Human Genome news

Linking Interdisciplinary Contributors and Users

Sponsored by the U.S. Department of Energy

ISSN: 1050–6101 Vol. 9, Nos. 1–2, January 1998

Groups Coordinate Gene Sequencing

cDNA Workshops Reduce Duplicative Work, Increase Production

When several research teams around the world announced plans in the fall of 1996 for full-length cDNA (gene) sequencing, investigators felt that the highly beneficial infrastructure provided since 1994 by the international Integrated Molecular Analysis of Genome Expression (I.M.A.G.E.) consortium [*HGN* **6**(6), 3] should be extended to the challenges of complete cDNA sequencing. A subsequent workshop for I.M.A.G.E. participants was held in May 1997 in Gaithersburg, Maryland. The meeting was organized and chaired by Greg Lennon [then at Lawrence Livermore National Laboratory (LLNL) and now at Gene Logic Inc.] with Marvin Stodolsky coordinating for the meeting sponsor, the DOE Office of Biological and Environmental Research. Scientists attended from France, Germany, Italy, Japan, Sweden, the United Kingdom, and the United States.

Several workshop participants are members of the subgroup EURO-IMAGE, whose goals include generating and sequencing a master set of unique full-length cDNA clones (based on I.M.A.G.E. consortium resources) representing 3000 transcripts and 6 Mb of finished sequence. Other EURO-IMAGE goals are to obtain high-resolution and comparative functional mapping in human and model organisms of 1000 master-set genes and to develop the I.M.A.G.E. consortium database for easy access to an integrated view of the sequence, map, and expression data generated.

U.S. funding agencies represented at the workshop included DOE, NIH, and the recently established nonprofit Merck Genome Research Institute

(continued, p. 2)

A report of the Ninth Genome Sequencing and Analysis Conference, held in September 1997 at Hilton Head, South Carolina, is on the Web (*http://www.ornl.gov/meetings/hilhead9.html*). The report was written by Darrell Doyle (The Institute for Genomic Research).

Editors' Note

Because of progress in the Human Genome Project, investigators have more tools to address the biological questions that prompted its establishment and are finding countless other applications in which genomic resources can be used. The convergence of new strategies and such resources as cDNA and clone libraries (see article at left), databases, and automation and array technology has provided usable information since early in the project.

Future applications in environmental molecular toxicology will help lead to understanding the links between genetic variation and environmentally influenced diseases. The articles below and on p. 10 discuss some applications in environmental genomics.◊

In This Issue

Environmental Genomics

Genome Resources Help Scientists Explore Life-Environment Interactions

Most current tests for human exposure to environmental mutagens are only indicators of genetic damage and cannot predict adverse outcomes for individuals. In the following article, Anthony V. Carrano [Lawrence Livermore National Laboratory (LLNL)] explains that the future of genetic toxicology and mutation research lies in studying genes and individual genetic variation to reveal risk factors that make some people more susceptible to disease. The basic topic addressed by scientists who explore these issues, he notes, is the nature and consequence of genetic change or variation, with the ultimate purpose of predicting or preventing disease.

This article is excerpted from a talk by Carrano at the Human Genome Project session at the 1997 Society of Toxicology meeting in Cincinnati, Ohio. Other speakers were J. Craig Venter (The Institute for Genomic Research), Henry Wagner (Johns Hopkins University), and Richard Woychik (now at Case Western Reserve University).

The first 6 years of the Human Genome Project are behind us, and now resources can be applied to functional genomic studies, the genomics of the future. Functional genomics will be facilitated by completing the entire human genome sequence as soon as possible and, along the way, sequencing a significant portion of the mouse and other model-organism genomes. We want to determine all

(see Carrano, p. 8)

Human Genome Project

Sequencing *(from p. 1)*

[*HGN* **8**(3–4), 9]. Selected highlights follow of technical progress in complete cDNA sequencing, as reported at the workshop.

Highlights of Technical Progress

Attendees addressed a wide range of topics, including the status of cDNA sequencing projects, future targets, data- and clone-release policies, quality criteria and assessment, and mouse and other model organism cDNAs. Speakers projected that, with adequate support from funding agencies, participating laboratories could generate up to 15,000 full-length cDNA sequences in the following year. With average cDNA lengths of 2 kb, this represents some 30 Mb of total sequence.

Researchers have long recognized that expression of a single gene may culminate in the production of several different messenger RNA (mRNA) transcripts, depending both on the gene and the source tissue. Added to this biological complexity are the technical challenges of converting fragile mRNAs to the sturdier cDNAs. Standard methods involve use of poly dT as a primer on the 3′ poly A end of purified mRNAs, with reverse transcriptase enzymes of viral origin polymerizing the synthesis of a single-stranded DNA complement of the mRNA. These initial DNA transcripts often fail to extend to the 5′ end of longer mRNAs. With the use of more routine biochemistries, the single-stranded DNA is converted into duplex DNA and combined with a DNA vector to support its propagation and maintenance as a DNA clone. The double-stranded DNAs produced are much more stable and less susceptible to degradative processes than their single-stranded mRNA predecessors. However, because the initial reverse transcription is often shortened, cDNA libraries with abundant truncated products are the common result, particularly for the longer source mRNAs. Strategies devised for alleviating this truncation problem were described by Takao Isogai (Helix Research Institute, Japan), Nobuo Nomura (Kazusa DNA Research Institute, Japan), John Quackenbush [The Institute for Genomic Research (TIGR)], and M. Bento Soares (University of Iowa).

A protocol that takes advantage of the unusual nucleotide “cap” on the 5′ end of mRNAs requires that the first cDNA strand’s extension be long enough to protect the cap as a contingency for final cDNA clone production. Soares reported, however, that about one-third of cDNA transcripts begin within the mRNA, as contrasted with preferred starts at the mRNA’s 3′ end, thus giving rise to 3′ truncations. This problem can be alleviated

Support in the United States

Several U.S. funding agencies provide support for cDNA-related projects. To annotate developing chromosome maps, DOE in 1990 began dedicated support for improved cDNA library production, early EST generation by J. Craig Venter’s team at TIGR, physical mapping of cDNAs onto chromosomes, and database support. High-throughput correlations of cDNAs with the new BAC resources are also in progress. Sequencing of cDNAs corresponding to genes recognized during genomic sequencing is often a component of major chromosome sequencing projects.

The NIH National Human Genome Research Institute also is supporting research and development in cDNA library improvement and mouse cDNA library production. In the NIH-supported chromosome map development using radiation hybrid methodologies, about one-third of the markers are derived from ESTs. The source genes thus are mapped onto the chromosomes. Recently the NIH National Cancer Institute (NCI) began providing substantial support for cDNA library production and analysis in a major effort to identify cancer-related genes [see article, *HGN* **8**(3–4), 8]. This effort, called the Cancer Genome Anatomy Project, CGAP, was described at the workshop by Carol Dahl and Robert Strausberg of NCI.

The Merck Genome Research Institute supports programs to characterize cDNAs representing disease genes, including full-length cDNA cloning and sequencing. The mouse model’s usefulness for studying human diseases is being advanced with diverse collaborative support, including that from NIH for library construction and from DOE for I.M.A.G.E. efforts at LLNL to array mouse cDNA libraries. Washington University (St. Louis) generates mouse ESTs for the clone arrays with support from the Howard Hughes Medical Institute.

I.M.A.G.E. Consortium Aiding Rapid Development of Human Gene Catalog

In a cDNA library, the numerical representation of particular cDNAs varies over a thousandfold. The predominant members of cDNA libraries from all tissues are the genes for cellular maintenance functions. I.M.A.G.E.’s coordination minimizes the unwanted and expensive repetitive analysis of the already characterized cDNAs.

A single sequencing read of a few hundred cDNA bases usually is sufficient to serve as a distinguishing identifier (EST) of the predecessor mRNA. This approach was pioneered by J. Craig Venter, now director of The Institute for Genomic Research, which made public a major EST data release in June 1997. High-throughput production of ESTs from I.M.A.G.E. cDNAs has been funded predominantly by Merck & Co., with sequencing at the Washington University (St. Louis) Genome Center Human EST Project. ESTs are deposited in the public database dbEST, which supports queries on the similarities of ESTs and cDNAs to cDNA molecules whose analysis is just beginning.

The I.M.A.G.E. consortium manages the distribution of reference sets of cDNAs. In the United States, libraries of cDNA clones representing many different tissues are donated to I.M.A.G.E. at LLNL, where the clones are placed into reference arrays; replicas are provided to genome research centers and private-sector resource distributors. Over 3 million clone replicas have been sent to more than 1000 laboratories worldwide; the end users analyze the clones and return data on them. All data is entered into public databases, enabling researchers to compare it with their preliminary cDNA sequencing data and eliminate redundant efforts.

EST analyses of over 500,000 cDNA I.M.A.G.E. clones suggest that more than 50,000 of the estimated 60,000 to 80,000 human genes are represented. I.M.A.G.E. researchers at LLNL are providing “subtracting cDNA reagents” to aid the production of new cDNA libraries by Bento Soares (see article above) that preferentially contain clones not already represented in the current I.M.A.G.E. collection.

substantially by size fractionating the mRNAs and later selecting out the cDNA products with lengths equal to the size-sorted mRNA templates. Hans Lehrach (Max Planck Institut für Molekulare Genetik, Germany) related the value of massively parallel oligomer fingerprinting of cDNAs. This is an economical way to screen a library for novel and longer, potentially full-length cDNAs. Optimal candidate cDNAs chosen by the Lehrach team at the Resource Center of the German Genome Project are being sequenced in the laboratory of Annemarie Poustka (Deutsches Krebsforschungszentrum).

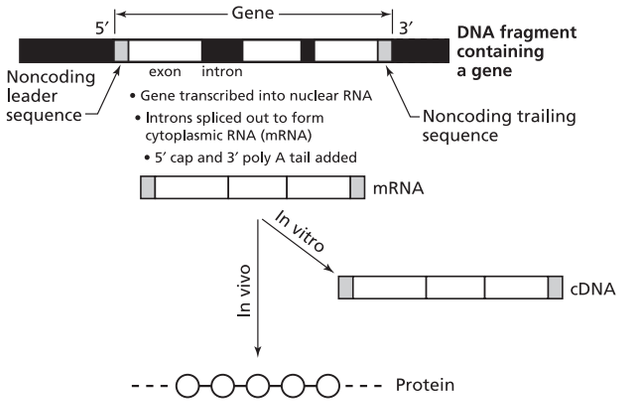
More than one sequencing read commonly is necessary to display the complete sequence for cDNAs longer than a few hundred bases. Strategies for economical full-length sequencing were discussed by Lennon and Richard Gibbs (Baylor College of Medicine). Sequence reads beyond 1000 bases now are being obtained with improvements to sequencing systems by Wilhelm Ansorge's team at the European Molecular Biology Laboratory. Ansorge suggested that, for cDNAs shorter than 2 kb, good coverage could be achieved by two overlapping reads on complementary strands.

Giuseppe Borsani (Telethon Institute of Genetics and Medicine) reported on the benefits of the easily manipulated *Drosophila* model for studies of development and function to reveal roles represented by human cDNAs.

Mark Boguski (National Center for Biotechnology Information) discussed the status of the dbEST cDNA sequence database and made recommendations for the evolution needed to meet the impending new demands of complete DNA sequencing. He observed that each group will have its own selection criteria and sequencing priorities, such as finding cancer genes, genes with *Drosophila* homologs, or genes that already have been mapped.

Boguski coined the expression "the slicing problem" to describe the difficulties in avoiding undesirable duplication and redundancy due to overlapping choice categories. A possible solution would be to establish a registration and tracking database modeled after the successful European Bioinformatics Institute's (EBI) RHAlloc-RHdb approach used in constructing the human transcript map. Patricia Rodriguez-Tomé (EBI) has accepted this responsibility. This data will include an investigator or center name and contact information,

Fast-Forwarding Through the Genome with cDNAs



Genes, which are housed in the DNA of the cell's nucleus, contain codes that essentially are recipes for tens of thousands of proteins. The code-containing regions of the gene (exons), however, are often separated by much noncoding DNA (introns). A cDNA molecule is a laboratory-made version of a gene that contains only its information-rich regions; these molecules provide a way for genome researchers to fast-forward through the genome to biologically important areas.

cDNA molecules are made using molecules of RNA (similar to DNA) obtained from living cells. In the cell, expression of the information from DNA into a protein first requires transcription of DNA into nuclear RNA molecules. These nuclear RNAs have noncoding regions that are processed out in the course of forming cytoplasmic RNAs (messenger RNAs). Because mRNAs are too fragile to withstand laboratory manipulations, scientists make sturdy double-stranded copies called complementary (or copy) DNAs, or cDNAs.

All DNA clones derived from a particular tissue constitute a library of clones representing the genes that were expressed when the source tissue was harvested. The analysis of libraries from many different tissues, obtained under a variety of physiological conditions, will be necessary to decipher the organ-specific patterns of gene expression.

Related Information

Baylor College of Medicine Human Genome Sequencing Center
http://gc.bcm.tmc.edu:8088/cgi-bin/seq/home

Caltech Genome Research Laboratory
http://www.tree.caltech.edu

Cancer Genome Anatomy Project
http://www.ncbi.nlm.nih.gov/ncicgap

dbEST Database
http://www.ncbi.nlm.nih.gov/dbEST/index.html

Deutsches Krebsforschungszentrum Division of Molecular Genome Analysis
http://www.dkfz-heidelberg.de/abt0840

Drosophila Related Expressed Sequences
http://www.tigem.it/LOCAL/drosophila/dros.html

European Molecular Biology Laboratory Ansorge Group Research Report
http://www.embl-heidelberg.de/ExternalInfo/ScientificProgrammes/Ansorge.html

German Human Genome Project Resource Center
http://www.rzpd.de

Helix Research Institute
http://www.hri.co.jp

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

Human Transcript Map
http://www.ncbi.nlm.nih.gov/science96
http://www.ncbi.nlm.nih.gov/Schuler/Papers/ESTtransmap

I.M.A.G.E.
http://bbrp.llnl.gov/image/image.html

Kazusa DNA Research Institute
http://www.kazusa.or.jp/KDRI/kazusa_home-e.html

Mouse cDNA Resources
http://www-bio.llnl.gov/bbrp/image/muslib_info.html

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

The Institute for Genomic Research
http://www.tigr.org

WashU–HHMI Mouse EST Project
http://genome.wustl.edu/est/mouse_esthmpg.htm

WashU–Merck Human EST Project
http://genome.wustl.edu:80/est/esthmpg.html

identifiers for the physical cDNA clones being sequenced and associated EST accession numbers, and sequencing status. When participants registered a clone that they intended to sequence, the database would detect and report overlaps with clones selected by other groups.

Attendees agreed that the I.M.A.G.E. consortium should convene every 6 months to maintain necessary coordination and efficiency. A subsequent meeting, organized by Quackenbush, was held in September 1997 in conjunction with the Ninth International Genome Sequencing and Analysis Conference in Hilton Head, South Carolina. Washington University scientists will organize the next meeting, tentatively planned to concur with the May 1998 Human Genome Workshop at Cold Spring Harbor Laboratory. [Meeting: *http://www.ornl.gov/meetings/wccs/index.html*] *[Marvin Stodolsky, DOE Human Genome Program, and Denise Casey, HGMIS]* ◊

JGI Sequencing Facility Progressing

Later in 1998, the DOE Joint Genome Institute (JGI) Production Sequencing Facility is expected to occupy 56,600 square feet of leased laboratory and office space now being renovated in Walnut Creek, California.

JGI was formed Jan. 1, 1997, by the DOE Human Genome Program to integrate high-throughput DNA sequencing among its three major human genome laboratories. DOE will invest about $150 million in JGI over the next 5 years. [JGI: *http://www.jgi.doe.gov*]◊

GDB to Cease Operations

Genome Database (GDB), which provides human gene mapping data to human genetics researchers from its base at Johns Hopkins School of Medicine, will cease operations by July 31, 1998. GDB's principal funder, the DOE Office of Energy Research, is discontinuing support to focus its informatics resources on the sequencing phase of the Human Genome Project. The servers for continued access to the current copy of GDB will be maintained in Oak Ridge National Laboratory's Computational Bioscience Section (*http://compbio.ornl.gov*), headed by Ed Uberbacher. This will be a static version of the database, and no further development is now contemplated. (Full GDB termination statement: *http://www.gdb.org/shutdown/notice.html*) ◊

FAQs

Contributed by

Daniel Drell
DOE Human Genome Program
daniel.drell@oer.doe.gov

Q. What is the value of the Human Genome Project?

A. My basic view is that the project will reap fantastic benefits for humankind, some that we can anticipate and others that will surprise us. Greater knowledge about the human genome will help us better understand the many diseases and heritable conditions that affect humans. Disease genes get all the attention, but much more profound is the need to understand normal biological functions. From this understanding will come insights into how to prevent diseases rather than rely on treatment after they start. The most cost-effective disease prevention ever invented is the vaccine. For the cost of smallpox vaccine or a tetanus shot, a life of incalculable value can be safeguarded *at no further expense*. Disease prevention—the gold standard for medicine—also represents the promise of genetics.

Another benefit will come from understanding genetic similarities between mammals and humans. There isn't that much difference between human biology and cattle or pork biology (or mouse biology for that matter). What we learn about human genetics will help us to raise healthier, more productive, disease-resistant farm animals that might, through wise and careful genetic engineering, produce drugs of value to us. (Additional benefits and applications to various areas of research are given in the sidebar on p. 5.)

Q: What concerns have been expressed about the Human Genome Project?

Genetic information can be used to make predictions about a person's medical future, and possible invasions of privacy by employers or insurers can be worrisome. A very serious possibility is that misunderstandings about the limits of genetic information may lead to discrimination, and people may not understand that having a predisposing gene mutation is not the same as being condemned to get the disease. Discrimination can happen because it often is less expensive to adopt a blanket policy excluding people with predisposing genes.

Educating judges and others in the court system about the nature and implications of genetics, including its limitations, is very important. Most judges are not and never have been scientists, so they are inhibited and uncertain when scientific matter is entered as evidence in a trial.

One of the most difficult challenges for geneticists will be to study multigenic or multifactorial conditions (not all are even diseases) and those in which genes and environmental factors interact. Diabetes mellitus is a good example of a complex disease. We know that quite a few genes (at least a dozen) are influential in determining which individuals develop diabetes. We also know that in genetically identical twins, when one twin has diabetes, the other twin has only a 1 in 3 chance of developing the condition. So there is a 2 in 3 chance that the other twin will *never* get diabetes, even though the twins have identical genes. So, in this case, genetics can account for about one-third of the causation, and external or internal environmental factors account for two-thirds. Environmental factors involved in susceptibility remain to be elucidated.

Such complex diseases are much more common than single-gene conditions including cystic fibrosis (CF), sickle-cell disease, and Huntington's Disease. But even understanding a "simple," single-gene disorder presents many challenges. For example, more than 600 alternative forms (alleles) of the CF gene have been identified, but their clinical effects are not yet known. Some alleles may give rise to the full-blown, fatal disease, whereas others apparently have little or no effect on the individual. Commercial gene tests available now present problems for doctors and patients in understanding the implications of a positive test, particularly when used prenatally. More research will be needed to determine the effects of each variant allele.

More than 250 alleles are associated with two genes called *BRCA*1 and *BRCA*2, which can cause a rare, inherited form of breast cancer. Is it appropriate to discriminate against every woman who bears a mutation in these genes? What use should be made of information that is not certain? Society is beginning to address these questions and many other implications related to the increased availability of genetic information.

Hardest of all, possibly, are questions surrounding the role that genes may play in human behavior. Which genes are they, how much do they affect behavior, and with what consequences? How would society use information about genes that affect behavior? If certain behaviors that may be influenced by genes are socially dangerous, what should we do about people who have these predisposing genes? Are those persons responsible for their behavior if brought before a judge and accused of criminal acts? If environmental factors such as drugs and alcohol are involved, where does responsibility reside?

These and other questions need answers that will come from more research and public discussions. The Ethical, Legal, and Social Implications component of the DOE Human Genome Program has been striving to find such answers since the beginning of the Human Genome Project.◊

Anticipated Benefits of Genome Research

Predictions of biology as "the science of the 21st century" have been made by observers as diverse as Microsoft chairman Bill Gates and U.S. President Bill Clinton. Already revolutionizing biology, genome research has spawned a burgeoning biotechnology industry and is providing a vital thrust to the increasing productivity and pervasiveness of the life sciences.

Technology and resources promoted by the Human Genome Project already have had profound impacts on biomedical research and promise to revolutionize biological research and clinical medicine. Increasingly detailed genome maps have aided researchers seeking genes associated with dozens of genetic conditions, including myotonic dystrophy, fragile X syndrome, neurofibromatosis types 1 and 2, inherited colon cancer, Alzheimer's disease, and familial breast cancer.

Current and potential applications of genome research will address national needs in molecular medicine, waste control and environmental cleanup, biotechnology, energy sources, and risk assessment.

Molecular Medicine

On the horizon is a new era of molecular medicine characterized less by treating symptoms and more by looking to the most fundamental causes of disease. Rapid and more specific diagnostic tests will make possible earlier treatment of countless maladies. Medical researchers also will be able to devise novel therapeutic regimens based on new classes of drugs, immunotherapy techniques, avoidance of environmental conditions that may trigger disease, and possible augmentation or even replacement of defective genes through gene therapy.

Microbial Genomes

In 1994, taking advantage of new capabilities developed by the genome project, DOE formulated the Microbial Genome Initiative to sequence the genomes of bacteria useful in the areas of energy production, environmental remediation, toxic waste reduction, and industrial processing. As a result of this initiative, six microbes that live under extreme conditions of temperature and pressure had been sequenced completely as of August 1997. Structural studies are under way to learn what is unique about the proteins of these organisms—the ultimate aim being to use the microbes and their enzymes for such practical purposes as waste control and environmental cleanup.

Biotechnology

The potential for commercial development presents U.S. industry with a wealth of opportunities. Sales of biotechnology products are projected to exceed $20 billion by the year 2000. The project already has stimulated significant investment by large corporations and prompted the creation of new biotechnology companies hoping to capitalize on the far-reaching implications of its research.

Energy Sources

Biotechnology, significantly fueled by insights reaped from the genome project, will play a significant role in improving the use of fossil-based resources. Increased energy demands, projected over the next 50 years, require strategies to circumvent the many problems associated with today's dominant energy technologies. Biotechnology promises to help address these needs by providing cleaner means for the bioconversion of raw materials to refined products. In addition, there is the possibility of developing entirely new biomass-based energy sources. Having the genomic sequence of the methane-producing microorganism *Methanococcus jannaschii*, for example, will enable researchers to explore the process of methanogenesis in more detail and could lead to cheaper production of fuel-grade methane.

Risk Assessment

Understanding the human genome will have an enormous impact on the ability to assess risks posed to individuals by environmental exposure to toxic agents. Scientists know that genetic differences make some people more susceptible—and others more resistant—to such agents. Far more work must be done to determine the genetic basis of such variability. This knowledge will directly address DOE's long-term mission to understand the effects of low-level exposures to radiation and other energy-related agents, especially in terms of cancer risk. *[Reprinted from the* DOE Human Genome Program Report, *in press.]*

Team Completes HGP Milestone: Human Chromosome 7 Map

In July 1997, researchers reported the construction of a high-resolution physical map for human chromosome 7 [*Genome Research* **7**(7), 673–92]. With an average STS spacing of about 79 kb, the new map exceeds the Human Genome Project goal and provides a framework for future chromosome 7 sequencing. The map, which covers virtually all the 170-Mb chromosome, was built using a YAC-based, STS-content approach and includes 22 YAC contigs integrated with the genetic, radiation hybrid, and cytogenetic maps generated for that chromosome.

Collaborators on the project included a team led by Eric Green at the NIH National Human Genome Research Institute and researchers at the Washington University School of Medicine (St. Louis), Généthon (France), and the University of Washington. [Information: *http://www.nhgri.nih.gov/DIR/GTB/CHR7* and *http://www.cshl.org/gr*]◊

Human Genome Project

Large-Insert Cloning Vectors Aid Function Studies

Collections of human DNA fragments are maintained for research purposes as clones in bacterial host cells. However, for unknown reasons, some regions of the human genome appear to be unclonable or unstable in bacteria. The team led by Jean-Michel Vos [University of North Carolina at Chapel Hill (UNCCH)] has developed a system using episomes (extrachromosomal, autonomously replicating DNA) that maintains large DNA fragments in human cells. This human artificial episomal chromosomal (HAEC) system may prove useful for coverage of these especially difficult regions. In the broader biomedical community, the HAEC system also shows promise for use in functional genomics and gene therapy. In the article below, Vos discusses some recent improvements to the HAEC system and its application to mapping, sequencing, and functionally studying human and mouse DNA.

Mapping and sequencing the human genome and model organisms are only the first steps in determining the function of various genetic units critical for gene regulation, DNA replication, chromatin packaging, chromosomal stability, and chromatid segregation. Such studies will require the ability to transfer and manipulate entire functional units into mammalian cells.

The development of extrachromosomal large-capacity cloning vectors for mammalian cells represents a powerful tool for functional genomic study. The UNCCH laboratory has developed the HAEC system to establish large DNA fragments as episomes in human cells, using the latent replication elements from the human herpes Epstein-Barr virus (EBV). A first-generation episomal vector based on the latent origin of replication oriP and its transactivator EBNA-1 from EBV was used to establish and maintain up to 350 kb of circular DNA in human cells. Such a system allowed the generation of a random clone library covering 10% to 20% of the human genome as extrachromosomal self-replicating episomes in human cells.

Current methods of transferring DNA into mammalian cells are particularly laborious and ineffective. To improve delivery of large genomic inserts into human cells, Vos's team developed a second-generation HAEC system based on engineering EBV. Such a mini-EBV vector carries the minimal *cis* elements for replicating and packaging the cloned DNA into infectious virus particles. In combination with a helper cell line, it was shown that up to 180 kb of non-EBV DNA can be packaged and delivered into target human cells as self-replicating HAECs.

Functional gene delivery by this mini-EBV HAEC system was demonstrated by introducing specific human genes coding for the hypoxanthine guanine phosphorybosyl transferase and Fanconi's anemia (FA) group C in lymphoblastoid cells from people suffering from Lesh-Nyhan and FA, respectively. In vitro correction of these defects supported the potential therapeutic usefulness of such a virus-based HAEC delivery system. By analogy to the bacteriophage lambda cosmid packaging system, the mini-EBV/HAEC technology represents an alternative to current DNA transfection techniques for studying and manipulating (within human cells) large human and model-organism chromosomal sections as nonintegrated and easily recoverable DNA clones.

The stable shuttling of any large human insert previously cloned in bacteria or yeast would be beneficial to future functional genomic analysis and interpretation of sequencing data. To establish specific clones isolated from current YAC and BAC-PAC large-insert libraries in human cells, the HAEC technology was upgraded recently by engineering the third-generation hybrid BAC-HAEC and retrofitting "POP-IN" vectors, respectively. The POP-IN HAEC system converts any bacterial-based P1, BAC, or PAC into a HAEC by cre-loxP recombination, while the hybrid BAC HAEC system is based on subcloning large YAC inserts into the BAC-HAEC vector. Such vectors can shuttle DNA inserts of up to 260 kb from bacteria or yeast into human cells using bacteria as an intermediate host. In addition, the extrachromosomal HAEC DNA isolated from human cells can be easily transferred back into bacteria for further analysis.

A table on large DNA cloning systems and a bibliography to accompany this article can be found on the Web (*http://www.ornl.gov/hgmis/publicat/hgn/v9n1/haec.html*).

The usefulness of these systems was illustrated by successfully shuttling expressed human housekeeping and tissue-specific genes as 100- to 200-kb genomic HAECs into human cells. Current efforts are focused on upgrading these respective technologies for high-throughput functional assaying in human cells of large human chromosomal regions.

The development of a technology analogous to the HAEC system for mouse cells, which are widely used as a surrogate system for studying human diseases, would enable additional levels of manipulation and analysis of the human genome. Mouse cells, however, have proven nonpermissive to the oriP EBNA-1 system. Through a strategy based on the transfer of HAECs as minichromosomes using microcell fusion, circular molecules carrying 100 to 200 kb of human DNA were stably shuttled into the mouse cells at low copy number. The establishment of a first-generation mouse artificial episomal chromosomes (MAEC) system will facilitate future functional and evolutionary studies of the human genome in a murine genetic background. Current efforts focus on generating MAEC-based transgenic mice for future developmental, genetic, and therapeutic studies. *[Jean-Michel H. Vos, University of North Carolina at Chapel Hill*, vos@med.unc.edu, http://www.med.unc.edu/lccc/voslab*]* ◊

The Gene Letter on Web

The latest issue of *The Gene Letter* is on the Web (*http://www.geneletter.org*). Designed to inform consumers and professionals about advances in genetics and to encourage discourse about emerging policy dilemmas, the newsletter is supported by a DOE grant to the Shriver Center.◊

Complete *E. coli* Genome Sequence in Public Databases

Meets HGP Goal, Will Aid Understanding of Toxic Strains

In September 1997, a team of scientists led by Frederick Blattner (University of Wisconsin, Madison) reported completing the sequence of the 4.6-Mb *Escherichia coli* K-12 genome. The paper published in *Science* (**277**, 1453–62) represents an analysis of data collected by more than 259 people over the project's 6-year duration.

Obtaining the complete DNA sequence of the *E. coli* genome has been a goal of the Human Genome Project, both to help develop sequencing and gene-finding technology and to facilitate studies on gene function and organization. More than 4200 *E. coli* genes have been identified, although the functions of over one-third of them remain unknown. Because of similarities found in genes across species, this work provides a valuable starting point for identifying and understanding genes in other organisms, including humans.

"Determination of the complete inventory of the genes of organisms is one of the major goals of biology—analogous to development of the periodic table of the elements in chemistry," said Blattner. "Once they are all known and relationships between them become evident, a classification system for understanding the basic functions of life can be erected."

For more than 70 years *E. coli*, a natural inhabitant of the lower intestinal tract of animals, has been one of the most studied organisms for scientists exploring fundamental processes in biochemistry, genetics, and physiology. In recent years it has become the workhorse of biotechnology and serves as a living factory for producing human insulin and other medicines.

The strain used in this study does not cause disease, but related toxic strains have been associated with an increasing number of human food poisonings. The new data provide a reference strain against which scientists are already comparing the genes of pathogenic strains.

At the Ninth International Genome Sequencing and Analysis Conference (Hilton Head, South Carolina) in September 1997, Blattner described his group's study of sequences from the genome of *E. coli* 0157:H7, the strain responsible for many recent outbreaks of fast-food poisoning associated with hemolytic uremic syndrome. A sample of 20,000 random sequences compared against the K-12 genome and the public databases revealed about 1 Mb of DNA not present in K-12. Homology searches revealed matches to virulence genes from a variety of species. These and future studies will lay the groundwork for developing drugs to help prevent or to treat diseases associated with toxic *E. coli* strains.◊

E. coli Web Site

A revised and updated *E. coli* sequence is now accessible at GenBank, NCBI (*http://www.ncbi.Nlm.nih.gov/cgi-bin/Entrez/framik?db=Genome&gi=115*), and the *E. coli* Genome Project Web site (*http://www.genetics.wisc.edu*). Annotation is ongoing, and comments to the authors are appreciated.

E. coli Genome Navigator is a Web-based visual interactive display and query resource for the complete *E. coli* genome. It graphically displays coding sequences and other genome elements by functional category and allows users to browse the genome and query external data sources about any of these elements (*http://www.Mpimg-berlin-dahlem.mpg.de/~andy/GN/E.coli*).

Completed Genomes

Organism	Genome Size (Mb)	Estimated Genes
Saccharomyces cerevisiae	12.1	6034
Escherichia coli	4.6	4288
Bacillus subtilus	4.2	~4000
Synechocystis sp.	3.6	3168
**Archaeoglobus fulgidus*	2.2	2471
**Pyrobaculum aerophilum*	2.2	N.A.
Haemophilus influenzae	1.8	1740
**Methanobacterium thermoautotrophicum*	1.8	1855
Helicobacter pylori	1.7	1590
**Methanococcus jannaschii*	1.7	1692
**Aquifex aolicus*	1.5	1508
Borrelia burgdorferi	1.3	863
Treponema pallidum	1.1	1234
Mycoplasma pneumoniae	0.8	677
**Mycoplasma genitalium*	0.6	470

*Supported by DOE. Also see genome sequencing projects on Web (*http://www.mcs.anl.gov/home/gaasterl/genomes.html* and *http://www.tigr.org/tdb/mdb/mdb.html*).

An article about the Lyme disease bacterium, *Borrelia burgdorferi*, appears on p. 16.

Web Sites for Microbial Sequence Data

The complete genome sequence of *Methanobacterium thermoautotrophicum* has been updated on the Genome Therapeutics Corporation (GTC) Web site, including annotation, comparative analysis, and maps (*http://www.Cric.com/htdocs/sequences/methanobacter/abstract.html*). GTC has also released preliminary data from the first full shotgun assembly of the 4.1-Mb *Clostridium acetobutylicum* genome, including BLASTP2 summaries for probable coding open reading frames (*http://www.Cric.Com/htdocs/sequences/clostridium/clospage.html*). This work is supported by the DOE Microbial Genome Program. [Contact: Douglas Smith (617/893-5007, Fax: -9535, *doug.smith@genomecorp.com*)] ◊

Environmental Genomics

Carrano *(from p. 1)*

transcript structures and gene-expression patterns in the human genome; ultimately, we want to understand the phenotypes resulting from mutations in every one of the open reading frames. Many groups have begun working in this area even before the genome project is completed, and much work reported at recent genome meetings is pointed toward functional genomics.

Finding Genes

As of March 1997, about 7000 genes had been identified, and around 5000 had been located and mapped to human chromosomes. Original estimates of total gene number were around 100,000, and although more recent estimates are as low as 50,000, I believe 75,000 to 80,000 genes will prove a more accurate figure. This translates to an average gene density of about 1 every 30 kb. Chromosome 19, which our laboratory has mapped and studied extensively since 1987 because it was the richest in G-C content, appears to have a higher gene density than many other chromosomes.

Positional cloning, the standard approach to disease-gene finding, requires many resources. The Human Genome Project originally was set up to create such resources. In positional cloning, researchers start with families having an inherited disease, then develop genetic (polymorphic) markers to localize disease traits on specific chromosomes with further, finer delineation to a region. After that, a set of clones—either YACs or BACs—is needed to get a contiguous DNA region that would localize or have candidate genes of interest in it. Then perhaps a higher-resolution set of clones such as cosmids would allow scientists to identify cDNA transcripts and ultimately use sequence information to identify the mutation associated with the phenotype. Many of these clones can be used now by the research community, and the cDNAs are available.

It is likely that these steps will be unnecessary in another 5 to 10 years because we will have the sequence for both normal and affected individuals. In some cases, we won't know the disease in individuals, but we can look at sequence information to predict the disease or function for the disease-associated gene. That's the goal we are shooting at—to bypass steps in mapping and sequencing.

cDNA Libraries

Collections of cDNA molecules, which represent the coding (gene) sequences of the genome, offer researchers a way to bypass the millions of bases of non-coding DNA to obtain the sequences having the greatest biological significance. The I.M.A.G.E. consortium has the largest collection of publicly available cDNAs, which are used throughout the world. I.M.A.G.E. was started as a collaboration by four individuals: Greg Lennon (then at LLNL), Mihael Polymeropolus (NIH), Bento Soares (then at Columbia University), and Charles Auffray (CNRS, France).

I.M.A.G.E. now has more than 56 human cDNA tissue libraries that are continually being expanded. They are collected at LLNL, where they are also arrayed and characterized, and information on them is sent to the Genome Database (GDB). Of the libraries in the I.M.A.G.E. collection, more than 500,000 clones have been arrayed. Using clustering algorithms to ask how many unique clusters of ESTs are present in GDB, about 60,000 have been identified.

These clones are sent from distribution centers to scientists throughout the world. At the same time, they are sent to Washington University scientists who determine the sequences on the 5′ and 3′ ends and enter the information in the dbEST database at the National Center for Biotechnology Information.

Recently, the I.M.A.G.E. group has been adding mouse cDNA libraries, many of which are normalized. For the mouse, cDNA libraries have been created from staged times during development to help biologists understand when certain critical genes are expressed (I.M.A.G.E.: *http://bbrp.llnl.gov/image/image.html*).

Some interesting uses of these resources for studying diseases with a toxicological or genetic component can be illustrated by our group's work a few years ago in collaboration with scientists studying the human *CYP* gene family. More than 60 genes code for the cytochrome P450 enzyme superfamily, which is involved in the metabolism of almost all chemicals to which we are exposed in our internal and external environments.

The *CYP2A* gene family codes for enzymes that are the first actors in a long pathway to detoxify and excrete xenobiotic chemicals (i.e., synthetic compounds foreign to living systems, such as drugs and insecticides). Available *CYP2A* family cDNAs were probed against the Livermore cosmid libraries developed as part of the National Laboratory Gene Library Project. Through a set of automated technologies, the cDNAs were built up very quickly into a contig spanning 350 kb. In that stretch are 11 genes from the *CYP2A* family, averaging 1 every 30 kb.

Looking specifically at the *CYP2A6* genes in 182 individuals of various ethnic backgrounds, investigators found two sequence variants. Variant 1 had a single amino-acid change, and the heterozygous state resulted in a reduction of activity to about 50% to 80%. Responsible for coumarin metabolism, *CYP2A6* is called coumarin hydroxylase. Coumarin, a drug used in the formulation of blood anticoagulants, also has been suggested and evaluated for the treatment of lymphoedema. Heterozygotes for this variant have a 50% to 80% reduction in activity for this particular enzyme. Homozygotes range from no activity up to 50% activity. Highly variable activity is associated with these heterozygotes; in fact, the pharmacological basis of this is quite well understood and determines drug treatment and outcomes. A second variant has several differences in its sequence. There are no known homozygotes for this variant. As far as we now know, it does not produce a functional protein.

The goal of these studies—to identify links between disorders and individual variabilities in DNA sequences ("polymorphisms")—will help researchers and clinicians identify people who may be predisposed to developing disorders such as cancer. (The converse also is probably true—some DNA variations may offer protection from these diseases.) With this knowledge, individuals can make informed decisions about aspects of their lifestyles (e.g., diet and

http://www.llnl.gov.llnl/llnl_org/bios/carrano.bio.html

occupation) that may help prevent or delay the onset of some diseases.

DNA Repair Genes

DNA repair genes, discovered by Richard Setlow and other DOE-funded researchers in the mid-1960s, also are important to toxicology and mutation research. LLNL has focused its research on chromosome 19 not only because of its high G-C content but because of the DNA repair enzymes encoded by genes on this chromosome. LLNL has over a 20-year history of DNA repair studies, beginning with Larry Thompson and others, that ties into the DOE goal of understanding mutations caused by exposure to ionizing radiation and other environmental pollutants associated with energy production and use.

Several genes involved in different DNA repair pathways such as base or nucleotide excision repair have been cloned, mapped, and the cDNAs isolated. In many cases, the genomic sequences have been determined at LLNL. Functions that have been ascribed to several of the DNA repair genes include helicase, endonuclease, polymerase, ligase, and others not fully understood but somehow involved in recombination and repair processes.

These known human genes have been found to be very similar to genes in yeast and even, in some cases, bacterial genes. Since we understand the DNA repair and metabolism systems much better in yeast and bacteria than in humans, we can begin with data from these model systems to understand human systems.

To take the DNA repair work a step farther, one of the DNA repair genes—*ERCC2*—is interesting because it has variable expression in individuals and is associated with three different diseases. The most severe of these diseases is xeroderma pigmentosum type D, which is characterized by high sensitivity to uv light, high cancer incidence, and some neurological disorders. Another of the diseases, Cockayne's syndrome, is characterized by slight photosensitivity and severe neurological defects but does not have a high cancer incidence.

The third *ERCC2* disease, trichothiodystrophy, presents a defect in metabolism that causes brittleness of hair, pale skin, slight photosensitivity in about 50% of patients, and some minor neurological defects. Severity can vary. Chris Webber at Livermore and collaborators have found mutations associated with these diseases throughout the entire gene but none can be associated absolutely with the phenotype. *ERCC2* is one protein in a multiprotein complex and is part of a transcription factor complex necessary for gene transcription into mRNA. Protein folding and interaction in the complex may affect the form and phenotype produced.

Does DNA variation mean anything? Can we use DNA variation, as we've seen in the previous examples, as an indicator of disease or susceptibility? LLNL scientists Harvey Mohrenweiser and Richard Shen have measured variability in the coding sequences of some DNA repair genes in people. Some variants in the DNA sequence occur at high frequency and lead to nonconservative amino-acid changes in the protein. The next step is to link this information to populations that have increased susceptibility to disease.

Migraine

Migraine is another example of a disease with strong genetic and environmental interactions, although the environmental component is not well understood. About 24% of females and 12% of males are affected by migraine, and some healthcare researchers believe it is the most common reason for seeking outpatient care in the United States. A subset of migraines, called hemiplegic migraine, is often preceded by an aura. Hemiplegic migraine has also been associated with another disease called episodic ataxia. Both of these are now known to be associated with the same calcium channel genes.

In a collaboration with researchers in the Netherlands, Mohrenweiser and others at Livermore were able to link the disorder with changes in a region of chromosome 19. In this case, resources available from the genome project included a clone collection, and it was suspected that a candidate calcium channel gene in a 1- to 3-Mb region of chromosome 19 was associated with familial hemiplegic migraine. With techniques to trap the cDNAs in that large region, the team found the alpha-1 subunit of the candidate gene. The alpha-1 subunit is present in four copies; it is four subunits long, and each subunit is composed of six alpha helical turns spanning the membrane, along with a pore section that controls and forms a channel in the membrane.

Family members were identified, the gene was sequenced for the alpha-1 subunit, and mutations were found. Several mutations are associated with the alpha helical units for familial hemiplegic migraine as well as with the pore unit. Mutations associated with episodic ataxia in the same region correlate closely with the condition in these families. We know that some calcium channel genes are involved in 5-hydroxytryptamine release, and their pharmacological role in migraine would be very interesting to pursue.

More important, certain environmental factors might help trigger migraine onset. Environmental factors that influence the gene constitute an area ripe for further studies.

Future Approaches

As mentioned, many novel human genes are being identified through comparisons with sequences from other species. In mouse-human comparisons, researchers use gene probes that will identify both human and mouse clones. For the DNA repair gene *XRCC1*, there are 17 exons in human and the same number of highly conserved exons in the mouse. Interestingly, some noncoding regions are also conserved. What function do these highly conserved, noncoding regions perform? Are they actually putative regulatory regions?

To understand the functions of gene sequences, researchers can use a suite of techniques including the standard biochemical approaches, structural approaches, or animal models such as knockout technologies in mice. In knockouts, a particular gene is disabled in the mouse, and any resulting changes in the animal are noted.

(continued next page)

Environmental Genomics

Technologies

The work discussed here could not have been done without automation in some technologies. Robotic systems are available that can put tens of thousands of DNA spots or DNA clones on a high-density filter, and hundreds of filters can be created automatically in a day. Using these filters is the fastest way to find a gene. The filters can be mailed anywhere in the world, and receiving laboratories can probe the filters with their own DNA sequences. Coordinates identifying the probe-positive clones containing a gene of interest can be sent back to the originating laboratory, and the genes can be pulled from the collections of clones in the refrigerator.

Sequencing will change dramatically. With the advent of new technologies, especially those based on microchannel or capillary sequencing, we will increase the rate at least 10- to 30-fold over the next 3 to 4 years. With totally automated refilling and minimal human intervention, it should be possible to produce 500 bases of sequence per lane every 2 hours.

As a final example, everybody wants to do the polymerase chain reaction (PCR), and forensic toxicologists want to do PCR in the field. A portable PCR machine was developed at LLNL with Department of Defense funding as a suitcase device that can do PCR in 20 minutes for a single gene or complex set of genes. This machine, based on a new silicon chamber device developed at LLNL, is now being commercialized.

So these are some technologies that are here now or coming on the scene in the next few months. It would be wonderful to see such resources and information in the hands of many more investigators. They have not yet taken full advantage of these opportunities. The time is past due for the mutation research community to use resources and capabilities produced by genome research.◊

Human Genome news

This newsletter is intended to facilitate communication, help prevent duplication of research effort, and inform persons interested in genome research. Suggestions are invited.

Human Genome Management Information System
Oak Ridge National Laboratory
1060 Commerce Park, MS 6480
Oak Ridge,TN 37830
423/576-6669, Fax: /574-9888
http://www.ornl.gov/hgmis

Managing Editor
Betty K. Mansfield
bkq@ornl.gov

Production Assistants
Murray Browne
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

Health Effects and Life Sciences Research Division, OBER
Marvin E. Frazier, Director
http://www.er.doe.gov/production/ober/hug_top.html

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

New HGP Spinoff Program to Study Genes for Environmental Risk

Although all people are equipped with the same basic set of genes and DNA regulatory regions, close comparisons among individuals reveal their true diversity, with a variation occurring about once every 500 to 1000 bp along the 3-billion–bp human genome. Some of these changes account for such obvious traits as the shape of the nose, height, and hair color, but some gene variations produce no apparent phenotypic differences. Other DNA variations affecting fundamental biological processes and gene-to-gene and environmental interactions can result in a wide range of susceptibility to diseases.

In a new program initiated by the NIH National Institute of Environmental Health Sciences (NIEHS), researchers in the Environmental Genome Project (EGP) will investigate how some genetic differences, called polymorphisms (for "many forms"), influence susceptibility to environmental exposure. Identifying and studying genes and other genetic regions that affect individual responses to environmental toxins can help scientists better predict health risks and develop environmental policies to protect the most vulnerable population groups.

EGP plans to systematically identify sequence variations in about 200 genes associated with environmental disease susceptibility in the U.S. population, develop a central database of polymorphisms for these genes, and foster studies of gene-environment interaction in disease susceptibility. NIEHS expects the multiyear effort to cost at least $60 million.

For more information, see *Science* **278**, 569–70 (October 24, 1997) and the Web site (*http://www.niehs.nih.gov/dirosd/policy/egp/home.htm*)

EGP's objective will be to sequence coding and regulatory regions of each of 200 genes in 1000 individuals. Susceptibility genes, to be chosen through a peer-reviewed process, are expected to include five broad classes: genes controlling toxicant distribution and metabolism, genes for nucleic acid–repair pathways, genes for the cell cycle control system, cell death and differentiation genes, and genes for signal transduction systems controlling gene expression in the other classes. A central database of polymorphisms found in these genes will be developed to support both functional studies of variants and population-based studies of disease risk. The latter studies are central to identifying specific alleles as well as the environmental exposures that cause disease.

Working with genetically susceptible subgroups will enable researchers to

(see Spinoff, next page)

Genetic Privacy Task Force Hears Testimony

The Task Force on Health Records and Genetic Privacy, established in 1997 by the Commerce Committee of the U.S. House of Representatives, held its first meeting on July 22 in Washington, D.C. Cochaired by Reps. Cliff Stearns (R-FL) and Gene Green (D-TX), the eight-member bipartisan task force is considering the following issues as a prelude to establishing policy:

- Effects on Americans of the many genetic tests created as a result of the Human Genome Project.
- Impacts of genetic research on medical practices; disease treatment, cure, and prevention; and test availability.
- Potential misuses of a person's genetic and other information by employers and health insurers.
- Risks to scientific research of addressing this potential misuse.
- Legislative steps that could maximize the success of future research.

The July meeting took the format of a congressional hearing and lasted for 3.5 hours, with testimony by 12 witnesses. These witnesses and members of several panels emphasized the importance of addressing genetic privacy within the context of all carefully protected medical records. Some argued for more awareness of the impact of policy decisions on research and against legislation that permits retrospective destruction of DNA samples or records.

One panel consisted of representatives of the health-insurance and pharmaceutical and biotechnological research industries. These speakers generally supported legislation to protect individuals from genetic discrimination and opposed efforts to separate genetic information from other medical records. They also favored broad medical-record confidentiality that protects medical research, has different standards for nonidentifiable data, and contains appropriate informed consent.

A second meeting of the task force is planned for early in 1998. [Information contact: Rep. Stearns' office (202/225-5744, Fax: -3973)] ◊

Spinoff *(from p. 10)*

identify more precisely the environmental agents with roles in disease causation as well as the true risks of exposure. These results could lead to public health programs for protecting susceptible populations and for targeting screening to groups at greater risk of disease. The project, which will use technology produced in part through the Human Genome Project, will also foster development of new high-throughput technology for a broader application of molecular genetics in epidemiology and environmental exposure.

In a presentation to the House Appropriations Committee in July 1997, NIEHS Director Kenneth Olden stated that it is time to take advantage of the tools developed and skills learned in 30 years of environmental research—to break with the past and lead in bold, new initiatives. In presenting EGP as part of a new "vision" for environmental health research, Olden also suggested a survey of chemicals taken up by humans, using blood and urine tests to determine American population exposure to specific agents; further development and approval of customized mice and other quick-throughput methods to screen chemicals and drugs; and a study of chemical mixtures to explore their effects on people. Olden hopes to launch EGP in 1998 with $10 million. [See related article by Anthony Carrano, beginning on p. 1.]◊

CME Genetics Conference for Physicians

The debut conference in the Continuing Medical Education (CME) series on genetics of the National Center for Genome Resources (NCGR) drew about 50 key educators from the nation's medical schools and societies to Santa Fe in July 1997. Participants at the conference, jointly sponsored by NCGR and the American Medical Association (AMA), explored the potential impact of clinical and ethical genetics issues on the practice of medicine.

"Bringing together prominent genetic scholars and clinicians to share their experiences and insight into real-life genetic dilemmas was educational for all of us," said Judith G. Ribble, NCGR's CME director.

Speakers at "Humans and the Human Genome Project: What Do Physicians Need to Know?" included Reed E. Pyeritz, president, and Joe Leigh Simpson, treasurer, of the American College of Medical Genetics; and Susan P. Pauker, a pediatric geneticist from Harvard Medical School. The program was chaired by Daniel L. Seckinger, AMA group vice president. In roundtable discussions and workshops, conference participants discussed how to integrate genetics into primary healthcare.

"When we make decisions about genetic testing, we're making decisions for people who are not yet born," Pauker told the educators, stressing the need for doctors to be able to counsel patients appropriately. "Our choices affect our children, our children's children, and their children."

NCGR's CME program is designed to disseminate information about genetics and genome research to physicians, focusing on doctors' ability to prevent genetic diseases, improve clinical outcomes of genetically related conditions, and empower consumers to make wise decisions regarding genetic testing. Experts agree that educating physicians is crucial to the success of genetic testing and counseling, and the NCGR CME employs conferences, computer-based training, Internet programming, and other teaching formats to accomplish this goal.

The next NCGR CME event is a symposium jointly sponsored with the American Society of Colon and Rectal Surgeons, "What Surgeons Need to Know About Inherited Colorectal Cancer," on May 3 in San Antonio. For more information, contact Ribble (505/995-4481, *jgr@ncgr.org*) or see the NCGR Web site (*http://www.ncgr.org/cme*).◊

Genetic Testing, Use, and Education

Coalition Promotes Genetic Education Among Healthcare Professionals

The need for all health professionals to have a basic competency in human genetics is underscored by the explosion of information on the role of genetics in human disease; the development of tests for specific disease genes; and the ethical, legal, and social issues associated with the application of scientific advances in genetics. In response to the rapid pace of human genetics research and its impact on health organizations, the National Coalition for Health Professional Education in Genetics (NCHPEG) was established in 1996. The coalition hopes to promote and achieve basic genetic literacy among all health professions through active and ongoing discussion, input, and participation from its members.

Catalyzed by the American Medical Association, the American Nurses Association, and the NIH National Human Genome Research Institute (NHGRI), the coalition consists of leaders from more than 100 healthcare professional organizations, consumer and voluntary groups, government agencies, industry, managed-care organizations, and genetics professional societies. The coalition's goal is to provide an organized, systematic, and national approach to the provision of genetic education for all healthcare professionals.

After the steering committee was organized in July 1996, the coalition held a 1-day meeting of the full membership on March 10, 1997. Top priorities identified by NCHPEG members included the following:

- Development of a comprehensive, Web-based genetics information center;
- Development of a core curriculum in genetics for health professionals, to serve as a template that can be modified according to discipline;
- Integration of genetics into continuing education, certification, and licensure examinations; and
- Development of a model tool to elicit comprehensive, multigenerational family histories.

The coalition has begun to address these priorities, and interdisciplinary working groups have been or are being formed to focus on each area.

One major activity has been the establishment of an NCHPEG Web site, which currently provides basic information about the coalition and its membership (*http://www.medsitenavigator.com/NCHPEG*). The site is being expanded to serve as a coalition communications hub with links to member organization Web pages, an NCHPEG newsletter, and high-quality genetics information sites of interest to healthcare professionals.

Plans are under way to develop a database of current and planned genetics education activities and materials for healthcare professionals. The coalition also is evaluating organizational structure models appropriate for NCHPEG and identifying critical procedures necessary for effective and efficient operations.

The Web site, working group progress and plans, and other coalition activities were presented and discussed at the NCHPEG Steering Committee meeting held November 7, 1997. [Contact: Karina Boehm (NHGRI, 301/402-0955, *kboehm@nhgri.nih.gov*)] ◊

HuGEM to Educate Health Professionals in Genetics

In a national survey of 329 health professionals who provide services to people with genetic disorders in university-affiliated programs across the country, investigators found that almost 70% reported having no course work in human genetics. The survey, conducted during the Human Genome Education Model (HuGEM) Project, targeted health professionals who furnish preventive, diagnostic, referral, advocacy, therapeutic, educational, or counseling services.

A grant from the NIH National Human Genome Research Institute will allow members of the HuGEM project team to provide educational training and resources to selected health professionals through their national organizations. The 3-year grant was made to Georgetown University and the Alliance of Genetic Support Groups. HuGEM II's goal is to increase knowledge of and sensitivity to human genetics; the Human Genome Project; and the ethical, legal, and social issues of genetic testing and research.

"We are in a unique position to make a major step toward educating health professionals to enter the genetic age of the 21st century," said Virginia Lapham (Georgetown University Medical Center), principal investigator of the project. "We hope the result is that healthcare professionals will be better able to serve individuals and families faced with the decisions, miracles, and disappointments of genetic testing, diagnoses, treatments, and promises of cures."

The project will focus on health professionals (other than physicians, nurses, and physician assistants) who traditionally serve persons with genetic conditions in hospitals, clinics, nursing homes, family agencies, and other health settings. This group will include dieticians, occupational therapists, physical therapists, psychologists, social workers, and speech and hearing pathologists. [Contact: E. Virginia Lapham (202/687-8245, *laphamv@gunet.georgetown.edu*, *http://www.dml/georgetown.edu*)]◊

¶ *Judges' Journal* Features DNA and Courts

The Summer 1997 issue of the American Bar Association's *Judges' Journal* is devoted to the challenges and controversies surrounding genetic evidence now entering the nation's courtrooms.

Many topics covered in this issue have been addressed at a series of conferences sponsored by the Einstein Institute for Science, Health, and the Courts (EINSHAC) to educate a core group of 1000 U.S. judges in the basics of genetics and gene testing [*HGN* **8**(1), 1–6].

(see Judges' Journal, p. 15)

CF Genetic-Testing Panel Emphasizes Education

A 14-member nongovernmental panel convened by NIH in April 1997 recommended that genetic testing for mutations that cause cystic fibrosis (CF) be offered to all pregnant couples, those planning a pregnancy, those with a family history of the disease, and partners of people with CF. The panel did not endorse genetic testing of newborns because current research does not show a benefit.

The panel made its recommendations at the close of a 3-day NIH Consensus Development Conference on Genetic Testing for Cystic Fibrosis. The Consensus Development Program was established in 1977 as a "science court" mechanism to resolve controversial topics in medicine and public health in an unbiased, impartial manner. The conference was sponsored by the NIH Office of Medical Applications of Research and the National Human Genome Research Institute.

CF, the most common inherited disorder in Caucasians, occurs when a child inherits two mutated copies of the CF gene, one from each parent. CF symptoms include lung, pancreatic, and intestinal complications ranging from mild to severe. Genetic testing for CF in adults usually involves identifying healthy carriers—people who have a single copy of the mutated gene and will never develop the disease. They are at risk of having a child with CF if their partner is a carrier.

Although more than 600 mutations have been identified in the CF gene, a single mutation (delta-F508) accounts for about 70% of CF alleles in Caucasians, 48% in African-Americans, 46% in Hispanics, and 30% in Asian-Americans and Ashkenazi Jews. Many CF gene tests are designed to detect about 70 of the most common mutations; by combining detection of delta-F508 with other mutations common to specific ethnic groups, it is possible to achieve 90% to 95% specificity for some ethnic groups, panel members explained. Few correlations have been made between particular mutations and disease severity. The panel emphasized the importance of education, counseling, and informed consent for all genetic testing. "As more and more genetic tests for a variety of diseases become available, it is important for both healthcare providers and patients to understand the limitations and implications of such tests," said panel chair R. Rodney Howell (University of Miami School of Medicine).

The full NIH Consensus Statement on Genetic Testing for Cystic Fibrosis is available online (*http://consensus.nih.gov*) or by calling 1-888/644-2667.◊

Helix Directory Lists Testing Laboratories

Since August 1993, Helix: A Directory of Medical Genetics Laboratories has been fulfilling clinician requests for names of laboratories that perform molecular genetic testing. The only resource of its type in North America, Helix was started by clinical geneticist Roberta A. Pagon (University of Washington), who recognized the need of genetic professionals for a centralized, computer-based directory of clinical and research genetic laboratories. Maxine Covington is the database manager.

Under contract with the National Center for Biotechnology Information of the National Library of Medicine, Helix has grown steadily to include 480 diseases for which testing is provided by over 300 laboratories. It lists laboratories by disease and provides information about testing methodologies and contact personnel. To discourage direct contact by patients and thus maintain the traditional relationship between healthcare providers and laboratories, Helix use is restricted to registered healthcare professionals, now numbering 4500. Providers are welcome to list their laboratories, register to use Helix, and provide feedback on the project. Helix services are free to both laboratories and clinicians.

In the first 3 years, most Helix users were genetic professionals such as counselors, medical geneticists, and researchers. Recently, more nongeneticists, particularly neurologists, have become users, presumably reflecting the large number of neurogenetic disorders for which molecular testing is used for routine diagnosis. Since the release of the Internet version in mid-October 1996, the use of Helix has more than doubled, with some 18,000 Internet inquiries in the first 7 months. Assisted access by telephone and fax yields another 20 to 30 inquiries a day.

Originally written in FoxPro, the Helix database was converted to an Informix-based system in the fall of 1997. The users' view of Helix is unchanged, but improved data entry will facilitate database management, and expanded data collection and reorganization of data fields will permit the creation of detailed, categorized reports on Helix data and use. Reports on available testing can be used by insurance companies to help determine reimbursement of testing costs.

Future directions for Helix include adding gene names as a database-search parameter, providing more information about laboratory methodologies (e.g., direct vs linkage molecular testing, FISH, uniparental disomy). Other improvements will make a clearer delineation between research and clinical laboratories and establish links between Helix and laboratory Web sites.

To promote the appropriate use of molecular genetic testing in patient care, Helix staff members are pursuing funds to add educational materials to the database. A recent article pointed out concerns about clinical use of testing, including questionable testing strategies, inaccurate interpretation of test results, and failure to provide genetic counseling [Giardiello et al., *N. Eng. J. Med*. **336**, 823 (1997)]. Access to educational information in Helix would help healthcare professionals who are not familiar with genetic-testing strategies and genetic counseling to understand the implications of testing.◊

Helix Directory of Medical Genetics Laboratories; Children's Hospital and Medical Center; P.O. Box 5371, MS CH-94; Seattle, WA 98105-0371 (206/527-5742, Fax: -5743, *helix@u.washington.edu*, *http://www.hslib.washington.edu/helix*)

Genetic Testing, Use, and Education

Task Force Reports on Genetic Testing

The Task Force on Genetic Testing, chaired by Neil Holtzman (Johns Hopkins University), was formed in 1994 by the NIH-DOE Working Group on Ethical, Legal, and Social Implications of Human Genome Research. The working group asked the task force to review genetic testing in the United States and make recommendations to ensure the development of safe and effective genetic tests to be delivered in high-quality laboratories and used appropriately by healthcare providers and consumers. The working group took this action after considering the imperfect predictability of tests, quality of laboratories providing clinical genetic tests, lack of proven interventions for many disorders, and the questionable ability of many healthcare providers to explain genetic tests accurately and nondirectively to patients.

Genetic Test: The analysis of human DNA, RNA, chromosomes, proteins, and certain metabolites to detect heritable disease-related genotypes, mutations, phenotypes, or karyotypes for clinical purposes. Such purposes include predicting risk of disease, identifying carriers, establishing prenatal and clinical diagnosis or prognosis, monitoring, and screening prenatally and in newborns, but they exclude tests conducted purely for research.

In 1995, the task force undertook a survey of organizations likely to be engaged in genetic testing and conducted in-depth interviews at 29 of the 463 organizations. From respondents whose organizations performed genetic tests, the task force collected informational materials distributed to providers and patients. The task force then commissioned papers on some of the widely used genetic-screening programs in the United States. Individuals, both professionals and consumers, also were asked to report their experiences with various aspects of genetic testing.

Halfway through its deliberations, the task force published interim principles, held a public hearing on them, and invited public comments. Taking these comments into consideration, the task force developed recommendations on which the public was again invited to comment. Final principles and recommendations were presented to the joint working group on May 9, 1997, and the final report was submitted in September (*http://www.med.jhu.edu/tfgtelsi*).

Summary of Recommendations

- The Secretary of Health and Human Services should appoint an advisory committee on genetic testing to be instrumental in implementing the task force's recommendations. The advisory committee or its designate should establish a system for determining which genetic tests require stringent scrutiny.
- Protocols for developing genetic tests that can be used predictively must receive the approval of an institutional review board if subject identifiers are retained and if the test is expected to be readily available for clinical use.
- To permit informed decisions about routine use, test developers must submit their validation and clinical data to external review and to interested professional organizations.
- The task force urges the newly created genetics subcommittee of the Clinical Laboratory Improvement Advisory Committee to consider creating a genetics specialty to ensure that problems specific to genetic testing are addressed in assessments of laboratory quality. If only a subspecialty is feasible for DNA- and RNA-based tests, the subcommittee should then address how to ensure the quality of laboratories performing non-DNA and non-RNA predictive genetic tests.
- The task force encourages the expansion and strengthening of genetics curricula in medical school, residency, and specialty training and the development and improvement of genetics programs by schools of nursing, public health, and social work.
- Hospitals and managed-care organizations should require evidence of competency before permitting providers to order predictive genetic tests that require stringent scrutiny or to counsel about them.
- Physicians who encounter patients with symptoms and signs of rare genetic diseases should have access to accurate information that will enable them to include such diseases in their differential diagnosis, to know where to turn for assistance in clinical and laboratory diagnosis, and to locate laboratories that test for rare diseases. The quality of laboratories providing tests for rare diseases must be ensured, and a comprehensive system must be established to collect data on rare diseases.◊

¶ New Journal: *Genetic Testing*

Genetic Testing is a quarterly, peer-reviewed journal dealing with all aspects of genetic testing: prenatal diagnosis, risk assessment, population screening, targeted testing, methodologies, strategies, and ethical and legal issues. The first issue (Summer 1997) was devoted to population studies of genetic disease in the Ashkenazim, a group that includes over 90% of American Jews. This population has a high frequency of such inherited diseases as Tay-Sachs, Niemann-Pick, Gaucher, and Canavan disease.

[Contact: Mary Ann Liebert, Inc. (914/834-3100, N.Y. State; 1-800/M-LIEBERT, outside N.Y. State; Fax: 914/834-1388, *liebert@pipeline.com*)] ◊

Natural History of Genes Web Site

The Natural History of Genes Web site is designed to take cutting-edge science from the laboratory to the classroom (*http://raven.umnh.utah.edu*). It features inquiry-based projects to encourage investigation and application of scientific concepts, ways to confront student misconceptions about DNA science, and protocols for conducting science experiments. The site also includes core genetic teaching activities, enrichment activities for teachers and students, fun science projects to do at home, and recent press articles that report current findings.◊

Legislation Sought Against Gene Bias

On January 20, 1998, Vice President Al Gore called for federal legislation to bar employers from discriminating against employees on the basis of their genetic makeup. "Progress in genetics should not become a new excuse for discrimination," Mr. Gore said. "Genetic discrimination is wrong—and it's time we ended it."

The Vice President also released an administration report, "Genetic Information and the Workplace," which documents problems of current and future genetic discrimination in the workplace and outlines principles for federal legislation to guard against these abuses (see box, upper right). Such legislation would forbid employers to request or require genetic information, prevent on-the-job discrimination, and ensure that genetic information is not disclosed without the explicit permission of the individual.

Workplace Report

The report states that more and more employers are using genetic testing and monitoring as a condition of employment. One study found that by the year 2000, 15% of employers plan to check the genetic status of prospective employees and their dependents before making job offers. The report also notes that genetic information already is being used to discriminate in the workforce. In addition to numerous individual cases of discrimination, nearly one-fifth of people who have a family member with a genetic disorder reported being discriminated against by employers, insurers, and others.

A 1995 Harris poll revealed that over 85% of Americans are concerned about insurers or employers having access to their genetic information. Another study showed that many high-risk people refuse to take advantage of new genetic tests for fear of losing their jobs or their insurance.

Legislative Protection

Although 14 states have widely varying laws to provide some protection against workplace discrimination, the need for federal protection has been recognized by Congress with the introduction of numerous bills with bipartisan support. Three stand-alone bills would amend existing civil rights or labor laws to protect workers against employment discrimination based on genetic information (S. 1045, Sen. Tom Daschle; H.R. 2275, Rep. Nita Lowey; and H.R. 2215, Rep. Joseph Kennedy). Two additional bills include worker protections against discrimination based on genetic information as part of broader proposals addressing the use of genetic information (S. 422, Sen. Pete Domenici; H.R. 2198, Rep. Cliff Stearns).

Health Insurance

In July 1997, President Clinton urged Congress to pass laws prohibiting health-insurance companies from discriminating against individuals on the basis of information contained in a genetic test or even the request for a genetic test. He endorsed H.R. 306, introduced by Rep. Louise Slaughter (D-NY). The comparable Senate bill, introduced by Sen. Olympia Snowe (R-ME), is S. 89. [Related article, *HGN* **8**(3–4), 1–3.] ◊

Workplace Report: National Human Genome Research Institute (301/402-0911, Fax: -2218) and *http://www.nhgri.Nih.gov:80/HGP/Reports/genetics_workplace.html*

Texts of All Pending Congressional Bills: *http://thomas.loc.gov/home/c105query.html*

Judges' Journal *(from p. 12)*

This project is supported by a grant from the Ethical, Legal, and Social Issues component of the DOE Human Genome Program.

Written in lay language, technical articles in the journal include overviews of the Human Genome Project, molecular biology, and gene testing. The adjudicatory perspective is explored in the context of criminal and civil cases presenting genetic evidence. Guest editors for this issue were Hon. Pauline Newman (U.S. Court of Appeals for the Federal Circuit), Hon. Rosalyn B. Bell (Maryland Court of Special Appeals, retired) and Franklin M. Zweig (EINSHAC).

Articles include "Introducing the Human Genome Project: Its Relevance, Triumphs, and Challenges" by Ari Patrinos and Dan Drell (both of the DOE Human Genome Program); "The Molecules of Life" by Mahlon Hoagland and Bert Dodson; "What Can the New Gene Tests Tell Us?" by Denise Casey (HGMIS); and "Interpreting Scientific Evidence" by John H. Ferguson (NIH), which compares the scientific method with the legal approach and discusses tensions between science and legal proof. [Single copies, $6.50; discounts on bulk purchases (800/285-2221 or 312/988-5522). Casey's article is available from HGMIS (see p. 10). Articles by Patrinos and Drell and by Casey also can be accessed via the Web (*http://www.ornl.gov/hgmis/publicat/judges/judgetoc.html*).]◊

¶ CSHL Book on Integrating Genetics, Medicine

In 1995 and 1996, leaders in genetics and primary healthcare met at Cold Spring Harbor, New York, to consider how genetic advances could benefit public health and to discuss the integration of genetic knowledge and technologies into mainstream medicine. The book *Toward the 21st Century: Incorporating Genetics into Primary Health Care* summarizes the meetings and lays out such key concerns as adequate training of healthcare providers; ensuring access to appropriate care; and patient autonomy including ethical, social, and cultural considerations. Paper, 100 pp., 1997. [Cold Spring Harbor Laboratory Press; 10 Skyline Dr.; Plainview, NY 11803-2500 (800/843-4388 or 516/349-1930, Fax: -1946, *cshpress@cshl.org*, *http://www.cshl.org*)] ◊

¶ *A Question of Genes*

A Question of Genes: Inherited Risks, a 2-hour PBS special produced by Noel Schwerin (NoelEye Documentaries), received rave ratings when it was aired in September 1997. It was recommended as the weekly top choice for viewing by several leading newspapers and magazines and won second prize in the San Luis Obispo Film Festival. Sponsored in part by the DOE Human Genome Program and SmithKline Beecham, the program featured seven stories exploring individual experiences with genetic testing and the resulting choices.

Video orders: 800/440-2651 or *http://www.pbs.org/gene*. A discussion guide to accompany each case history can be downloaded from Schwerin's Web site (*http://www.Pbs.org/gene/educator/41_discussion.html*).

A print copy of the free educator's guide can be ordered from 800/991-1441 or through the extensive Web site, which contains many additional resources for teachers and genetic professionals.◊

Genetic Testing, Use, and Education

¶ New *MIM* Edition

For many years, *Mendelian Inheritance in Man (MIM): A Catalog of Human Genes and Genetic Disorders* by Victor McKusick and his colleagues (Johns Hopkins University School of Medicine) has been the principal source of information on inherited diseases for clinical geneticists and other medical practitioners. The expanded 12th edition, issued in three volumes, includes over 9000 entries (2000 new), along with new and enhanced features, mapping information on some 4000 genes of known function, and information on specific point mutations responsible for 700 genetic disorders or neoplasms (1997, 3792 pp.). An online version is also accessible (*http://www3.ncbi.nlm.nih.gov/omim*).

The portable CD-ROM edition, *MIM CD-ROM*, includes all the data grouped by mode of inheritance. Updated and issued four times a year, the CD-ROM also includes a synopsis of the human gene map, a powerful search engine, and a full-color archive of over 500 relevant images (1997). [Johns Hopkins University Press, 800/537-5487, Fax: 410/516-6998] ◊

McKusick Receives Lasker Award

In recognition of his pioneering work in founding a new branch of medicine—medical genetics—Victor McKusick was honored as winner of the 1997 Albert Lasker Award for Special Achievement in Medical Science. His association of genes and phenotypes in *MIM* and its online version has led medical genetics to the mapping of tens of thousands of genes and to the Human Genome Project. In a lighthearted comment, McKusick attributed his research discoveries in part to his "chauvinistic, opportunistic, dilettante, and parochial" character.

Known as "America's Nobels," the Lasker Medical Research Awards for 50 years have celebrated the scientists, physicians, and public servants who have contributed to major advances in the understanding, diagnosis, treatment, prevention, and cure of many of the great cripplers and killers of this century (*http://hopkins.med.jhu.edu*). Many consider the Lasker Awards, founded by advertising tycoon Albert Lasker and his wife Mary, to be the most significant biomedical science prizes in the United States.◊

¶ *Genetic Secrets*

Genetic Secrets: Protecting Privacy and Confidentiality in the Genetic Era, edited by Mark Rothstein (University of Houston), was published in December 1997. The book arose from a workshop on Medical Information and the Right to Privacy held in 1994 in Washington, D.C. In a comprehensive exploration of the ethical, legal, and social issues surrounding new discoveries in genetics, distinguished experts in diverse fields consider the many contexts in which issues of genetic privacy arise. The contexts range from research and clinical settings to workplaces, insurance offices, schools, and the courts. In the final chapter, Rothstein and others discuss flaws in existing and proposed legislation designed to protect genetic privacy and offer a new set of guidelines for policymakers. Both the book and the original workshop were sponsored by DOE's Office of Energy Research and Office of Environment, Safety, and Health. 448 pp. [Order Department; Yale University Press; P.O. Box 209040; New Haven, CT 06520-9040 (800/987-7323, Fax: /777-9253)] ◊

¶ Special Genome Issues of *Your World/Our World*

"Exploring the Human Genome" is an expanded 24-page issue of *Your World/Our World: Biotechnology and You*. The colorful magazine, developed by the Pennsylvania Biotechnology Association (PBA) for teaching biotechnology to students in grades 7 through 10, is also appropriate for adult lay audiences. Articles and learning activities explore the Human Genome Project; DNA structure and function; genes, proteins, and genetic disease; mapping; invention of the polymerase chain reaction; informatics; Huntington's Disease; and ELSI implications of genome research. This special genome issue is made available in part by DOE, which provided a complimentary copy to every 7th- to 10th-grade U.S. science teacher (about 40,000).

Building on concepts introduced in the Human Genome issue, the 15-page "New Diagnostics" issue of *Your World/Our World* is devoted to techniques used to detect disease. Articles cover such topics as immuno- and genetic diagnostics, strep throat control, and safeguarding the blood bank. Sample copy of Human Genome issue and price list available from HGMIS (see contact information on p. 10). [PBA, 800/796-5806 or 814/238-4080, Fax: -4081, *73150.1623@compuserve.Com*] ◊

¶ NSGC Paper on Predisposition Testing

"Predisposition Genetic Testing for Late-Onset Disorders in Adults: A Position Paper of the NSGC" appeared in the October 15, 1997, issue of the *Journal of the American Medical Association (JAMA)* **278**(15), 1217–20. Written by Wendy McKinnon and other members of the National Society of Genetic Counselors, the paper puts forth NSGC's recommendations for comprehensive genetic counseling and education by professionals who offer predisposition testing.◊

¶ Journals Publish Special Genome Issues

- *Science* **278**(5338), October 24, 1997 *http://www.sciencemag.org/content/vol278/issue5338* (requires registration)
- *Journal of the American Medical Association* **278**(15), October 15, 1997 *http://www.ama-assn.org/sci-pubs/journals/archive/jama/vol_278/no_15/toc.htm*

ExPASy Web Server

The ExPASy molecular biology Web server is dedicated to the analysis of protein and nucleic acid sequences and 2-D PAGE (*http://www.expasy.ch*). The site includes numerous links to databases such as SWISS-PROT, PROSITE, and ENZYME; tools and software packages such as Swiss-Model, Melanie, and TOOLS; and gateways to other data, molecular biology servers, and interesting places to visit.◊

Microbial Genomics

Lyme Disease Bacterium Genome Sequenced

Researchers at The Institute for Genomic Research (TIGR) announced the completion of an 18-month effort to sequence the genome of *Borrelia burgdorferi*, the bacterium that causes Lyme disease. The most common tick-transmitted illness in the United States, Lyme disease is difficult to diagnose and cure. If not treated early with antibiotics, it causes nerve damage and progressive arthritis. The work, described in a paper by Claire Fraser and others in the December 11, 1997, issue of *Nature*, represents the sixth genome completely sequenced at TIGR in 2.5 years. Sequence data and related annotation are available on the TIGR Web site (*http://www.tigr.org/new*).◊

Cloning: From DNA Molecules to Dolly

The possibility of human cloning, raised when Scottish scientists at the Roslin Institute created the much-celebrated sheep "Dolly" (*Nature* **385**, 810–13, 1997), has aroused worldwide interest and concern because of its scientific and ethical implications. The feat, cited by *Science* magazine as the breakthrough of 1997, also has generated uncertainty over the meaning of "cloning"—an umbrella term traditionally used by scientists to describe different processes for duplicating biological material.

Cloning DNA, Cells, and Animals

To Human Genome Project researchers, cloning refers to copying genes and other pieces of chromosomes to generate enough identical material for further study. Cloned collections of DNA molecules (called clone libraries) enable scientists to produce increasingly detailed descriptions of all human DNA—the aim of the Human Genome Project. Bacterial or yeast cells are used routinely by scientists to make these extra copies of human DNA molecules. Researchers also can exploit the natural process of cell division to make many copies of an entire cell. The genetic makeup of these cloned cells, called a cell line, is identical to the original cell.

Two other types of cloning produce complete, genetically identical animals. Blastomere separation (sometimes called "twinning" after the naturally occurring process that creates identical twins) involves splitting a developing embryo soon after fertilization of the egg by a sperm (sexual reproduction) to give rise to two or more embryos. The resulting organisms are identical twins (clones) containing DNA from both the mother and the father.

Dolly, on the other hand, is the result of another type of cloning that produces an animal carrying the DNA of only one parent. Using somatic cell nuclear transfer, scientists transferred genetic material from the nucleus of an adult sheep's udder cell to an egg whose nucleus, and thus its genetic material, had been removed. (All cells that are not egg or sperm cells are somatic cells.)

Why Clone?

One goal of this and similar research is to develop efficient ways to alter animals genetically and reproduce them reliably. Alterations have included adding genes (such as those for human proteins) to create drug-producing animals as well as inactivating genes to study the effects and possibly create animal models of human diseases. Cloning technology also may someday be used in humans to produce whole organs from single cells or to raise animals having genetically altered organs suitable for transplanting to humans.

The technique used to produce Dolly and other cloned animals is an extension of 40 years of research using DNA from nonhuman embryonic and fetal cells. Before this demonstration, scientists believed that once a cell became specialized—a liver, heart, udder, bone, or any other type of cell—the change was permanent and other unneeded genes in the cell became inactive. Dolly's creators demonstrated that nuclei of an adult animal's specialized cells can be made to revert to a nonspecialized, embryonic state, thus restoring the ability to give rise to any kind of cell. Explorations into how cells revert to an undifferentiated state may provide insights into the process by which cells become cancerous.

NBAC Recommendations

Following the birth of Dolly in 1997, public reaction prompted U.S. President Bill Clinton to ask the National Bioethics Advisory Commission (NBAC) for recommendations on the science of human cloning and its ethical, religious, legal, and regulatory implications. Released in June 1997, the NBAC report concluded that attempts to clone humans are "morally unacceptable" for safety and ethical reasons. NBAC advised continuing the current moratorium on using federal funds for human cloning but recommended that any legislation should be temporary and that the situation should be reviewed again in 3 to 5 years. The NBAC report is available on the Web (*http://www.nih.gov/nbac/nbac.htm*).

In January 1998, the U.S. Food and Drug Administration (FDA) announced its authority to regulate human cloning, thus making it a violation of federal law for anyone to try the procedure without FDA approval. Acting FDA Commissioner Michael Friedman stated that the agency considers the manipulations involved in human cloning to be serious health and societal issues for the mother and fetus.

Also in January 1998, 19 European countries signed a ban on human cloning. Although still supporting areas of cloning research that could lead to significant medical benefits, President Clinton continues to press Congress for a ban on human cloning in the United States. *[Denise Casey, HGMIS]*◊

¶ More on Cloning

http://ethics.acusd.edu/reproductive_technologies.html#Internet Resources — Page down for links to related newspaper and magazine articles and *some* background resources.

http://stlr.stanford.edu/stlr/symposia/cloning/contents_f.htm — Transcript of a discussion panel on the implications of human cloning.

Clone: The Road to Dolly and the Path Ahead by Gina Kolata (William Morrow and Company, 1997). Book review and first chapter: *http://www.nytimes.com/books/97/12/28/reviews/971228.28turnert.html*

Remaking Eden: Cloning and Beyond in a Brave New World by Lee M. Silver (Avon Books, 1997). Book review: *http://www.nytimes.com/books/98/01/11/reviews/980111.11raeburt.html*

LLNL and Onyx Collaborate on Proteome Project

Lawrence Livermore National Laboratory (LLNL) Human Genome Center and Onyx Pharmaceuticals, Inc., recently signed an agreement to collaborate in developing automated, high-throughput methods for generating proteins from gene sequences. To help identify and characterize new therapeutic targets for Onyx's current drug-screening programs, research will focus on genes known to play a causative role in cancers. Joanna Albala of LLNL and Robin Clark and Anthony Davies of Onyx are principal investigators in the collaboration.

LLNL brings to the collaboration the Integrated Molecular Analysis of Gene Expression (called I.M.A.G.E.) collection of arrayed cDNA libraries, along with the expertise and infrastructure for miniaturizing and automating biological sample handling. Onyx has developed expertise and reagents for expressing proteins using the baculovirus expression system and for purifying a wide variety of human proteins using epitope tags. Investigators expect the new system to reduce drug-development time by providing a ready source of proteins for assay development and drug screening.

The 2-year collaboration is the first at LLNL to be funded under the Biotechnology Strategic Targets for Alliances in Research (BioSTAR) project. BioSTAR provides for university matching of private-sector funding for biotechnology research.◊

Genetics in the News

DOE Genome Researchers Win R&D 100 Awards

DOE researchers in 12 facilities across the country won 36 of the 100 awards given by *R&D* magazine for 1996 work. DOE award-winning research ranged from advances in supercomputing to the biological recycling of tires. Announced in July 1997, these awards bring DOE's R&D 100 total to 453, the most of any single organization and twice as many as all other government agencies combined.

Two genome-related research projects sponsored by the DOE Office of Biological and Environmental Research, Office of Energy Research, received R&D 100 Awards:

- **Richard Keller** and **James Jett** (Los Alamos National Laboratory) with **Amy Gardner** (Molecular Technologies, Inc.): "Rapid-Size Analysis of Individual DNA Fragments." Speeds determination of DNA fragment sizes, making DNA fingerprinting applications in biotechnology and other fields more reliable and practical.
- **Edward Yeung** (Ames Laboratory): "ESY9600 Multiplexed Capillary Electrophoresis DNA Sequencer." Allows simultaneous use of multiple capillary tubes for DNA sequencing and has potential for significantly lowering the cost and increasing the speed of gene sequencing.

R&D began making annual awards in 1963 to recognize the 100 most significant new technologies, products, processes, and materials developed throughout the world during the previous year (*http://www.rdmag.com/rd100/100award.htm*). Winners are chosen by the magazine's editors and a panel of 75 scientific experts in a variety of disciplines. Previous winners of R&D 100 Awards include such well-known products as the flashcube (1965), antilock brakes (1969), the automated teller machine (1973), the fax machine (1975), digital compact cassette (1993), and Taxol anticancer drug (1993).◊

Hollaender Fellows Named

DOE has announced the award of 11 FY 1997 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 sciences.

Fellowship winners were chosen from a field of 47 applicants who received their doctoral degrees after April 30, 1995. Two of the 11 were in genome-related topics. Listed below are each fellow's name, university of doctoral degree, host laboratory and research mentor, and proposed research topic.

Winners in Genome-Related Topics

- **Jeffrey Koshi** (University of Michigan, Ann Arbor): Los Alamos National Laboratory, William Bruno. Construction, Analysis, and Use of Optimal DNA Mutation Matrices.
- **Sandra McCutchen-Maloney** (Texas A&M University): Lawrence Livermore National Laboratory, Michael Thelen. Structure and Function of a Damage-Specific Endonuclease Complex.

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

TIGR Terminates Relationship, Releases Data

On June 24, 1997, The Institute for Genomic Research (TIGR) announced termination of its relationship with its profit-making arm, Human Genome Sciences, Inc. Also on June 24, TIGR released 40 Mb of microbial genome data (*http://www.tigr.org*).

In addition, TIGR released data from the *Borrelia burgdorferi* (**completed**) chromosome on July 25.

All data are accessible by anonymous ftp (*ftp.tigr.org*). Users should log in as "anonymous" and provide their e-mail address as a password. The data also may be accessed through the Web (*ftp://ftp.tigr.org/pub/data*).

Searching the Data

TIGR's completed and published genomes are searchable through the TIGR site (*http://www.tigr.org/tdb/mdb/mdb.html*). A new BLAST service is now available at the National Center for Biotechnology Information for searching the incomplete microbial genomes, which are not yet accessible through GenBank or Entrez (*http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-tigrbl*). TBLASTN (a user's protein query vs a six-frame translation of the microbial DNA sequences) uses the new gapped BLAST algorithm.

The South African National Bioinformatics Institute has put up a search engine for TIGR contigs and Sanger cosmids of *Mycobacterium tuberculosis*, finished and unfinished (*http://www.sanbi.ac.za*). In addition, a first pass at ORF-predicted sets can be used for searching.◊

Bermuda Sequencing Strategy Report on Web

A summary of the Second International Strategy Meeting on Human Genome Sequencing held in Bermuda on February 27–March 2, 1997, is on the Web (*http://hugo.gdb.org/bermuda2.htm*). The summary includes principles endorsed by attendees on sequence quality, submission and annotation, and claims and etiquette. The next meeting is planned for the end of February 1998 in Bermuda.◊

Web Site for DOE Santa Fe Meeting Proceedings

The proceedings of the sixth DOE Human Genome Program Contractor-Grantee meeting, held November 9–13, 1997, in Santa Fe, New Mexico, are on the Web (*http://www.ornl.gov/hgmis/publicat/97santa/santafe.html*). The site includes 158 abstracts from the meeting. The print copy is available from HGMIS (see p. 10).◊

HGP: Reaching Maryland's Minority Communities

About 150 leaders of minority communities came together June 20–21, 1997, at the University of Maryland, Baltimore (UMB) to learn about the Human Genome Project. The meeting was organized by Carmen Nieves (Center for Minority Health Research, UMB) and Ray Zilinskas (UM Biotechnology Institute, College Park). Its goal was to (1) inform minority communities about the Human Genome Project by explaining its potential benefits and clarifying its possible ethical, legal, and social implications (ELSI); and (2) make the aspirations and interests of these communities known to genome project scientists and policymakers. This program grew from organizers' concerns about an information vacuum among minorities regarding the genome project and the possibility that suspicions will arise about the project's intent.

Minorities (specifically African-Americans, Native Americans, and Hispanic Americans) generally are under-represented among scientists involved in Human Genome Project research and its applications. Many believe that "the" human genome being sequenced contains genes primarily from Northern Europeans and, therefore, that findings will benefit mostly Caucasians. [See Editor's Note, p. 20]

Minorities also have a history of being under-represented in clinical trials involving treatment of inherited diseases—especially diabetes, hypertension, and heart disease, which are prevalent among these populations. This lack of participation may stem from the mistrust of scientific projects focusing on health. Most minorities know about the infamous Tuskegee Study of Untreated Syphilis in the Negro Male (1932–72), involuntary sterilization programs in state mental institutions during the 1920s and 1930s, and discrimination against African-Americans as a result of sickle-cell screening in the 1970s.

Meeting Highlights

The meeting's major objective was to convey information to attendees, so most of the time was allocated to formal presentations followed by question-and-answer sessions. Some highlights follow.

The first speaker, Martha Krebs, director of the DOE Office of Energy Research, reviewed DOE's involvement in the genome project and discussed how people might become "responsible ancestors." She stated that "all of us, not just the scientists, have a responsibility to address the tough issues that modern science raises. We must take upon ourselves the charge to be responsible ancestors as we support and utilize the science of today, which undoubtedly impacts the children of tomorrow."

Karen Nelson (The Institute for Genomic Research) explained to the largely lay audience that alterations that can cause disease have been found in an estimated 5000 genes. Although the development of tools for diagnosing human diseases is likely to outpace new therapeutics, she said, scientists will learn how to use various interventions to prevent or treat many genetically linked diseases.

In his presentation, "The Genetics of Behavior and IQ," Jonathan Beckwith (Harvard University) argued that research has yet to provide convincing evidence for strong deterministic genetic influences on human behavior. (Genetic determinism emphasizes the role of genetics on behavior and minimizes the role of environment.) Beckwith noted that the "new, more sophisticated understanding of human genetics resulting from the Human Genome Project gives us an expanding picture of complexity that challenges the simplistic notions presented in *The Bell Curve* and similar works."

Fatimah Jackson [UM, College Park] presented a fervent discourse on the potential benefits and costs of the Human Genome Project. Jackson asked two central questions: "Will comprehensive molecular genetic testing be harmful to African-Americans and other groups that are under-represented among molecular genetic researchers and funders? Once initiated, will such testing be helpful to these groups?" She spoke of a history of "bad science" in which testing has been used as a tool of "coercion, subjugation, and oppression" against certain groups, as in the Tuskegee syphilis study, and warned against repeating the ethical and moral errors of the past. "It is essential that [minorities] be permitted to collaborate meaningfully in the development of hypotheses and research designs; collection, analysis, and evaluation of data; and development of subsequent policy initiatives." She encouraged African-Americans and members of other minority groups to have more input into genome databases and data interpretation.

Jonathan Marks (Yale University) discussed three categories of ideas that often affect genetic research because they are strongly embedded in public consciousness. These ideas, stemming from "folk heredity," are racism (the assessment of individuals' worth based on properties stereotypically assigned to their groups), hereditarianism (the belief that innate variation is the root cause of specific observed differences), and essentialism (the tendency to ignore visible diversity in favor of an imaginary uniformity).

Robert F. Murray, Jr. (Howard University) stated that genetic screening can be to the patient's advantage by allowing early detection, intervention, and treatment. However, gathering information about susceptibility long before the disease is clinically evident makes screening socially sensitive and possibly dangerous. Hindrances to employment, health insurance, and professional advancement can be intensified. The past history of

The meeting was supported by the ELSI component of the DOE Human Genome Program, with supplementary funding from the NIH National Human Genome Research Institute. Local sponsoring organizations were the Center for Minority Health Research, University of Maryland (UM), Baltimore; and the Center for Public Issues in Biotechnology, UM Biotechnology Institute in College Park. Government officials, legislators, educators, attorneys, officers of community-based cultural and social organizations, and scientists were among the participants.

Meetings

genetic screening (as, for example, in screening for sickle cell anemia) indicates that without certain protective measures, information may be used to stigmatize or discriminate against individuals, especially members of minority groups.

In his talk on "Economic Opportunities and Minorities," Arche McAdoo (UM Biotechnology Institute) pointed out that Maryland has the third-largest concentration of biotechnology companies in the United States and, therefore, the potential to become a major player in the industry. To ensure that all groups gain from economic opportunities, youths should be positioned to achieve the skills needed to become stakeholders and major players in biotechnology. This means that parents must take an interest in their children's education, making certain that it is of high quality and includes the necessary science for participation in the field.

A community empowerment model for genetic counseling was presented by Ilana Mittman (Howard University). She focused on designing culturally appropriate means to provide genetic services to Asians and Hispanics in San Francisco and to African-Americans and Russian immigrants in Baltimore. Mittman said that members of ethnic minorities face formidable cultural, financial, educational, and physical barriers to receiving medical services based on the new understanding of genetics. Often, she continued, they are either unable to access these services or face culturally insensitive encounters with the genetic-service delivery system.

Like other speakers before her, Mittman mentioned a lack of trust in healthcare providers and the low number of minorities in the medical genetics and genetic-counseling professions and in government agencies responsible for developing public policy in this area. She then presented a model program, conducted under the auspices of Howard University, that encourages minorities to enter the genetic-counseling field and genetic sciences in general.

In addition to formal presentations, one keynote address was presented each day by a political figure. Maryland's lieutenant governor, Kathleen Kennedy Townsend, spoke of the genome project's potential economic importance to Maryland and the measures needed to make certain its benefits are shared equitably by all sectors of society. Participants were impressed with her thorough understanding of the Human Genome Project and the ELSI issues it might engender.

Larry Young, Chairman of the State Senate Health Subcommittee, discussed how the Human Genome Project may come to benefit the health status of all. After having candidly admitted that until very recently he knew nothing about the project, he promised to inform the next session of the Maryland legislature of its existence and importance.

Formal presentations were followed by a panel discussion. The panel, which included two African-Americans, a Native American, two Asians, a Caucasian, and a Hispanic, served to broaden the meeting's scope by bringing up important ELSI concerns that had not been covered in formal presentations. As the conference organizers expected, the panel discussion led to a wide-ranging, impassioned debate involving many of the attendees.

Two breakout sessions were held to allow participants to discuss their concerns about ELSI issues generated by the genome project and what they wanted to do about them. Each group included scientists who acted as resource persons. The consensus was to organize a task force that will develop a plan for continuing involvement to influence the Human Genome Project's direction and ensure its equitable application.

Michael Carter (Maryland State Department of Health and Mental Hygiene) indicated that the conference was extremely useful as a communication device and that many attendees were very interested in following genome project developments and in ensuring that all groups benefit equally. He said the conference had led participants to think about equity, privacy, and other issues; realize that they should become involved; and plan to take the steps to do so.

Summary

Attendees departed with mixed feelings of apprehension and excitement. On the one hand, they perceived that although government agencies, scientists, and bioethicists have been involved in the genome project since 1986, minorities have not been sufficiently aware of the research and technological advances that will significantly affect how diseases are diagnosed and treated in the future. The concern is that these benefits will not be shared equitably. Furthermore,

(see Minorities, p. 21)

Editor's Note: "Reference" Genome to Contain Basic Set of Genes

Except for identical twins, each human has his or her own unique genome—the complete set of DNA, or genetic material, found in the 46 chromosomes of each cell. Scientists estimate that individuals differ in about 0.1% of their 3 billion DNA base pairs. Although people who make up a particular population group share common ancestors and are more likely to share some genetic sequences, scientists believe that individuals within a group are genetically more variable than the groups are.

Given these differences, all humans still share the same basic set of genes and genomic regulatory regions that control the development and maintenance of their biological structures and processes. The Human Genome Project's goal is to determine the DNA sequence for a complete "reference" human genome that will help orient researchers and provide them with tools for further studies of fundamental human biology. Because the genome of each person is unique and different samples will be used for sequencing, the reference sequence will not represent an exact match for any one person's genome.

Some researchers outside the Human Genome Project are beginning to look more closely at differences in DNA sequences of particular genomic regions to study the role of genetic variation in disease and susceptibilities. Researchers in the Environmental Genome Project, for example, plan to sequence about 200 genes from 1000 individuals to investigate how some genetic differences influence susceptibility to environmental exposures (see article, page 10). Another group seeks to catalog genetic differences in groups around the world (see the Human Genome Diversity Project, *http://lotka.stanford.edu/research*). Through these and other future studies, scientists will begin to identify and understand factors influencing health.◊

Workshop Focuses on Sequence Annotation

New Paradigms and Organizational Models Needed

The Fifth International Conference on Intelligent Systems for Molecular Biology held June 21–25, 1997, in Porto Carras, Greece, ended with a workshop on Automatic Annotation of Genome Sequence Data.

Automatic annotation of large amounts of genomic DNA sequence clearly is and will continue to be a formidable challenge. When completed, the human genome sequence will consist of 24 strings of As, Ts, Cs, and Gs with a combined length of 3 billion characters. Without marking the locations of such biologically important parts of the sequence as the genes and their regulatory elements, this string of characters has little usefulness. Annotating the genome sequence in parallel with its determination is critical.

Attendees felt this problem will be addressed properly only by developing very efficient computational tools for initial sequence annotation, treating the annotations as hypotheses, and testing and verifying them in the laboratory. Additionally, for maximum usefulness, the generated annotation results must be stored in an easily retrievable and queryable form in well-curated databases. The "If you sequence it, the community will annotate it" approach is unlikely to produce desired results, and new paradigms and possibly new organizational models will be needed to present genomic sequence in its most useful form.

Workshop Speakers

Eight workshop speakers addressed the challenges and technologies in automatic annotation and the most efficient division of labor between biology and computer science.

Introductory remarks by session chairman Chris Sander [European Molecular Biology Laboratory–European Bioinformatics Institute (EBI)] made clear that no one yet has the experience to know *the* right way to proceed with automatic annotation. Richard Durbin (Sanger Centre) stressed an often-repeated theme that proper annotation will require wet-laboratory work as well as computational annotation. He also stressed the need for curated databases. Michael Ashburner (EBI) discussed his experience in annotating *Drosophila* sequences and the need for hierarchial controlled vocabularies. He suggested the possibility of an annotation database that would be separate from but seamlessly linked to the sequence databases.

Three other speakers addressed general problems in genomic-sequence annotation: Antoine Danchin (Institut Pasteur) discussed annotation of the *Bacillus subtilis* genome, Terry Gaasterland (Argonne National Laboratory) described annotating microbial genomes, and Chris Overton (University of Pennsylvania) shared experiences from a project to annotate genomic sequence from human chromosome 22. Other speakers discussed annotation efforts and tools being developed in the bioinformatics industry. *[Richard Mural, Life Science Division, Oak Ridge National Laboratory,* muralrj@ornl.gov*]* ◊

Minorities *(from p. 20)*

information generated by the genome project may be used by unscrupulous organizations to discriminate against persons who possess what are defined as less-than-optimal genetic profiles.

On the other hand, findings resulting from the Human Genome Project can significantly enhance the health status of all persons, including minorities. Thus, informed groups can act on genetic information to make certain their communities share equitably. *[Carmen Nieves, University of Maryland, Baltimore, and Ray Zilinskas, University of Maryland, College Park]* ◊

¶ New Publications Available from HGMIS

The following publications are available from HGMIS at the address on page 10.

Human Genome Program Report

The 116-page DOE *Human Genome Program Report, Part 1*, published in November 1997, explores the genome program's history and describes research highlights, technology transfers, research at major DOE centers, coordination with other programs, and management of the program. A separate volume, Part 2, contains 1996 research abstracts.

A Vital Legacy

A Vital Legacy: Biological and Environmental Research in the Atomic Age is a 48-page booklet prepared by Douglas Vaughan (Lawrence Berkeley National Laboratory) to commemorate the 50th anniversary of the founding of DOE's Biological and Environmental Research (BER) program. The colorful booklet describes highlights from five decades of research and looks ahead to the program's exciting future.◊

Web Sites

50th Anniversary BER Exhibits

Exhibits from the DOE Biological and Environmental Research Program's 50th Anniversary Symposium held May 21 and 22, 1997, at the National Academy of Sciences are on the Web (*http://www.er.doe.gov/production/ober/samplexh.html*).◊

✔ Contact Updates

- Cancer Genome Anatomy Project: *http://www.ncbi.nlm.nih.gov/ncicgap*
- Pacific Southwest Regional Genetics Network: Fax: 510/540-3293, *pcohen@genetic.dhs.cahwnet.gov, http://www.best.com/~psrgn*

Calendar of Genome and Biotechnology Meetings*

More comprehensive lists of genome-related meetings and organizations offering training are available at *http://www.ornl.gov/hgmis* or from HGMIS (see p. 10 for contact information).

March 1998

6–8. Cancer Genetics for the Clinician; Clearwater Beach, FL [H. Lee Moffitt Cancer and Research Inst., 813/632-1775, Fax: /979-3787; *diamontm@moffitt.usf.edu*]

9. 2nd NIH Gene Therapy Policy Conf.; Bethesda, MD [Strategic Results, Fax: 410/377-6429; *orda2@strategicresults.com*; *http://www.strategicresults.com/nih/orda2*]

10–11. Functional Genomics: From Identifying Proteins to Faster Drug Discovery; Washington, DC [NMHCC, 888/882-2500, Fax: 941/365-0157; *http://www.biotech.nmhcc.org*]

19–20. Seizing Opportunities in Emerging Biochip Technologies; San Francisco [GBR, 916/773-3236, Fax: -9321; *globalbr@global8.com*; *http://www.globalbusinessresearch.com*]

21–24. ABRF '98. From Genomes to Function: Technical Challenges of the Post-Genome Era; San Diego [ABRF, 301/530-7010, Fax: -7014; *abrf98@faseb.org*; *http://www.faseb.org*]

22–25. RECOMB '98: 2nd Annu. Intl. Conf. on Computational Molecular Biol.; New York [G. Benson, 212/241-5777, Fax: /860-4630; *benson@ecology.biomath.mssm.edu*; *http://www.mssm.edu/biomath/recomb98.html*]

25–29. 39th Drosophila Research Conf.; Washington, DC [L. Raftery, 617/726-1825, Fax: -4453; *lraftery@cbrc.mgh.harvard.edu*; *http://cbbridges.harvard.edu:7081/docs/news/announcements/meetings*]

26. TIGR/NRC/DOE Distinguished Speaker Series: Carla Shatz (Stanford Univ.); Rockville, MD [D. Hawkins, 301/838-3501, Fax: -0209; *dhawkins@tigr.org*; *http://www.tigr.org*]

26–27. 2nd Intl. Symp. on Fungal Genomics: Novel Developments and Approaches in Structure, Function, and Evolution of Genomes; Athens, GA [IBC, 508/481-6400, Fax: -7911; *reg@ibcusa.com*; *http://www.ibcusa.com*]

28–29. 1998 CORN Meeting; Washington, DC [C. Hinton, 404/727-1475, Fax: -1827; *cfh@rw.ped.emory.edu*; *http://www.cc.emory.edu/PEDIATRICS/corn/corn.htm*]

28–30. HGM '98–HUGO's Human Genome Meeting; Torino, Italy [Secretariat, +44-171/935-8085, Fax: -8341; *hugo@hugo-europe.org.uk*; *http://hugo.gdb.org/hgm98.htm*]

April 1998

6–7. Trinucleotide Repeats; Santa Fe [CHI, 617/630-1300, Fax: -1325; *chi@healthtech.com*; *http://www.healthtech.com/conferences*]

9. TIGR/NRC/DOE Distinguished Speaker Series: Floyd Bloom (Scripps Research Inst.); Rockville, MD [see contact Mar. 26]

16–17. Partnering for Functional Genomics Research; Oak Ridge, TN [R. Mann, 423/574-5845, Fax: -3036; *mannrc@ornl.gov*; *http://lsd.Ornl.gov/fgconf*]*

16–19. DNA Repair: Bacteria to Humans; Warrenton, VA [GSA, 301/571-1825; *http://www.faseb.org/genetics/gsa/dna-main.htm*]

23–24. The HGP: Science, Law, and Social Change in the 21st Century; Cambridge, MA [G. Cervini, 617/258-0633; *cervini@wi.mit.edu*; *http://www.wi.mit.edu/bio/genetics_policy.html*]

27–28. Intl. Workshop on Advanced Genomics: Expression Profiling and Related Technol.; Tokyo [Y. Sakaki, +81-3/5449-5622, Fax: -5445; *sakaki@ims.u-tokyo.ac.jp*]

29–May 3. Zebrafish Development and Genetics; Cold Spring Harbor, NY [CSHL, 516/367-8346, Fax: -8845; *meetings@cshl.org*; *http://www.cshl.org*]

May 1998

4–6. 2nd Annu. HGP–Europe; Cannes, France [see contact: Apr. 6–7]

11–29. CLMA Internet Symp. On Genetic Predisposition Testing [C. Steward, 610/995-9580; *http://www.clma.org*]

13–17. Genome Mapping, Sequencing, and Biol.; Cold Spring Harbor, NY [see contact: Apr. 29–May 3]

14. TIGR/NRC/DOE Distinguished Speaker Series: Bailus Walker (Howard Univ.); Rockville, MD [see contact Mar. 26]

15–16. Genome Horizons: Public Deliberations and Policy Pathways; Washington, DC [T. Hartman, 734/647-8304, Fax: /936-0927; *genome.horizons@umich.edu*; *http://www.sph.umich.edu/genome/ghconf.htm*]

17–21. ASM 98th Annu. Meeting; Atlanta [ASM, 202/737-3600; *http://www.asmusa.org*]

19–21. 6th Gene Amplification and Detection; Washington, DC [see contact: Apr. 6–7]

28–31. ASGT Annu. Meeting; Seattle, WA [G. Stamatoyannopoulos, 206/543-3526, Fax: -3050; *gstam@u.washington.edu*; *http://weber.u.washington.edu/~asgt*]

June 1998

8–10. BGA 1998 Meeting; Stockholm [N. Pedersen, +46-8/728-7418; Fax: /31-3961; *Nancy.Pedersen@imm.ki.se*; *http://www.nyas.org/nyasconf.htm*]

11. TIGR/NRC/DOE Distinguished Speaker Series: Shirley Tilghman (Princeton Univ.); Rockville, MD [see contact Mar. 26]

15–16. 7th Annu. Bioinformatics Conf.; Boston [see contact: Apr. 6–7]

28–July 1. 6th Intl. Conf. on ISMB; Montreal [J. Glasgow, 613/545-6058, Fax: -6513; *ismb98@qucis.queensu.ca*; *http://www-lbit.iro.umontreal.ca/ISMB98*]

28–July 3. 8th Intl. Symp. on the Genetics of Industrial Microorganisms; Jerusalem [Organizing Committee, Fax: +972-3/517-5674; *gim@kenes.com*; *http://www.tau.ac.il/~biotwww/gim98.html*]

July 1998

28–Aug. 2. 1998 Yeast Genetics and Molecular Biol. Conf.; College Park, MD [M. Ryan, 301/571-1825, Fax: 530-7079; *mryan@genetics.faseb.org*; *http://genome-www.stanford.edu/Saccharomyces/yeast98/index.html*]

August 1998

19–23. Cancer Genetics and Tumor Suppressor Genes; Cold Spring Harbor, NY [see contact: Apr. 29–May 3]

September 1998

2–6. Mouse Molecular Genetics; Cold Spring Harbor, NY [see contact: Apr. 29–May 3]

10–13. 2nd Intl. Conf. on DNA Sampling. Commercialization of Genetic Research: Legal, Ethical, and Policy Issues; Edmonton, Alberta [Health Law Institute; *hli@ualberta.ca*; *http://www.law.ualberta.ca/centres/hli*]

17–20. 10th Intl. Genome Sequencing and Analysis Conf.; Miami [TIGR, 301/838-3515, Fax: -0229; *seqconf@tigr.org*; *http://www.tigr.org*]

23–27. Gene Therapy; Cold Spring Harbor, NY [see contact: Apr. 29–May 3]

October 1998

27–31. ASHG; Denver [see contact July 28–Aug. 2; *http://www.faseb.org/genetics/ashg/ann-meet/ashgmeet.htm*] ◊

*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*

March 1998

4–10. Advanced Molecular Cytogenetics; Cold Spring Harbor, NY [CSHL, 516/367-8346, Fax: -8845; *meetings@cshl.org*; *http://www.cshl.org*]

16–20. Basic Genetic Linkage Analysis; Zurich [K. Montague, 212/327-7979, Fax: -7996; *montagk@rockvax.rockefeller.edu*; *http://linkage.rockefeller.edu/course/here.html*]

18–31. Advanced Genome Sequence Analysis; Cold Spring Harbor, NY [see contact: Mar. 4-10]

April 1998

5–8. Genetic Analysis Methods for Medical Researchers; Durham, NC [V. Roberts, 919/684-2470; Fax: -2275; *genclass@genemap.mc.duke.edu*; *http://www2.mc.duke.edu/depts/medicine/medgen/course/index.html*]

May 1998

31–June 7. Medical Informatics; Woods Hole, MA [C. Hamel, 508/289-7401; *admissions@mbl.edu*]

June 1998

7–13. NFCR Annu. Flow Cytometry Course: Applications in Immunobiology and Cell Biol.; Brunswick, ME [M. Munson, 207/729-6767, Fax: -5443; *mem@vsh.com*]

10–16. Genetic-Epidemiological Studies of Complex Diseases; Cold Spring Harbor, NY (appl. deadline: Mar. 15) [see contact: Mar. 4–10]

15–19. Basic Linkage Course; New York City (appl. deadline: May 1) [see contact: Mar. 16–20]

July 1998

3–11. Bioinformatics: Computer Methods in Molecular Biology; Trieste, Italy [M. Di Blas, +39-40/3757333, Fax: /226555; *courses@icgeb.trieste.it*; *http://www.icgeb.trieste.it*]

7–17. 20th WTSS. Human Genome Analysis: From Genome to Function; London [P. Faik, +44-171/403-6998, Fax: /407-5281; *wss@umds.ac.uk*; *http://www.umds.ac.uk/wlmg*]

19–31. 39th Annu. Short Course in Medical and Experimental Mammalian Genetics; Bar Harbor, ME [Jackson Lab., 207/288-6262; *education@jax.org*; *http://www.jax.org*]

25–31. 21st WTSS. Human Genome Analysis: Genetic Analysis of Multifactorial Diseases; London [see contact: July 7–17]

28–Aug. 17. Eukaryotic Gene Expression; Cold Spring Harbor, NY [see contact: Mar. 18–31] ◊

For Your Information

SBIR 1997 Human Genome Awards Announced

In July 1997 the DOE Office of Biological and Environmental Research announced three Phase I and two Phase II awards in human genome topics of the Small Business Innovation Research (SBIR) program (see box below). The highly competitive SBIR awards are designed to stimulate commercialization of federally funded research and development for the benefit of both private and public sectors. SBIR emphasizes cutting-edge, high-risk research with potential for high payoff in hundreds of areas, including human genome research. (See box at right for contact information.) ◊

SBIR Awards in Genome, Structural Biology, Related Technologies

Phase I

- Cimarron Software, Inc. (Salt Lake City, Utah): (1) An Integrated Genetic Analysis; (2) A Workflow-Based LIMS for High-Throughput Sequencing, Genotyping, and Genetic Diagnostic Environments
- Premier American Technologies Corp. (State College, Pennsylvania): A Fully Automated 96-Capillary Array DNA Sequencer

Phase II

- Genaissance Pharmaceuticals, Inc. (New Haven, Connecticut): Peptide Nucleic Acid Modular Probe Technology for DNA Sequencing and Genetic-Variation Discovery
- Promega Corporation (Madison, Wisconsin): An Engineered RNA/DNA Polymerase to Increase Speed and Economy of DNA Sequencing

Merck Genome Research Institute

■ Functional Genomics Technologies

Topics: Merck Gene Index, full mammalian cDNA clone and sequencing technology, bioinformatics, disease models, and gene expression and functional assays. Proposals accepted for 1- to 2-year projects ranging from $100,000 to $150,000 annually and innovative, high-risk pilot projects up to $50,000.

Application receipt dates: March 1, June 1, September 1, and December 1.

Contact: Finley Austin; MGRI; P.O. Box 4, WP 42-300; Sumneytown Pike; West Point, PA 19486 (215/652-4180, Fax: -6913, *mgri@merck.com*; *http://www.mgri.org*) ◊

HGN Reprinting Encouraged

HGMIS encourages readers to duplicate and reprint any part of *HGN*. Contacting HGMIS is not necessary, permission is not required, and no charge is made. We would like to have a copy of the publication, if possible. When reprinting an article, please add a credit such as "Reprinted from the U.S. DOE newsletter *Human Genome News* (for a free subscription to *HGN*, contact Betty Mansfield at 423/576-6669 or *bkq@ornl.gov*)."◊

Calendar Submissions

Items for the *HGN* "Calendar of Genome and Biotechnology Meetings" and "Training Events" should be submitted to HGMIS by mail, e-mail, or fax as soon as information is finalized (see p. 10 for HGMIS contact information). *HGN* is published quarterly.◊

Meeting Reports. In addition to these advance calendar listings, *HGN* staff welcomes reports on past chromosome workshops and sequencing, mapping, informatics, and ELSI meetings.◊

Gene Sequence Analysis Tool

A new computer-based gene sequence–analysis tool used to identify DNA regions involved in attachment to the nuclear matrix has been created by Gautam B. Singh [National Center for Genome Resources (NCGR)]. MAR-Finder, which uses statistical inference to deduce the presence of matrix-association regions in DNA sequences, is available through the center's Web site (*http://www.ncgr.org/MarFinder*).

MAR-Finder is among several new Web-based tools developed by NCGR's scientific staff to analyze gene sequences or improve access to sequences and valuable annotation in the center's Genome Sequence DataBase. Scientists in academia and private industry use the database and NCGR's bioinformatics services to conduct leading-edge research in genetics, medicine, and functional genomics.◊

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@oer.doe.gov* or 301/903-6488
- Relevant documents: *http://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 1999
- Contact: Barbara Dorsey, Oak Ridge Institute for Science and Education (423/576-9975, Fax: /241-5220; *dorseyb@orau.gov*; *http://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 laboratory.

- 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, *http://www.nhgri.nih.Gov/About_NHGRI*
- Program announcements: *http://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@oer.doe.gov*. Next DOE SBIR due March 2; STTR due December 15.SBIR, *http://sbir.er.doe.gov/sbir*; STTR, *http://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

SBIR/STTR conferences: April 3–5, 1998, San Jose, CA (800/382-4634, *http://sbir.foresnt.com*); Nov. 3–5, 1998, Boston, MA (360/683-5742, *http://www.zyn.com/sbir)* ◊

Human Genome Management Information System Subscription/Document Request* (Vol. 9, 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) _____

1. ___ *Human Genome News* ___New Subscriber ___Change of Name/Affiliation/Address (circle all that apply) ___Drop Subscription
2. ___ *To Know Ourselves* (DOE) ___DOE *Primer on Molecular Genetics* (also see *http://www.ornl.gov/hgmis/publicat/publications.html*)
3. ___ Reprint of "A New Five-Year Plan for the U.S. Human Genome Project" (*Science*, October 1, 1993) by Francis Collins and David Galas
4. ___ Reprint of "What Can the New Gene Tests Tell Us?" by Denise Casey (*http://www.ornl.gov/hgmis/publicat/judges/judgetoc.html*), *Judges' Journal* **36**(3), Summer 1997.

***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.**

SELECTED ACRONYMS

ABRF Assoc. of Biomolecular Resource Facilities
ASGT Am. Soc. for Gene Therapy
ASHG Am. Soc. for Hum. Genet.
ASM Am. Soc. for Microbiology
BAC bacterial artificial chromosome
BGA Behavior Genetics Assoc.
bp base pair
CHI Cambridge Healthtech Inst.
CLMA Clinical Laboratory Management Association
cM centimorgan
CORN Council of Regional Networks for Genetic Services
CSHL Cold Spring Harbor Lab.
DHHS NIH Dept. of Health and Human Services
DOE Dept. of Energy
ELSI ethical, legal, and social issues
EST expressed sequence tag
FASEB Fed. of Am. Soc. for Exp. Biol.
FISH fluorescence in situ hybridization
GBR Global Business Research
GDB Genome Database
GSA Genetics Societies of America
HGM Human Genome Meeting
HGMIS Human Genome Management Information System
HGP Human Genome Project
HUGO Hum. Genome Org.
IBC Intl. Business Communications
ISMB Intelligent Systems for Molecular Biology
MBL Marine Biological Laboratory
NFCR Natl. Flow Cytometry Resource
NIH Natl. Institutes of Health
NMHCC Natl. Managed Health Care Congress
NRC Natl. Research Council
NSTA Natl. Science Teachers Association
PAC P1 artificial chromosome
RECOMB Conference on Computational Molecular Biology
STS sequence tagged site
TIGR The Inst. for Genomic Res.
WTSS Wellcome Trust Summer School
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