conf. on Yeast Genetics and Molecular Biol.; Rimini, Italy [L. Frontali, +39-06/445-3950, Fax: /446-1980; frontali@axcasp.caspur.it; *www.icgeb.trieste.it*]

29–June 1. 31st Annu. Meeting of the European Society of Human Genetics; Geneva [J. van Goyenkade,+31-20/679-3411, Fax: /673-7306; eshg@eurocongres.com; *eurocongres.com/eshg*]

30–June 3. ASM General Meeting; Atlanta [ASM, 202/942-9248, Fax: -9340; MeetingsInfo@asmusa.org; *www.asmusa.org/jnlsrc/news/mtgconf.htm*]

June 1999

9–10. Proteome; San Francisco [see contact: April 19–20]

10–13. American Society of Gene Therapy Annu. Meeting; Washington, DC [M. Stallings, 609/848-1000 ext. 264, Fax: -5274; mstallings@slackinc.com; *www.asgt.org/natmtg.html*]

14–15. Bioinformatics; San Francisco [see contact: April 19–20]

14–15. DNA Forensics; McLean, VA [see contact: April 19–20]

16. Protein Structure; San Francisco [see contact: April 19–20]

17. Whole Genome Analysis by Optical Mapping. NHGRI Lecture Series: David Schwartz (New York University); Washington, DC [see contact: April 15]

19–24. 26th FEBS Meeting; Nice, France [G. Dirheimer, +33-388/417-055, Fax: /602-218; Guy.Dirheimer@ibmc.u-strasbg.fr; *coli.polytechnique.fr/febs99*]

24–25. Genomics, Functional Genomics, Proteomics, and Beyond: New Horizons for the 21st Century Academia-Pharmaceutical Collaborations; Paris [L. Drye, Fax: +33-140/613-405; euroconf@pasteur.fr; *www.pasteur.fr/Conf/euroconf.html*]

July 1999

4–7. 1999 Behavior Genetics Assoc. Meeting; Vancouver, B.C., Canada [R. Rose, 812/855-8770, Fax: -4691; rose@indiana.edu; *www.bga.org/1999*]

11–15. 9th European Congress on Biotechnol.; Brussels [Secretariat, +32-2/706-8174, Fax: -8170; secretariat@ecb9.be; *www.ecb9.be*]

August 1999

5–9. Microbial Biodiversity; Chicago [see contact: May 30–June 3]

8–13. Human Molecular Genetics; Newport, RI [GRC, 401/783-4011, Fax: -7644; grc@grcmail.grc.uri.edu; *www.grc.uri.edu*]

17–22. Yeast Cell Biol.; Cold Spring Harbor, NY [see contact: May 19–23]

September 1999

13–14. NIH NHGRI Advisory Council Meeting; Bethesda, MD [K. Malone, 301/402-2205, Fax: -0837; kimberly@od.nhgri.nih.gov]

18–21. 11th Intl. Genome Sequencing and Analysis Conf.; Miami [TIGR, 301/838-3515, Fax: -0229; seqconf@tigr.org; *www.tigr.org*]

October 1999

5–10. 10th Intl. Congress on Genes, Gene Families, and Isozymes: Advances in Genome Research and Their Implications for Biol. in the 21st Century; Beijing [N. Wang, +8610/6255-1158, Fax: -1951; xuegx@public.east.cn.net]

6–10. Neurobiology of *Drosophila*; Cold Spring Harbor, NY [see contact: May 19–23]

16–19. NSGC 18th Annu. Educ. Conf.; Oakland, CA [NSGC, 610/872-7608; nsgc@aol.com; *www.nsgc.org*]

19–23. ASHG; San Francisco [C. Galkin, 301/571-1825, Fax: /530-7079; cgalkin@genetics.faseb.org; *www.faseb.org/genetics/ashg/ashgmenu.htm*]

February 2000

27–Mar. 2. Eighth DOE Human Genome Program Contractor-Grantee Workshop; Santa Fe, NM [Sylvia Spengler, 510/486-4879, Fax: -5717, sjspengler@lbl.gov] ◊

*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.

Training Events*

April 1999

23–25. 1999 Genetics Review Course; Schaumburg, IL [M. Greenfield, 301/571-1887, Fax: -1895; mgross@faseb.org; *www.faseb.org/genetics/acmg*]

30–May 2. Medical Genetics and Genetic Counseling Review Course; Pittsburgh (May 14–16, Oakland, CA) [NSGC, 610/872-7608; nsgc@aol.com; *www.nsgc.org*]

June 1999

9–29. Advanced Bacterial Genetics; Cold Spring Harbor, NY (app. deadline Mar. 15) [CSHL, 516/367-8346, Fax: -8845; meetings@cshl.org; *www.cshl.org*]

12–17. Contemporary Challenges in Health Care Ethics. Intensive Bioethics Course; Washington, DC [Kennedy Inst. of Ethics, 202/687-5477; kicourse@gunet.georgetown.edu; *guweb.georgetown.edu/kennedy/courses/ibc98.htm*]

October 1999

4–14. Gene Expression Analysis. Theoretical and Practical Course; Monterrey, Mexico [H. Barrera-Saldaña, +52-8/329-4173, Fax: /333-7747; hbarrera@ccr.dsi.uanl.mx; *www.icgeb.trieste.it*]

13–26. Genome Informatics; Cold Spring Harbor, NY (app. deadline July 15) [see contact: June 9–29]

13–26. Macromolecular Crystallography; Cold Spring Harbor, NY (app. deadline July 15) [see contact: June 9–29] ◊

DOE: Office of Science Grants and Contracts

www.er.doe.gov/production/grants/grants.html

Comprehensive ELSI Program Notice: Notice will be issued in early summer, 1999.

99-03: Environmental Meteorology Program—Vertical Transport and Mixing

99-06: Environmental Management Science Program—Research Related to Subsurface Contamination/Vadose Zone Issues

99-07: Financial Assistance Program, Energy Biosciences

99-08: Next-Generation Internet Research in Basic Technologies

99-09: Next-Generation Internet Applications, Network Technology, and Network Testbed Partnerships

99-10: Next-Generation Internet University Network Technology Testbeds

99-11: Fundamental Research in Carbon Management

99-14: Low-Dose Radiation Research Program ◊

NHGRI: National Research Service Award Fellowships

Topic: To engage in research relevant to the Human Genome Project. Fellowships for postdoctoral, senior postdoctoral, and predoctoral minorities or persons with disabilities are available to U.S. citizens or permanent residents; research in ethical, legal, and social issues (ELSI) is not open to predoctoral students through this program.

- **Applications for postdoctoral and senior postdoctoral due:** December 5, April 5, and August 5.
- **Applications for predoctoral minorities or persons with disabilities due:** May 1 and November 15.
- **Contacts:** ELSI topics, Elizabeth Thomson (301/402-4997, *elizabeth_thomson@nih.gov*); all other topics, Bettie Graham (see NHGRI contact information in box) ◊

Informatics *(from p. 21)*

¶ Bioinformatics Guide

Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins is intended to help the molecular biologist design and implement a successful sequence-analysis strategy using the overwhelming array of tools available, including Internet resources. Edited by Andreas D. Baxevanis (NIH NHGRI) and B.F. Francis Ouellette (then at NIH National Center for Biotechnology Information), the book is a collection of chapters on relevant topics from 16 contributors. Paperback, 362 pp. 1998.◊

NCI: Technologies for Molecular Analysis

PAR-98-066

www.nih.gov/grants/guide/pa-files/PAR-98-066.html

The National Cancer Institute (NCI) invites small business applications for basic, clinical, and epidemiological research to develop novel technologies for molecular analysis of cancers and their host environment. Funding mechanisms will be SBIR, STTR, and NCI's recently developed Phased Innovation Award (PAR-98-067).

- **Letter of Intent due:** March 5, 1999
- **Application due:** April 9, 1999
- **Contact:** Carol Dahl (301/496-1550, *carol_dahl@nih.gov)* ◊

NIH: Network for Large-Scale Mouse Sequencing

RFA HG-99-001 *(released 12/98)*

www.nih.gov/grants/guide/rfa-files/RFA-HG-99-001.html

Topic: Establish a Mouse Genome Sequencing Network that will support mouse genome mapping and sequencing. Network goals are to generate necessary mapping resources and begin a working draft of the mouse genome DNA sequence. Applications are encouraged for pilot sequencing projects in new groups and the capacity expansion of existing sequencing centers.

- **Letter of Intent due:** March 1, 1999
- **Application due:** April 29, 1999
- **Contacts:** See contact, NHGRI. Sequencing: JanePeterson (*Jane_Peterson@nih.gov*; Mapping Resources: Bettie Graham (*Bettie_Graham@nih.gov)* ◊

NHGRI: Genomic Technology Development

PAR-99-047

www.nhgri.nih.gov/Grant_info/Funding/Research/advtech.html

Topic: Advanced development of high-throughput methods, hardware, and software in support of genomic research. Initial emphasis will be on DNA sequencing technologies.

- **Letter of Intent due**: February 15 and July 15
- **Application due**: April 26 and September 21
- **Contact**: Jeffery Schloss (see contact, NHGRI) ◊

U.S. Genome 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 Genome Program

- Funding information, inquiries: *genome@science.doe.gov* or 301/903-6488
- Relevant documents: *www.er.doe.gov/production/ober/hug_top.html*

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.

- Next deadline: January 2000
- Contact: Barbara Dorsey, Oak Ridge Institute for Science and Education (423/576-9975, Fax: /241-5220, *dorseyb@orau.gov, www.orau.gov/ober/hollaend.htm)*

Computational Molecular Biology Postdoctoral Fellowships

Topic: Support career transitions into computational molecular biology from other scientific fields. Funded by DOE and the Alfred P. Sloan Foundation to give young scientists an intensive 2-year postdoctoral opportunity in an appropriate molecular biology facility.

- Contact: Christine Trance; Alfred P. Sloan Foundation; 630 Fifth Ave., Ste.2550; New York, NY 10111 (212/649-1649, Fax: /757-5117, *trance@sloan.org)*

NIH National Human Genome Research Institute

- NHGRI program: 301/496-7531, Fax: /480-2770, *www.nhgri.nih.gov/About_NHGRI*
- Program announcements: *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*. SBIR applications due March 2, 1999. STTR due April 8, 1999. SBIR: *sbir.er.doe.gov/sbir*; STTR: *sttr.er.doe.gov/sttr*
- Bettie Graham (see contact, NHGRI). NIH SBIR due April 15, August 15, and December 15. STTR, April 1, August 1, and December 1

National SBIR/STTR conference: April 9–11, 1999, Washington, D.C. (*www.zyn.com/sbir*; *teddy@seeport.com*). For regional conferences, see Web site. ◊

Human Genome Management Information System Subscription/Document Request* (Vol. 10, Nos. 1–2)

Name _____
(First) (MI) (Last)

Affiliation _____

Department/Division _____

Street/P.O. Box/Building _____

City/State/Zip Code _____

Country _____ **Area of Interest** _____

Phone _____ **Fax** _____

E-Mail Address (important to list if you have one) _____

***Please type, print carefully, or enclose a business card to ensure efficient shipping. To change name/address/affiliation or drop your subscription to *Human Genome News*, enclose your current *HGN* address label. Send to HGMIS address shown below and on p. 10.**

Request for Print Subscription and Documents (all online documents listed at *www.ornl.gov/hgmis/publicat/publications.html*)

1. *Human Genome News*

___New Subscriber

___Change of Name/Affiliation/Address (circle all that apply)

___Drop Subscription

2. ___*Human Genome Program Report, Parts 1 and 2* (*www.ornl.gov/hgmis/publicat/97pr*)

3. ___*To Know Ourselves* (*www.ornl.gov/hgmis/tko/index.html*)

4. ___Reprint of **"What Can the New Gene Tests Tell Us?"** by Denise Casey, HGMIS (*www.ornl.gov/hgmis/publicat/judges/judgetoc.html*), *Judges' Journal* 36(3), Summer 1997

Documents Available Online Only

Primer on Molecular Genetics (*www.ornl.gov/hgmis/publicat/primer/intro.html*)

Five-Year Plan (*www.ornl.gov/hgmis/hg5yp*)

Medicine and the New Genetics (*www.ornl.gov/hgmis/resource/medicine.html*)

SELECTED ACRONYMS

ASBMB Am. Soc. for Biochem. and Mol. Biol.
ASHG Am. Soc. for Hum. Genet.
ASM Am. Soc. for Microbiology
BAC bacterial artificial chromosome
BIO Biotechnology Industry Organization
CHI Cambridge Healthtech Inst.
cM centimorgan
CME continuing medical education
CSHL Cold Spring Harbor Lab.
DOE Dept. of Energy
ELSI ethical, legal, and social issues
EST expressed sequence tag
FEBS Fed. of Eur. Biochem. Soc.
FISH fluorescence in situ hybridization
GRC Gordon Res. Conf.
HGMIS Human Genome Management Information System
HGP Human Genome Project or DOE Human Genome Program
IBC Intl. Business Communications
I.M.A.G.E. Integrated Molecular Analysis of Gene Expression
INRIA French Natl. Inst. for Research in Computer Science and Control
kb kilobase
Mb megabase
NCBI Natl. Center for Biotechnology Information
NHGRI Natl. Human Genome Research Institute
NIH Natl. Institutes of Health
NRC Natl. Research Council
NSGC Natl. Soc. of Genetic Counselors
OBER Office of Biological and Environmental Research
PAC P1 artificial chromosome
PCR polymerase chain reaction
RFA Request for Applications
RECOMB Conference on Computational Molecular Biology
RH radiation hybrid
SNP single-nucleotide polymorphism
STC sequence tag connector
STS sequence tagged site
TIGR The Inst. for Genomic Res.
WWW World Wide Web
YAC yeast artificial chromosome

OAK RIDGE NATIONAL LABORATORY • MANAGED BY LOCKHEED MARTIN ENERGY RESEARCH CORP. • FOR THE U.S. DEPARTMENT OF ENERGY

Betty K. Mansfield
HGMIS
Oak Ridge National Laboratory
1060 Commerce Park, MS 6480
Oak Ridge, TN 37830, U.S.A.

Postmaster: Do Not Forward.
Address Correction Requested.
Return Postage Guaranteed.

BULK RATE
U.S. POSTAGE
PAID
OAK RIDGE, TN
PERMIT NO. 3

Recycled
Recyclable