Human Genome news

Linking Interdisciplinary Contributors and Users

Sponsored by the U.S. Department of Energy

www.ornl.gov/hgmis

ISSN: 1050-6101

Vol. 9, No. 3, July 1998

In This Issue

Santa Fe Workshop Highlights

DOE Genome Program Contractors and Grantees Present Progress at the Sixth Santa Fe Meeting, held in November 1997.

Reported by Denise Casey, HGMIS

Other Topics

Santa Fe Workshop

Eyes on the Prize: Deliver the Sequence

Complete, Accurate Sequence Most Important, Patrinos Says

Ari Patrinos
Associate Director
DOE OBER

The message delivered by Ari Patrinos last November at the Sixth DOE Human Genome Program Contractor-Grantee meeting in Santa Fe was clear and unequivocal: The Human Genome Project needs to stay focused on the commitment to obtain a highly accurate, complete human DNA sequence by 2005 (see New 5-Year Plan, p. 2). More than 400 genome program grantees, managers, and guests attended the workshop in the city of the "holy faith," high above the southwestern desert.

Patrinos, associate director of the DOE Office of Biological and Environmental Research (OBER), explained the rationale for dramatically increasing support for large-scale sequencing at the expense of other projects in the genome program. It's simple, he said. "We need to take our human genome goals seriously, or public support may evaporate and bring potentially serious budgetary repercussions."

1999 DOE Workshop Moves from Santa Fe to San Francisco Bay Area (p. 6)

(See Patrinos, p. 5)

In the News

Private-Sector Sequencing Planned

New Venture Aims for Rapid Coverage of Genome

In May, J. Craig Venter [The Institute for Genomic Research (TIGR)] announced plans to form a new company with Perkin-Elmer's Applied Biosystems Division (PE-ABD) to sequence a large portion of the human genome in 3 years for $300 million. The company plans to use several resources generated in the government-sponsored Human Genome Project (HGP). (See related article, p. 4.) A report by Venter in *Science* [**280**(5369), 1540–42] called the plan a "mutually rewarding partnership between public and private institutions."

Delivering high-quality DNA sequence to the entire scientific community at the lowest possible cost has been the ultimate goal of HGP since its inception. Although DOE and NIH program managers welcomed the promise of substantial private-sector investment, they noted that the planned venture will be more like a rough draft rather than the publicly available, detailed "A-to-Z" recipe book promised by the genome project. Positioning of sequencing subclones into the chromosomal puzzle is expected to be less certain sometimes than in more conservative strategies. Data release will be quarterly rather than immediate, as required of government-funded sequencing centers.

(see Venter, p. 4)

HGP Planning

New 5-Year Plan in Process as HGP Passes Midpoint

Midlife traditionally presents an opportunity for making evaluations, as progress toward milestones is measured and courses are altered. So, too, with the Human Genome Project (HGP), the massive 15-year biological undertaking begun in 1990 to obtain the sequence of all 3 billion bases in human DNA. Rapid progress and technology developments during the first half of the project have affirmed researchers' optimism that the task can be completed on time and within budget. A new set of U.S. goals for the next 5 years will be presented to Congress this fall.

The new plan was developed during a series of individual and joint DOE-NIH workshops on various aspects of the project. The DOE planning committee, chaired by Raymond Gesteland (University of Utah), met May 28–29 with its NIH counterpart and a broad group of 120 researchers for a final evaluation of the plan. At that meeting, the consensus was that (1) the U.S. HGP should stick with its original goal of achieving full and highly accurate human sequence and (2) improving sequencing capacity is paramount.

Priorities for the next 5 years include the following:

- Clone and sequence full-length cDNAs of humans and model organisms, especially mouse.
- Develop and improve software for determining and assembling sequences and recognizing expressed genes.
- Identify single-nucleotide polymorphisms as measures of human variation.
- Continue to study ethical, legal, and social issues related to the project.

Human Genome Project Value

Although initially controversial in the scientific community, the genome project's value has been proved beyond question. The wider biological and scientific communities in the United States and around the world are developing tools and applications for the new data in such wide-ranging fields as medicine, agriculture, bioremediation, and industrial enzymology.

International efforts have played a critical role in the project's success, with at least 18 countries now supporting programs for analyzing the genomes of a variety of organisms ranging from microbes to economically important plants and animals to humans.◊

JASON Review of DOE Genome Program

Last year the DOE Office of Biological and Environmental Research invited an independent group of physicists and engineers (JASON group, run by Mitre Corporation) to review the technology, quality-control, and informatics components of the DOE Human Genome Program. The group released its first advisory report in January. Recommendations emphasize development of advanced technology; improvements in current technology; and establishment of a standardized, quantitative program for data-quality assessment (*www.ornl.gov/hgmis/publicat/miscpubs/jason/index.html*).

A report summary by JASON member Steven Koonin (California Institute of Technology) and commentary were published in the January 2 issue of *Science*. A response by Phil Green (University of Washington) to specific points in the report was published in the February 20 issue of *Science*, and a more detailed commentary by Green is available on the Web (*www.mbt.washington.edu/ScienceLetter/JasonComments.html*).

The DOE Human Genome Program recently announced a Request for Applications for instrumentation research to develop substantial evolutionary improvements in current systems and novel revolutionary technologies for efficient genomic analyses. See p. 3.◊

DOE, NIH Discuss Informatics Goals

Since the beginning of the Human Genome Project, informatics has been widely regarded as one of the project's most important elements. The vast quantity and wide variety of generated information dictate the use of computational tools for data collection, management, storage, organization, access, and analyses.

On April 2–3, the DOE and NIH human genome programs convened a workshop in Herndon, Virginia, to identify informatics needs and goals for the next 5 years. Attending were 46 invited informatics and genomics experts and 17 agency staff from DOE, NIH National Human Genome Research Institute, NIH National Institute of General Medical Sciences, and National Science Foundation (NSF).

Both DOE and NHGRI support the philosophy that the needs of data users are foremost and must drive the goals of genome informatics. At the meeting, the wide-ranging viewpoints of large sequencing centers, smaller specialized groups, biotechnology industry users, researchers exploring comparative and functional genomics, and medical geneticists were presented (see box on medicine and genome data, p. 3).

Not all uses for these data can be anticipated today, thus implying the need for building structural flexibility into current and planned databases that support the genome project. Additionally, because knowledge will grow over time, curating the data—correcting it and adding new functional and useful links (annotation)—must be done on a continuous basis.

Meeting attendees identified priorities and made suggestions and policy recommendations on these and other issues.

Priorities and Issues

Major priorities identified by the group included the development of a reference genome map and sequence database and databases of individual variation and functional expression. Sequence data should be continuous and annotated, linked to maps, and structured to allow all conceivable

data-supported queries. Data should be updated and curated by editors. The variation database should be organized according to population and individual genotype and haplotype; it should include or link to information on individual phenotypic variation. Functional expression databases should include such pathway and regulatory data as in the databases WIT, KEGG, and EcoCyc.

Standardization. Much current data are highly heterogeneous in format, organization, quality, and content. This is not surprising, given the wide diversity of genome-research investigators who are generating the data. An identified priority is to comprehensively capture raw, summary, or processed data in standard, well-structured formats using controlled vocabularies. Additionally, databases must be integrated and linked.

Intelligent consensus standards should be defined and implemented by academia, government, and industry working together. Today, industry standards are very distinct from the few that exist in the genome project. The Object Management Group, now composed largely of industry representatives, also should involve personnel from academia and government. Explicit object definitions and access methods are needed desperately. Component-oriented software standards would promote systems integration, interoperability, flexibility, and responsiveness to change (adaptability). A balance is needed, however, between maintaining standards and allowing change and flexibility.

Tools. Tools to speed up the data-finishing bottleneck in sequencing are critical; still other tools are needed for production, research, access, annotation, data capture, functional genomics, and data mining. A Web site that collects and annotates these tools would be very useful.

Availability of Underlying Data, Especially for Individual Genotypes. Given the expense of phenotyping, the ability to see ABI traces and check on the possible association with a particular single-nucleotide polymorphism would be valuable. ABI traces are not necessary for the reference sequence because questionable regions can be resequenced.

Making Genome Data Useful in Medicine

At the informatics meeting, Anne Spence (University of California, Irvine), provided some insights into challenges facing medical geneticists who use genome data. She stressed the need to capture already discovered knowledge and make it easily available. She noted that a typical query to databases might be, "Tell me everything about gene X." Currently, this type of query would involve interrogating several Web sites, not always interlinked and often containing uncurated data of varying reliability. What medical geneticists need, she observed, is a user-friendly database of diseases and gene entries with links to other resources; regular rapid updates; and accurate, curated, and annotated information and population data.◊

Annotation. Automated annotation analyses should use clearly defined standard operating procedures, consistent application, and sufficient documentation for a more detailed understanding of particular chromosome regions. Automated annotation is a way to generate intelligent hypotheses about sequence functions and must be regarded critically as overall annotation improves with time. For this reason, human participation in the annotation process is still vitally important for getting the most out of genomic information.

Quality Checks. Attendees suggested regular checks of database quality. Users are frustrated by incorrect data and the unwillingness or inability of database providers to correct these mistakes. Official editors who curate information could resolve errors and improve data quality. Successful quality assessment at sequence centers serves as a model.

Training and Environment Issues. NSF science and technology centers are models for needed genome informatics centers. Three to five such centers were proposed to facilitate interactions among various disciplines and the training of students.

Policy Recommendations

- Open competition should be used for most database and informatics needs.
- No single database can be expected to do everything for everybody; users, however, should feel that they are interacting with only one entity. Data submission should be uniform.
- Existing frameworks such as database schema and submission tools should be used where possible.
- Model-organism databases should continue to be supported.
- Raw data should be captured to the maximum extent possible before the information is irretrievable.
- Investments should be made in optimizing and exporting software tools from genome centers.

[Daniel Drell (301/903-4742, daniel.drell@oer.doe.gov*) and Lisa Brooks (301/496-7531,* lisa_brooks@nih.gov*)]* ◊

DOE Refocuses Instrumentation Program

In April the DOE Office of Biological and Environmental Research (OBER) announced its interest in receiving new applications in genome instrumentation research for both substantial evolutionary improvements in current systems and revolutionary technologies for the post-2005 era. To stimulate contributions from investigators not previously involved in DOE's Human Genome Program, OBER invited applications from a broad range of scientists with backgrounds in biology, chemistry, physics, and engineering. At press time, preapplication response had been excellent.

DOE's transition to production sequencing has been based largely on gel electrophoresis, with data acquisition by laser-induced fluorescence. However, this in no way decreases the necessity for innovative long-term basic research in the area of instrumentation support for genome studies. In this context, OBER is refocusing its current Genome Instrumentation Program, taking stock of current progress and considering likely future needs. [See related article on Jason Review, p. 2.]

(see Instrumentation, p. 17)

In the News

Bang for the Buck: Government-Backed Research Underpins Potentially High Payoff Ventures

Spinoffs of Human Genome Project technologies continue to impact U.S. Industries, including medicine, environmental technology, agriculture, chemicals, and energy production. U.S. leadership in science and technology reaffirms the value of publicly funded research such as that supported at universities and national laboratories and in industry. Two recent spinoffs from the DOE Human Genome Program follow.

Biochip Agreement Aimed at Commercial Use

Companies to Refine Genome Technology for Mass Production

In June DOE announced that Argonne National Laboratory (ANL), Motorola Inc., and Packard Instrument Company have agreed to develop and mass-produce biochips. Motorola and Packard will contribute a total of $19 million over 5 years, making this collaboration one of the largest biotechnology research agreements ever signed by a DOE national laboratory.

Like computer chips that execute millions of mathematical operations per second, biochips can quickly perform thousands of biological reactions. "This process, developed for DOE's Human Genome Program, provides miniaturized, faster, and more economical methods to analyze DNA samples," said former Secretary of Energy Federico Peña.

"By combining biochips with robots and computers, we can find one genetic variation among 3 billion DNA bases in a matter of minutes. Conventional methods take days," said Andrei Mirzabekov, a biologist who developed the biochips at ANL and at the Russian Engelhardt Institute of Molecular Biology. "In addition to being faster than conventional gene-sequencing methods, biochips provide a 3-D platform that allows greater sensitivity and accuracy in assaying proteins, RNA, and DNA," he noted.

Argonne's contribution, in conjunction with its Moscow research partner, consists of 19 inventions related to biological microchips that have been licensed exclusively to Motorola and Packard. These inventions are the result of more than $10 million in research support since 1994 by DOE, Defense Advanced Research Projects Agency, Russian Academy of Sciences, and Russian Human Genome Program. Motorola will develop manufacturing processes to mass-produce biochips, and Packard will develop and manufacture analytical instruments to process and analyze them.

Biochips have immediate practical applications for analyzing polymorphisms, studying gene expression, and monitoring clinical trials. Richard McKernan, president of Packard, noted that within the next few years commercial biochips should bring "better, more rapidly developed pharmaceuticals; faster and more accurate medical diagnostics; a heightened ability to assess and possibly repair environmental damage; and better, more hardy, and healthier crops." The transition of biochips into the clinical diagnostics market is expected in 4 to 5 years.◊

For more information, see *www.doe.gov/biochip.htm*

Venter *(from p. 1)*

Sequencing Strategy

Plans for the new venture hinge on using BAC clone libraries being produced in the genome project. The capillary sequencing machine, now in beta testing, depends on sheath-flow detector technology (see article on Fluorescence Detector at right). Venter expects to begin large-scale sequencing next spring, with 230 capillary sequencing machines running 10 times a day and generating 100 Mb of raw bases at a cost of about $.10 per base.

Processing raw data to produce finished sequence continues to be a major bottleneck in the sequencing process. The most productive sequencing centers each generated about 20 to 25 Mb of finished sequence last year, for a total of about 120 Mb. Around 200 Mb is expected for 1998.

The Venter-PE plan is to use whole-genome shotgun sequencing, a technique developed by TIGR to sequence much smaller and less complex bacterial genomes that typically have not presented the difficulties already encountered in some regions of human DNA. These include centromeric and telomeric areas and duplicated regions found throughout the genome.

Going for the Gold

Because commercial interests will guide the path of the new venture, the focus probably will be on such potentially lucrative genomic regions as susceptibility- and disease-associated sequences that can guide the development of new diagnostic and pharmaceutical products. The company expects to seek intellectual-property protection for 100 to 300 of these sequenced regions. Generating data on biologically important genomic locations can be of great value to researchers and consumers alike. However, genome maps at the highest level of resolution—those promised by the genome project—still will be needed as the ultimate tools for scientists to embark on a thorough investigation into human biological function in all its complexity.

Genome Project to Continue on Course

When members of the DOE and NIH advisory committees and program staff met with researchers in May to review a draft of a new 5-year plan for the U.S. genome project (FY 1999–FY 2003), they reaffirmed their commitment to delivering a full and highly accurate sequence of the human genome (see New 5-Year Plan, p. 2). The approved plan will appear in *Science* in the fall.◊

Sheath-Flow Fluorescence Detection for DNA Capillary Electrophoresis

In sheath-flow systems, analytes exiting a capillary are transported in a flow of buffer and moved across interrogating laser beams. Higher sensitivity is achieved by avoiding a major source of background noise caused by direct laser action on capillaries in which DNA fragments have been separated.

Norman Dovichi [University of Edmonton, Alberta (UEA)] contributed to the development of detection systems for flow cytometry while at Los Alamos National Laboratory. Subsequently at UEA his team pioneered sheath-flow fluorescence readout technology for DNA capillary electrophoresis. In 1996 Dovichi received the American Chemical Society Award in Chemical Instrumentation for research in this project, which had some early support from DOE and major funding from Canadian sources. An article in *Science* **280**, 995 (1998) shows a Dovichi sheath-flow detector. Detailed explanations can be accessed on the Web (*www.chem.ualberta.ca/faculty/dovichi.htm*).◊

See other "In the News" articles on pp. 15–19.

Santa Fe Workshop

JGI Comes of Age: Goals, Progress, and Challenges Outlined

Scientific Director Branscomb Offers "State of JGI" Message

With its multicultural Hispanic, Anglo, and Indian heritages, Santa Fe seemed an appropriate venue for discussing the challenges of forging a union from various independent cultures. Joint Genome Institute (JGI) Scientific Director Elbert Branscomb acknowledged the formidable challenges in joining DOE's three genome centers.

John Wooley (DOE), Ari Patrinos (DOE), and Elbert Branscomb (JGI Director, Lawrence Livermore National Laboratory)

Quality First

Two 1997 reviews of JGI prompted a major redesign and sharpened goals, Branscomb said, especially for the first year, and high-quality, production-level sequencing was defined as JGI's single priority for 1998. Ambitious JGI sequencing goals are to submit 20 Mb of unique, "Bermuda-quality" DNA sequence to GenBank by October 1 (see box, Bermuda-Quality Sequence, p. 7.) [Editor's Note: As of July 7, JGI had completed 11.9 Mb, for a projected throughput rate of 32 finished Mb per year.]

JGI is committed to immediate and full public data release, with data and quality assessments computed automatically and presented in a common internal Web interface. Monthly goals and results are available on the JGI Web site (*www.jgi.doe.gov/Docs/JGI_Seq_Summary.html*).

Branscomb highlighted the importance of an ongoing review of sequencing priorities in terms of amount, quality, and cost, noting the inevitability of some problematic trade-offs. Plans are under way for JGI to participate in sequence-evaluation programs with other major sequencing centers, and ➡

Santa Fe Workshop

Patrinos *(from p. 1)*

OBER's decisions, Patrinos continued, are aimed specifically toward DOE's pledge to complete at least human chromosomes 5, 16, and 19 (about 340 Mb or 10% of the genome) over the next 7 years. He then outlined the steps DOE is taking toward its daunting goal.

Sequencing Factory on Track

DOE made the first and most important change in its genome program in late 1996 by joining genome center sequencing work at Lawrence Livermore National Laboratory (LLNL), Lawrence Berkeley National Laboratory (LBNL), and Los Alamos National Laboratory into the Joint Genome Institute (JGI). This move is aimed at exploiting individual strengths, reducing redundancy, and creating the critical mass needed. "The only way to do production sequencing on the required competitive scale is with a factory approach," Patrinos asserted. Work on JGI's new Production Sequencing Facility (PSF) began in January in Walnut Creek, California, about 35 minutes from LLNL and LBNL.

With PSF operations scheduled to begin in late fall 1998, Patrinos said he felt "very optimistic." He credited the hard work and strong support of principal scientists, senior management at the three laboratories, and DOE advisors, many of whom were at the meeting.

Sequence quality generated by JGI will conform to or exceed community standards (see box, Bermuda-Quality Sequence, p. 7) and include full and immediate data release. JGI will be held to the highest standards of quality assurance and control and database sharing. Expectations are for strong academic collaborations with sequencing centers funded under the NIH National Human Genome Research Institute, Patrinos said.

Informatics and Technology Development

An effective program of technology and informatics development is essential to success in production sequencing, Patrinos continued. He also stressed the importance of coordinating increased informatics efforts not only within the program but also with NIH and the National Science Foundation. "This is critical," he said, "for dealing with the 3 billion bp of human DNA and for the post-project challenges that will confront us in understanding the biology of long strings of sequence, a new field called functional genomics." (See box, New Awards, p. 8, and article, p. 12.)

Preparing for the Future

Patrinos expressed his strong belief that DOE's Human Genome Program is important for the future of biology, science, and society and that it requires the participation of many disciplines to bring its promise to fruition. He pointed to the Biological and Environmental Research Program's 50-year tradition of supporting a diverse portfolio of research that drives science at disciplinary interfaces where most advances occur. "DOE has done its part with training physicists, engineers, and computer scientists to strengthen the program in ways that can help us meet genome project objectives and also generate other applications," Patrinos said.◊

Fifty Years of BER Progress

- *www.er.doe.gov/production/ober/ober_top.html*
- *www.er.doe.gov/production/ober/ber50.html*
- *A Vital Legacy*: Contact HGMIS, p. 10

Santa Fe Workshop

Branscomb pledged JGI to objective cost reporting and predicted cost-efficiencies comparable to those of the rest of the community.

An 11-member external board (*www.jgi.doe.gov/Docs/JGI_Advisor.html*) advises JGI on managerial, strategic, technical, and scientific matters. JGI also seeks input from "end users" of genomic information, governmental policymakers, and the academic community.

Sequencing Strategies and Goals

Because the new JGI Production Sequencing Facility (PSF) will not begin operations until later in 1998, JGI is using capabilities and strategies already in place at current facilities. "We need to go with what we know will work," Branscomb stated.

The three laboratories have significant experience in both directed and random (shotgun) sequencing strategies:

- *Berkeley* has been a pioneer in exploiting a directed sequencing approach based on transposon mapping,
- *Livermore* has developed a highly efficient finished-sequencing operation based on shotgun sequencing in M13 and plasmid clones, and
- *Los Alamos* has extensive experience with double-end sequencing of plasmid clones to a low level of redundancy and then finishing with a primer-walking approach.

To reach the 20-Mb sequencing goal, 45% will be done at both Berkeley and Livermore and the remaining 10% at Los Alamos. A rate increase to around 35 Mb is projected by the end of FY 1998. ➡

From left, Lisa Brooks (NIH), Lloyd Smith (University of Wisconsin, Madison), and Jeff Schloss (NIH)
(Photo submitted by Ken Beattie, Oak Ridge National Laboratory)

1999 DOE Workshop Set for California

The seventh DOE Human Genome Program Contractor-Grantee workshop will be held in January 1999 in the San Francisco Bay area of California. At least one investigator from each funded project is expected to attend the entire meeting and represent the project at poster sessions. Some projects also will be represented in platform presentations. Abstracts should be submitted through the Web site (*www.lbl.gov/Conferences/DOE_HGP*).

Abstract deadline: October 1.
Contacts: Sylvia Spengler or Kelcey Poe (*kjpoe@lbl.gov*, 510/486-4879, Fax: -5717).◊

JGI Sequencing: Moving Toward a Consensus Strategy

Chris Martin (LBNL) and Jane Lamerdin (LLNL) outlined progress toward forming efficient collaborations among the member laboratories and establishing a facility capable of sequencing 100 Mb per year. They observed that this ramp-up will require increasing the scale of operations (hiring more people) and improving the processes in terms of cost, throughput, and quality.

At LBNL, emphasis is on the up-front shotgun phase, using double-end plasmid subclone sequencing and moving toward increasingly automated finishing processes. The primary templates are BACs for chromosome 5, with gaps finished via transposons. Martin reported working closely with LANL to establish sequence-annotation and -submission processes. LBNL ships completed clones to LANL, where restriction-digest verification required for JGI quality is performed.

At LLNL, sequencing targets are chosen from regions of the high-resolution chromosome 19 physical map, which offers 42 Mb (fivefold to sixfold coverage) of sequence-ready cosmid and BAC clones, Lamerdin said. About 90% of genomic targets are 1 Mb, and the remaining 10% are regions of at least 100 kb that contain genes of interest. All regions are flanked by publicly available markers. JGI targets can be found via the National Center for Biotechnical Information (NCBI) Web site (*www.ncbi.nlm.nih.gov/HUGO*).

Investigators follow a standard shotgun strategy using M13 and plasmid vectors and a mixture of sequencing chemistries, with automated base calling and clone sequence assembly by Phred and Phrap. Sequences are edited using the CONSED package, and gaps are closed with a variety of techniques, including PCR and an in vitro transposon technology. Each sequence is verified relative to three independent restriction digests, and sequence analysis and annotation are done using local BLAST and GRAIL coding-prediction tools. Results are parsed manually and submitted through SEQUIN to NCBI.

Lamerdin said the member laboratories are instituting quantitative quality standards for JGI, which will be adopting these processes for all generated sequence. (For details, see box, Bermuda-Quality Sequence, p. 7.) This standard will enable data to be compared within JGI and with sequence generated by NIH centers. David Nelson is monitoring data-quality processes and helping to implement them. All data are available on the Web site, and submitted data are in GenBank. All sites have implemented a Web-accessible list of accessions, submit lengths, and Phrap-quality statistics. Because investigators are focusing on genomic continuity, they also have a "unique length" category indicating sequence new to the database.

Martin and Lamerdin also described some successful methods for increasing sequence-data quality and throughput and for decreasing costs. At LBNL, significant improvements in automation were made with commercial Tecan and custom Prep Track robots. Other changes include switching to a Qiagen real prep for DNA preparation before the shotgun phase, using BigDye Terminators in sequencing reactions, and moving to 64-lane gels. Changes at LLNL instituted by Paula McCready include incremental chemistry improvements (i.e., use of the polymerase Omnivase from Promega), in-house reagent quality control for all buffers, and production and finishing operations. In collaboration with JGI, the Whitehead Institute is developing front-end automation and sequencing on chromosome 19.◊

The sequencing goal for FY 1999 is 20 to 24 Mb of high-quality sequence and an additional 70 to 80 Mb of draft sequence. Branscomb observed that sequencing goals and cost economies will not be achieved unless, within 3 to 4 years, PSF is generating well above 100 Mb of Bermuda bases per year.

A Consensus Strategy for PSF

Branscomb outlined the basic plan, which focuses on using a combined shotgun strategy with directed-sequencing approaches. PSF was designed to be flexible to accommodate changes in technology, and optimization will begin after the first year. The initial sequencing targets are chromosomes 19, 16, and 5, which have been mapped largely by the three member laboratories.

Sequencing Support

Activities in support of production sequencing at Los Alamos, Berkeley, and Livermore will include sequencing-technology development, large-insert clone production and mapping, sequence-annotation submission, and overall informatics support and technology development. In general, PSF will obtain sequence-ready clones from the three laboratories and return assembled sequence.

Branscomb emphasized the need for well-designed informatics that integrates (within a single functional entity) support for the work at the different sites, especially PSF. Four goals for informatics are to achieve as much uniformity as possible in practices and tools, provide seamless management of and access to critical data across all sites, maintain organizational and administrative unity and coherence, and define a uniform role for informatics and computational approaches in support of quality maintenance.

Beyond Production Sequencing

The ultimate value of high-throughput sequencing depends largely on what is learned about the revealed genes. Logical adjuncts to production sequencing, therefore, include the generation of full-length or nearly full length cDNA sequences, obtaining various kinds of expression data in mouse and human, and

JGI and "Bermuda-Quality" Sequence

The international sequencing community holds an annual meeting in Bermuda, sponsored by the Wellcome Trust, DOE, and NIH to set standards for DNA sequence with respect to cost, quality, timeliness of submission, and level of annotation. For details, see the Web site (*hugo.gdb.org/bermuda2.htm*).

Standards for JGI meet or exceed those for "Bermuda-quality."

- *Sequencing targets*: Megabase-sized (or larger) regions are the preferred targets to maximize biological impact.
- *Coverage*: Goal is to complete sequence continuity across a target region, as feasible (no more than 1 gap in 200 kb on average).
- *Sequence accuracy*: The acceptable error rate in finished sequence is 1 in 10,000 bases. JGI is using a rules-based approach for achieving this standard, which requires a minimum Phrap consensus value of 40 for each base and greater than 95% double-stranded coverage, with a minimum coverage of 2 high-quality reads with 1 read on the opposite strand.
- *Clone assembly verification*: Two independent approaches will be used to verify accuracy of a clone's finished sequence.
- *Data submission and annotation*: Minimum submission is the size of the starting clone, with 95% of sequence represented on both strands and all ambiguities resolved; sequences will be annotated to the extent feasible at the time of submission, largely automated; immediate release of finished annotated sequence.◊

JGI Clone Resource Task

Large-insert clones for sequencing are being mapped, selected, and validated at JGI member laboratories. Jan-Fang Cheng (LBNL) discussed JGI development of clone-based maps for production sequencing; other team leaders are Norman Doggett (LANL) and Anne Olsen (LLNL). Among the major goals are the complete closure and validation of existing maps of chromosomes 5, 16, and 19; selection of new mapping targets; and creation of pools and high-density filters of newly approved BAC libraries for STS screening.

Cheng observed that this year's challenge is to build templates for the FY 1999 ramp-up and beyond. Optimistic goals are to generate 70 Mb of contigs longer than 1 Mb, with associated restriction mapping data and over tenfold genome coverage. Summarizing the current map status of chromosomes 16, 19, and 5, Cheng reported that more than 70 Mb of contigs greater than 300 kb has been generated, with over 25 contigs (40 Mb) larger than 1 Mb.

Cheng pointed out that before JGI was established, the three laboratories used different clone types as well as different mapping approaches. Currently, the mapping plan calls for three autonomous map-production teams, although ideally one major production site will generate templates for PSF; a centralized clone repository will be set up with 1-Mb contigs prioritized for sequencing.

Breakdown for the clone resource task is resource production (15%), pilot R&D projects (15%), STS-content mapping (35%), and restriction mapping (35%). Cheng observed that, unlike sequencing strategies, mapping techniques evolve quickly and scientists want to retain flexibility.◊

Linda Ashworth (JGI, Lawrence Livermore National Laboratory)

Santa Fe Workshop

obtaining mouse genomic sequence of all homologous regions conserved between mouse and human. On the last topic, Branscomb noted that "there are areas in the human genome where we already know the cost of sequencing will pay off richly, and in those same areas we'd like to know all the conserved elements in mouse—a critically important thing to find out; the question is how to do that affordably when we can't afford to sequence the mouse right now in parallel with human."

Branscomb pointed to another trade-off for high-quality and high-productivity sequencing goals: Postpone originally planned efforts to "functionalize" the sequence data (that is, annotate it with additional, experimentally derived information to make it more useful to biologists). Investigators are exploring ways to do comparative sequencing in the mouse at a much lower level of redundancy and quality to locate and sequence mouse-human conserved regions accurately enough to learn their biological significance. In the first year, researchers will perform a small amount of mouse physical mapping to support future mouse-human comparative genomic sequencing if it proves affordable.

DOE OBER also has funded a pilot project to functionalize fruit fly sequences. "The practical challenge," Branscomb said, "is to find out what scaleable methods can be devised to annotate fly sequence data for about $1000 per gene or less. It's an interesting challenge." ◊

New DOE Awards for Collaboration with JGI

Following two competitions, DOE awarded grants for collaborations with JGI. The awards, announced in February and March, went to investigators at universities, industries, and national laboratories to provide additional technologies, expertise, and resources for human genome research and functional genomics.

- Jack Barber (Immusol Inc.) and Gerald Rubin (University of California, Berkeley): Grants for research on human gene function via use of ribozymes and comparison with *Drosophila*, respectively.
- Edwin Bradbury (LANL): Integrated approach to functional genomics (phage display library).
- Ron Davis (Stanford University), Trevor Hawkins (University of Florida, Gainesville), and Skip Garner (UTSWMC): Grants for instruments and technologies to increase sequencing efficiency and reduce costs.
- Eric Lander (Whitehead Institute for Biomedical Research): Automated system to prepare DNA samples for sequencing and perform sequence-data analysis.
- Ed Michaud and Tuan Vo-Dinh (Oak Ridge National Laboratory): Germline deletion complexes in embryonic stem cells for mapping gene function in mouse-human homology regions.
- Lisa Stubbs (LLNL): Resource libraries (full-length mouse cDNAs).
- F. William Studier, John Dunn, Joel Sussman, and Otto Ritter (Brookhaven National Laboratory): Systematic determination of archetypical structures for protein families.
- David Torney (LANL): Annotation of coding DNA with protein domains.
- Ed Uberbacher, Richard Mural, and Manesh Shah (ORNL): Computing infrastructure for large-scale functional annotation of DNA sequence.◊

JGI Informatics: Tracking a Moving Target

Providing informatics support for achieving "dream" targets of 100 Mb a year of Bermuda-quality sequence is an evolving process, said Tom Slezak, director of the JGI informatics team. "It can't happen in a single leap. It will ride a learning curve similar to all the other scientific and technological ones going on in parallel," he said. But a lag time for informatics support on these processes is inevitable because support requirements are not yet clear.

In general terms, informatics for sequencing encompasses clone resources and validation, sequence production, sequence analysis and annotation, informatics integration, and systems administration. Slezak gave an overview of various short-, medium-, and long-term solutions being implemented to meet these challenges.

Focusing on JGI's charter and scope, investigators are working toward jointly designing, developing, and deploying processes; and sharing hardware, software, expertise, and production goals. Prime areas of informatics concern include improving sequence quality-control tracking and reporting, increasing automation of finishing tasks, linking mapping data with sequencing, and starting some pilot informatics projects to support early functional genomics efforts. As to the last, Slezak cautioned, "We'd better get a head start on these, or we're going to end up with five different spreadsheets or other forms of information and a lot of duplication."

The JGI informatics team is using systems already in place at the three participating laboratories, sharing code, and standardizing where feasible. "The three sites will meet on the Web," Slezak said.

Medium-range goals (FY 1999–FY 2000) include adapting the most robust JGI shotgun system to scale up to 20 to 40 Mb per year. Major challenges are to modify process changes and automate sequence finishing. Slezak observed that large increases in automation present severe informatics challenges, such as moving from many small robots to automating individual processes to an automated factory. He also noted the difficulty of predicting how much these efforts will speed up finishing, and he warned that the system will hit inherent limits.

Turning to long-term plans, Slezak observed, "We can't scale up by bolting things together." A complete overhaul of the entire process—the biology, automation, informatics, perhaps even management—will be needed, he said. One promising avenue for meeting long-term goals is participation by commercial and academic groups in developing some automation and laboratory information-management systems (see New Awards box above). Other long-term goals are to develop centralized databases, connect them with existing ones, and provide capabilities for Web-based navigation and display of all the data from any starting point.◊

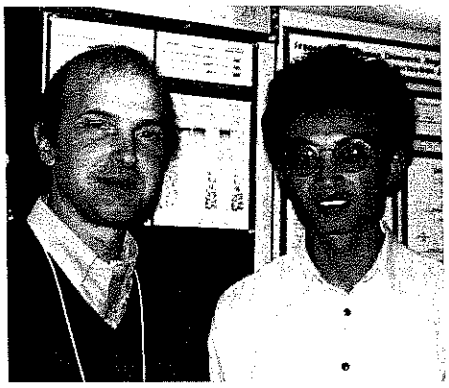
Santa Fe Workshop

Much Progress Reported in All Areas

Researchers Outside JGI Describe Work on Sequencing, Resources, Functional Genomics, Informatics, ELSI

The Santa Fe workshop offers all DOE genome grantees a unique opportunity to discuss and share research successes, problems, and challenges as well as new material resources and software capabilities. The meeting also gives scientists and administrative staff an overview of the program's progress and content, a chance to assess the impact of new technologies, and a forum for initiating collaborations.

Brief summaries begin below of several plenary talks on research outside the Joint Genome Institute (JGI). *See p. 6 for research specific to JGI.*



From left, Stanley Tabor (Harvard Medical School) and Yanlong Li (University of California, Berkeley)

Sequencing Strategies and Tools

Meeting Human Genome Project sequencing goals on time and within budget will require major improvements in speed, reliability, and costs. Last year, the most efficient sequencing centers achieved outputs of around 20 to 25 Mb at a cost of about $.50/bp, with a total community-wide sequencing capacity close to 100 Mb/year. Projections are that several hundred megabases of finished sequence will have to be generated each year to meet goals by 2005. Workshop speakers presented some successes and challenges they are encountering.

Large-Scale Projects

Shawn Iadonato [University of Washington (UW)] discussed the advantages of using detailed sequence-ready restriction maps produced with a high-resolution, multiple complete-digest method. He presented data from a large-scale project, based on Eric Green's (NIH) STS and YAC-based map, to sequence a contiguous 2-Mb human chromosome 7q31.3 region. Advantages of investing in upstream mapping, he noted, include (1) clone validation, (2) assembly checking, and (3) optimal tiling path. In addition, UW emphasizes high-quality raw and finished sequence data.

Iadonato also suggested a way for large-scale sequencing facilities to measure sequencing cost and efficiency. He divided the factors into three areas: development activities universally applicable to the genome community, typically the smallest investment; production-related development that enhances a particular facility's efficiency; and production of mapping and sequencing data, generally the largest investment. The real measure of cost efficiency, he said, is the number of finished base pairs per read of generated data; at the UW center it fluctuates between 45 and 55 bp.

Owen White [The Institute for Genomic Research (TIGR)] reported on high-throughput microbial sequencing projects at TIGR and on implementation of enhanced data-annotation techniques. "Biology is not just data acquisition," he reminded attendees; "it also attempts to draw relevant conclusions." White observed that sequence generation and annotation are coupled tightly and that people generating sequence should be enhancing their submissions with data such as database and orthology matches. (Orthology refers to genes occupying the same genetic locus in different species.) He noted that "orthologs are the central kernel of information we will be using instead of individual genes." Other possible annotation includes frame shift analysis, laboratory management database systems, noncoding information such as DNA repeats and regulatory regions found upstream of a gene, and literature citations. Because an increasing number of genome sequences are coming online, a robust, flexible system of data management across genomes will be needed to handle the numerous kinds of data. Development of such a system will enable new entries to update the annotation of other genomes as applicable.

Chemistries, Strategies, Technologies

Current DNA sequencing methods use a DNA polymerase to extend a primer in the presence of the four natural nucleotides. Two important polymerase properties are its ability to incorporate dideoxynucleotides onto a growing DNA strand and the length of time the polymerase remains associated with the DNA template (known as its processivity). For 12 years, Stanley Tabor (Harvard Medical School) and colleagues sought to capture a picture of the replicating complex in action; their efforts were rewarded last year with an elegant determination of a T7 DNA polymerase structure at a 2.2-Å resolution. The T7 is locked in a replicating complex with a dideoxy-terminated primer-template, an incoming dNTP, and the processivity factor thioredoxin. The work was reported in the January 15, 1998, issue of *Nature* (*www.eurekalert.org/releases/3dna-rephms.html*).

"The structure has been a gold mine for helping us understand the polymerization mechanism and for facilitating further studies to define critical features that will enable more precise engineering of mutant polymerases with enhanced properties," Tabor said. The group's past successful applications of structural studies include development of an improved polymerase, now commercially available, which reduces the amount of

expensive reagents required and supports popular cycle-sequencing protocols.

John Dunn (Brookhaven National Laboratory) reported on the development of vectors and protocols to allow simple and reliable production of nested deletions for rapidly sequencing across one strand of a cloned fragment using a universal primer. The strategy has advantages for sequencing gaps and repetitive DNA. Exploratory work has demonstrated its effectiveness in sequencing fragments at least as large as 17 kb, cloned from a human BAC. Imaging and sizing software is being tested for automated selection of an appropriate set of deletions for sequencing.

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
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
www.er.doe.gov/production/ober/hug_top.html

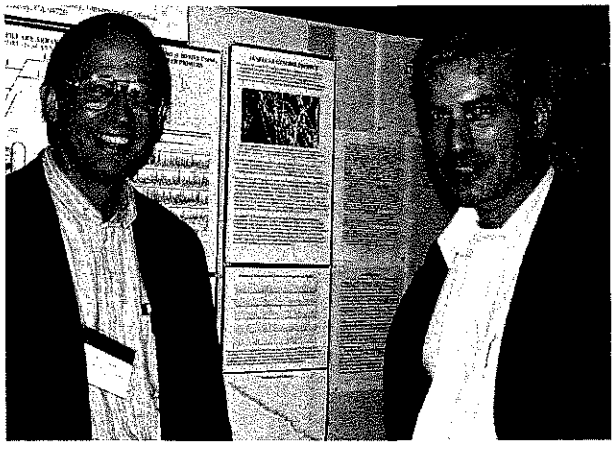
Contact: Daniel W. Drell
301/903-6488, Fax: -8521
daniel.drell@oer.doe.gov or *genome@oer.doe.gov*

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.◊

Alex Glazer [University of California, Berkeley (UCB)] discussed improvements resulting from the use of energy-transfer (ET) fluorescent primers for DNA sequencing and analysis. Fluorescent labels are critical components of conventional automated sequencing approaches, and ET primers provide more distinct and intense fluorescence emissions than single dye-labeled primers. This improvement has led to significant advances in DNA sequencing and analysis, including short tandem repeat (STR) typing often used in diagnostics and forensics.

Glazer described a collaboration that he and Richard Mathies (UCB) have with David Sidransky (Johns Hopkins University) in which two-color ET primer sets are applied to bladder-cancer diagnosis. The technique is based on electrophoretic analyses of PCR-amplified STRs from bladder epithelial cells shed in the urine. Diagnosis depends on detecting loss of heterozygosity (variation) at particular loci, and multiplex analyses allow quantitative determination of amplified fragments from two different samples (normal and tumor cell). The noninvasive assay facilitates the monitoring of surgery's effectiveness in eliminating cancer cells and the detection of a relapse.

Significant increases in sequencing throughput can be achieved using higher electric fields in the fragment-separation (electrophoresis) step. Although conventional slab gel systems retain too much heat under these conditions, sets of gel-filled glass capillaries that dissipate heat more efficiently are being developed as an effective alternative to conventional methods. Another advantage to capillary systems is the potential for eliminating the labor-intensive gel-pouring and -loading steps.



From left, Richard Mathies (University of California, Berkeley) and Norman Doggett (Los Alamos National Laboratory)

Several groups discussed advances in various approaches that use many capillaries in parallel [called capillary array electrophoresis (CAE) systems]. Barry Karger (Barnett Institute, Northeastern University) discussed the use of replaceable polymers and capillary electrophoresis with high-resolution automated fraction collection for picking out differentially expressed mRNAs or cDNA systems. Indu Kheterpal (UCB) reported progress in developing a second-generation CAE scanner with Ron Davis (Stanford University) that can detect up to 1000 capillaries in an array. Jian Jin [Lawrence Berkeley National Laboratory (LBNL)] described a beta test version of a 96-well capillary system that employs a sheath-flow excitation-detection geometry. A prototype of a fully automated 96-well system is ready for testing, according to Qingbo Li (SpectruMedix Corporation, formerly Premier American Technologies Company). Ed Yeung (Iowa State University) won an R&D 100 award for developing the technology.

Chip-based CAE approaches are being explored by groups such as the team led by Mathies. Their device, featured in the February 15, 1998, issue of

Web Shows Worldwide Sequencing Progress

- Genome-MOT: *www.ebi.ac.uk/~sterk/genome-MOT*
- Human Genome Sequencing Index: *www.ncbi.nlm.nih.gov/HUGO*
- Genome Channel: *compbio.ornl.gov/gac/datactrs.shtml*
- Jared's Molecular Biology Page: *weber.u.washington.edu/~roach/human_genome_progress2.html* ◊

Analytical Chemistry, combines an electrophoretic injection and separation system with an electrochemical detector in a microfabricated apparatus. The technology represents the first example of integrating onto a single chip a miniaturized detection system with injection and separation components of an electrophoretic chemical-analysis system.

Sequence-Ready Map Strategy

Because of their higher stability as compared with their YAC or cosmid counterparts, clone libraries constructed in BAC, PAC, and P1 vectors have become the choice for clone sets in high-throughput genomic sequencing projects. A strategy was proposed in 1996, and pilot projects were begun for using end sequences from BACs or PACs to support just-in-time contig extension for directed sequencing [*Nature* **381**, 364–66 (1996)]. The strategy requires collection of end sequences from clones representing a 15-fold coverage. DOE-funded pilot projects are being carried out at TIGR and UW.

Mark Adams (TIGR) provided a progress report on a pilot BAC end sequencing (BES) project to explore the strategy's feasibility, optimize technologies, establish quality controls, and design the necessary informatics infrastructure. Adams reported a success rate of around 75% and, using four ABI 377 sequencers, daily production of about 400 high-quality BAC end sequences having an average edited length of about 475 bases. Researchers are running into some large duplicated regions in the chromosome 16 end-sequencing project (about 40 kb in one BAC), and Adams stressed the importance of understanding the targeted region's genomic structure. So far, about 20% has been accomplished toward the goal of 15-fold genome coverage.

Details of prep and sequencing methods are on the TIGR Web site (*www.tigr.org*), and all components are available commercially. Data from a BES companion project led by Gregory Mahairas (UW) are also on the Web (updated information on BES and BAC-PAC resources: *www.ornl.gov/meetings/bacpac/95bac.html*). ◊

Santa Fe Workshop

Gene-Discovery Resources

Over the years, the team led by Bento Soares (University of Iowa) has optimized methods for producing cDNA libraries, a technically challenging undertaking, and is continuing to produce libraries of the highest quality. Individual clones from these libraries have been arrayed at Livermore and distributed worldwide for characterization by the international I.M.A.G.E. (Integrated Molecular Analysis of Gene Expression) Consortium. To date, over 3 million I.M.A.G.E. clone replicas have been sent to more than 1000 laboratories worldwide; end users analyze the clones and return data on them (*www.ncbi.nlm.nih.gov/dbEST/index.html*).

At the Santa Fe meeting, Soares described further progress in developing cDNA libraries and serial subtractive hybridization strategies (within and across different pooled libraries). These strategies are expected to minimize redundancy and identify cDNAs not yet represented in publicly available collections of human, mouse, and rat cDNAs.

Soares observed that finding novel cDNAs is increasingly challenging as researchers approach completion of human and mouse gene-discovery efforts. His group uses pools of I.M.A.G.E. clones, from which ESTs have been derived, as drivers in hybridizations with single or multiple normalized libraries, thus generating subtracted libraries enriched for new cDNAs. Sequence analysis of two subtracted libraries indicated a fourfold reduction in representation of the driver clones. ➨

Santa Fe Workshop

Sequencing at NIH NHGRI

Lisa Brooks and Jeff Schloss [NIH National Human Genome Research Institute (NHGRI)] updated attendees on current projects and future plans for NIH-supported large-scale sequencing. Third-year awards for pilot projects were made this summer after a review that included sequence-quality assessment of clones selected by NHGRI staff. Two checker groups reassembled trace files, assessed the assembly quality overall and at the single-base level, and then sequenced to resolve discrepancies found in the GenBank record. The original sequencer had the opportunity to review and respond to the evaluation.

NHGRI released an RFA in January (updated in June) for participation in a cooperative research network comprising three areas: sequence-production centers; specialized centers to sequence difficult regions or close gaps, for example, or test new technology, methodology, and instrumentation; and a quality-control center (*www.nhgri.nih.gov/Grant_info/Funding/rfa-hg-98-002.html*). Awards are expected to cover up to 3 years for specialized centers and the quality-control center and up to 5 years for production facilities. The network's goal is to complete 1.8 billion bp (60%) of human DNA sequence by 2005, a rate that will require 300 Mb of finished sequence annually between 1999 and 2005. NHGRI will set aside $80 million a year for production sequencing.

Brooks discussed the need for genomic-scale technologies and pilot or large-scale projects for the discovery and scoring of single-nucleotide polymorphisms (SNPs). An RFA was subsequently released by NIH in January (*www.nhgri.nih.gov/Grant_info/Funding/rfa-hg-98.001.html*). Analysis of such DNA sequence variation is an increasingly important source of information for identifying genes involved both in disease and normal biological processes such as development, aging, and reproduction. A public SNP database, under development at the National Center for Biotechnology Information, is expected to come online later this year.◊

Santa Fe Workshop

Charles Auffray (Centre National de la Recherche Scientifique) reviewed the goals of EURO-IMAGE, which 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. The European consortium is integrating its efforts with those in the United States and Japan. A meeting of I.M.A.G.E. participants is tentatively planned for March 1999 in Kazusa, Japan.

Auffray outlined some major challenges of systematic large-scale efforts to obtain the human "transcriptome" (the complete collection of all unique sequenced gene transcripts). He also discussed generating insights into gene function by exploring similarities and linking the information to the proteome (complete set of proteins; see also p. 20).◊

From left, Glen Evans (University of Texas Southwestern Medical Center) and Debbie Nickerson (University of Washington, Seattle)

Santa Fe Workshop

Functional Genomics

Deciphering the function of each human gene is a daunting prospect that will continue far past the projected 2005 deadline for completing the Human Genome Project. Researchers use model organisms such as the laboratory mouse to help guide these explorations. In Santa Fe, Eddy Rubin (Lawrence Berkeley National Laboratory) and Monica Justice (Oak Ridge National Laboratory) discussed the value of detailed comparative mouse and human linkage maps and the technologies available for manipulating the mouse genome.

Rubin described an in vivo approach for tracking down gene function. His group has generated YAC and P1 transgenic mice panels that contain defined contiguous regions of the human genome. The mice are assessed for a phenotype attributed to a candidate genomic region such as the human chromosome 21 Down's-syndrome critical region, which has been linked to learning and behavior. Using three lines of transgenic mice containing up to 2 Mb of human DNA, the group identified a human gene that affects learning in mice. Supporting evidence points to the gene's expression in the developing mouse nervous system and also to a homologous fruit fly gene shown to impact learning in that insect as well. Other potential applications for in vivo libraries include mutation cloning and identifying genes rather than just map locations (quantitative trait loci) in the study of such common complex human disorders as asthma and schizophrenia.

Justice spoke about the potential for studying human disease by using deletions of mouse genomic regions coupled with chemical mutagenesis using ethylnitrosourea (ENU), a supermutagen that primarily causes single base changes. She observed that a series of overlapping deletions generated by ENU can be useful for generating fine-structure physical maps, gene cloning, and further mutational analysis of the region. Her group's work focuses on the mouse chromosome 7 albino (c) region, which is homologous to human chromosome 11q13-q21. This region is linked to the mouse homologue of the human disorder oculocutaneous albinism. Justice showed dramatic photographs of a baby and mouse, both bearing the disorder's characteristic coloring. Using this paradigm, she observed that such mouse-human comparative analyses, with additional induced and targeted deletions and chemical mutagenesis, will provide essential data for large-scale expansions of the mouse functional map in parallel with human gene maps.◊

Santa Fe Workshop

Data Surge Challenges Informatics Developers

The explosive growth of sequence and biological information poses pressing challenges for data acquisition, representation, access, and analysis. Some highlights from informatics sessions at the Santa Fe workshop follow.

bioWidgets: Adaptable, Reusable Modules for Viewing Data

Many software analysis applications commonly are tailored to fit resources available at a particular site. The bioWidgets toolkit philosophy of Chris Overton's team [University of Pennsylvania (Penn)], however, is to use a component-based approach to design adaptable and reusable software, easily incorporated in a variety of applications and deployable in modules, that promotes interaction among applications. Jonathan Crabtree described the team's efforts to develop and deploy graphical user interfaces for visualizing molecular, cellular, and genomic information. The current implementation includes widgets that display sequences, maps, BLAST results, chromosomes, and sequence alignments. The group also is developing interfaces for data stored in such distributed heterogenous databases as the Genome Database, Genome Sequence DataBase, Entrez, and ACeDB and is creating a consortium of bioWidget developers and users to create standards. All bioWidgets are implemented in Java for Web distribution (*agave.humgen.upenn.edu/bioWidgetsJava*).

Querying Across Databases with BioKleisli

Sue Davidson (Penn) described a new suite of tools that permits researchers to pose complex questions over the distributed, heterogenous sources housing most genome-related data. ➨

Answering the query, "Find human sequence entries on human chromosome 22 overlapping q12," for example, would now require access to three separate databases. The new system, which performs integration "on the fly" while allowing simultaneous structural source-data transformations, is based on the powerful Kleisli integration system developed at Penn. Together with the high-level Collection Programming Language (called CPL), bioKleisli can be used to integrate data through dynamic user-defined views or to create specialized data warehouses allowing fast access (*www.pcbi.upenn.edu*).

Improved BCM Search Launcher

Kim Worley [Baylor College of Medicine (BCM)] reported on the enhanced sequence-analysis search services provided by the BCM Search Launcher (*gc.bcm.tmc.edu:8088/search-launcher/launcher.html*). Search Launcher is an easy-to-use interface that organizes Web sequence-analysis servers according to function and provides a single point of entry for related searches. It adds hypertext links for easy access to Medline abstracts, related sequences, and other information. A BLAST Enhanced Alignment Utility (BEAUTY) tool makes it easier to identify weak but functionally significant matches in BLAST protein database searches. Recent enhancements make BEAUTY searches available for DNA queries (BEAUTY-X) and for gapped alignment searches (using WU-BLAST2). For users who need to perform a particular search on a number of sequences at once, the Batch Client provides access to all searches available from the BCM Search Launcher Web pages in a convenient drag-and-drop (Macintosh) or command line (UNIX, PC) interface. Future developments are focusing on the analysis of large-scale sequences to support the efforts of the Genome Annotation Consortium (see sidebar above).

WIT/WIT2: Reconstructing Metabolism

Analysis of the increasing number of fully or partially sequenced small genomes can serve as the foundation from which to look at more complex genomes. Evgeni Selkov and Ross Overbeek (both at Argonne National Laboratory) discussed the reconstruction of accurate metabolism models for 29 of these small organisms. Using sequence data supplemented with biochemical and phenotypic data, the group has made reconstructions (some based on still-incomplete sequence data) available via the WIT/WIT2 system. WIT2 is a UNIX-based system in two parts: a Web-based, data-access system and a set of batch tools offering extensible data-query access (*wit.mcs.anl.gov/WIT2/wit.html*).

(see Informatics, next page)

Genome Annotation: Informatics Advances Needed for Age of Functional Genomics

Sharply increasing rates of sequence-data production are placing greater and greater demands on information systems for new ways to view and better understand the meaning of the growing strings of As, Cs, Ts, and Gs piling up in GenBank and community databases (see also p. 2). Enriching data with such information as gene features and locations, gene-control regions, related sequences, gene-expression patterns, gene and protein families, pathways, and phenotypes can help pave the way for a successful transition from the current structural genomics phase of DNA mapping and sequencing to functional genomics studies.

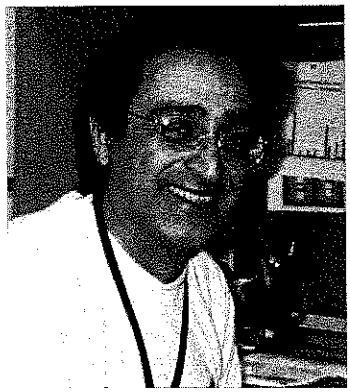
Genome Annotation Consortium

Ed Uberbacher (Oak Ridge National Laboratory) described several pilot projects in the multi-institutional Genome Annotation Consortium, which was established to minimize some problems posed by genome-scale sequencing and to build a shared infrastructure for integrating diverse biological information. Four basic components of the pilot projects are daily sequence and biological data acquisition from 19 major genome centers; automated data analysis to link biological information to sequence using tools for exon prediction, gene modeling, and sequence comparison; a storage, maintenance, and update component; and a series of methods for browsing, querying, and accessing other tools of value to researchers. An important goal is to build a level of interoperation using CORBA, which has not yet been implemented into the system.

In outlining some current challenges in sequence annotation, Uberbacher noted that no community-wide annotation processes exist and that much of the annotation does not describe the methods and evidence used to create the data. Moreover, even if the sequence were annotated extensively when submitted to the database, long-term update and maintenance are challenges. New ESTs that may be important to understanding a genomic region of interest, for example, may have been entered into the dbEST database but are not represented in the original annotation. Annotation by end users is difficult because it requires multiple tools that use different formats and lack interoperability.

Genome Channel

Uberbacher also demonstrated the Genome Channel, a prototype graphical user tool for browsing and querying the annotated reference genome (*compbio.ornl.gov/tools/channel/index.html*). The Java interface relies on a number of underlying data resources, analysis tools, and data-retrieval agents to provide an up-to-date view of genomic sequences as well as computational and experimental annotation. Designed to be simple enough for a layperson, the channel also offers sophisticated capabilities for hypothesis testing. The system had about 6000 GRAIL-EXP and 4000 GENSCAN predicted human genes as of June.◊



Morey Parang (Oak Ridge National Laboratory)

Santa Fe Workshop

ELSI Grantees Address Accelerated Societal Impact of Genome Data

Rapid worldwide progress in human genome sequencing has heightened the urgency of addressing the many complex ethical, legal, and social issues (ELSI) surrounding genetic data. Some topics presented at the Santa Fe workshop are summarized below.

Testing, Managed Care, and Confidentiality

Jeroo Kotval (University of Albany) spoke about the threat to patient welfare and confidentiality created by the confluence of three elements: DNA-based testing, the rise of market-driven managed-care organizations (MCOs), and the availability of the medical record on networked computers. Confidentiality of DNA-based tests provides new and heightened concerns because some of these tests can predict future healthcare costs and also implicate relatives. Such information handlers as secretaries and data-entry clerks are not licensed professionals and, therefore, not bound by medical confidentiality laws in most states, she pointed out.

Central to confidentiality concerns in the MCO setting, Kotval suggested, is the practice of utilization review, which tracks each physician's referral and test-ordering practices and sometimes even treatment protocols. She observed that the MCO setting presents some unique ethical dilemmas because physicians and other personnel are MCO employees or contractors and because payor and provider functions are contained within the same entity. Physicians, no longer free agents, may be caught between competing MCO and patient interests.

She suggested that traditional concepts of interpersonal morality with regard to confidentiality may not apply to institutional decisions because institutions are not moral beings. Organizational decisions are made for the institution's good, and values implied by such decisions may differ from those held by individuals in their personal lives.

DNA-based predictive tests for adult-onset disorders or predispositions may, therefore, be used by insurance companies to discriminate in the interest of cutting costs. Kotval emphasized that cost-tracking is not restricted to for-profit, market-driven managed care.

Kotval's group seeks to (1) understand the context in which DNA-based tests will be used by MCOs, (2) identify policy gaps that could allow misuse of confidential medical information, and (3) make practical recommendations to remediate these gaps. She stressed that genetic information increasingly will be an inseparable part of the medical record. If individuals are to avail themselves of the benefits of genetic testing, however, they must be assured that the medical record is confidential.

"In the popular imagination," she said, "one's genetic makeup is perceived as fundamental and integral to the self, revealing something deep, basic, and even final about a person, adding to [the genetic data's] sensitivity and raising concerns about its possible misuse. Our genes are fraught with both personal and cultural significance."

Anguish of Genetic Testing

Gene testing's profound challenges to a person's sense of self, family, and future were well illustrated in *A Question of Genes*, last fall's 2-hour nationally televised Public Broadcasting Service special sponsored by the DOE Human Genome Program and SmithKline Beecham. At the Santa Fe meeting, producer and director Noel Schwerin (NoelEye Documentaries) presented a short excerpt. The program follows the lives of several individuals and families as they confront genetic testing for such conditions as heart disease, Alzheimer's disease, breast cancer, and cystic fibrosis. The decisions and dilemmas of a range of personalities and perspectives are explored, including those of the sole survivor of four sisters who experiences tremendous guilt on learning that she does not harbor the gene mutation associated with a rare inherited form of breast cancer. [A print copy of the free educators' guide can be ordered from 800/991-1441 or through the extensive Web site (*www.pbs.org/gene*), which contains numerous additional resources for teachers. Discussion guides to accompany the video can be downloaded from the Web (*www.pbs.Org/gene/educator/41_discussion.html*).]

Mental Retardation Organization Viewpoint

Sharon Davis represented The Arc, a national organization of 140,000 members concerned with the welfare of people with mental retardation and their families. The Arc, funded by the DOE ELSI program to increase awareness of the Human Genome Project, is examining critical issues related to new genetic discoveries. More than 750 genetic disorders have been identified as causing mental retardation; two of the most common are Down's and Fragile X syndromes.

In discussing the future possibility of gene-based cures, Davis noted that most of The Arc's workshop participants support increased funding for research to cure mental retardation and that this does not devalue those already affected. Davis emphasized the need for education to (1) promote widespread discussion before policy is enacted to govern the use of future technologies and (2) allow informed personal choices regarding testing

Informatics *(from p. 13)*

WIT/WIT2 reconstructions are based on the metabolic pathway (MPW) collection, which includes over 2800 diagrams covering primary and secondary metabolism, membrane transport, signal-transduction pathways, intracellular traffic, transcription, and translation. Selkov observed that identifying universal metabolic aspects and gene families will lead to integrated understanding of metabolic evolution and to technologies for developing higher-level functional models. In the current public release of MPW (*wit.mcs.anl.gov/MPW*), the coding, based on the pathways' logical structure, is represented by objects commonly used in electronic circuit design. Such design facilitates diagram drawing and editing and enables automation of basic simulation operations.◊

and participation in research. She concluded her presentation by reminding the audience that "the potential for cure for some must not make us less accepting of those living with the condition." (See *www.ornl.gov/hgmis/resource/arc.html* for the full text of Davis's talk.)

Physician Education

Transitional periods can be uncomfortable, noted Sara Tobin (Stanford University) as she described the current context in which genetic advances are emerging: changes in the healthcare system; marketing pressures; uneven distribution of genetic resources, especially of genetic counselors; limited public understanding; and inadequate training of physicians in the "new genetics." A study published last year in the *New England Journal of Medicine* reported that most physicians who ordered a particular DNA-based test did not obtain prior informed consent and that one-third interpreted the results incorrectly to their patients.

Tobin is developing an interactive CD-ROM course to aid physicians who have had little or no training in clinical applications of molecular genetics. This course, "The New Genetics: Courseware for Physicians; Molecular Concepts, Applications, and Ramifications," will provide continuing medical education credits for practicing physicians in basic genetics, molecular techniques, clinical applications, and ELSI. Probable release date is spring 1999 (*www-leland.stanford.edu/dept/scbe/cdrom.htm*).◊

Abstracts of Research in Progress

DOE Human Genome Program research in progress since 1991, including research abstracts from the 1997 Santa Fe meeting, can be found on the HGMIS Web site (*www.ornl.gov/hgmis/research.html*). Print copies of the workshop proceedings are available from HGMIS (see address on p. 10).◊

In the News *(from p. 4)*

Palmisano Joins DOE OBER

Anna Palmisano
DOE OBER

On July 20 Anna Palmisano, a microbiologist and microbial ecologist, joined the Environmental Sciences Division of the DOE Office of Biological and Environmental Research. In her new position, Palmisano will continue her program management activities for the Natural and Accelerated Bioremediation Research program, the Microbial Genome Program, and such biological aspects of ocean sciences as the Biotechnological Investigations–Ocean Margins Program. She worked on these projects for 9 months in 1997 as a detailee from the Office of Naval Research (ONR), where she served as a program officer in environmental biology for 6 years. Before joining ONR, she conducted research on biodegradation in freshwater streams, soils, and landfills for the Environmental Science Department of Procter & Gamble company.

Palmisano received her B.S. in microbiology from the University of Maryland and M.S. and Ph.D. in biology from the University of Southern California, where she studied the physiological adaptation of microorganisms in Antarctica. She was a National Research Council postdoctoral fellow in planetary biology at the National Aeronautics and Space Administration–Ames Research Center, investigating the biogeochemistry of mat-like structures formed by microbes.◊

DNA Files on National Public Radio

The DNA Files: Unraveling the Mysteries of Genetics is a series of nine 1-hour nationally syndicated documentaries to be distributed this fall by National Public Radio (*www.best.com/~ringo/strp*). Hosted by NBC *Dateline* reporter John Hockenberry and supported in part by the DOE Human Genome Program, the series will explore both the science and the social, ethical, and legal implications of genetic developments. The voices of prominent researchers, people affected by advances in the clinical application of genetic medicine, members of the biotechnology industry, and others from related fields will provide real-life examples of the impact of genetic discoveries. In addition to public radio audiences, the series will target educators, scientists, and involved professionals.

The programs will include such topics as DNA and behavior; prenatal genetic testing; predictive genetic tests; gene therapy; law and the genetics of identity; genetics and biotechnology; genetics of human evolution; plants, animals, and transgenics; and the Human Genome Project. The series will be available after November 5 to local public radio stations, which should be contacted for broadcast schedules. *The DNA Files* Web site will be expanded by November 1 to feature program information, audio excerpts, resources, and interactive scenarios (*www.dnafiles.org*). [Contact: *strp@aol.com* or *mills015@tc.umn.edu*] ◊

Human Genome News

All current and past issues of *Human Genome News* are archived and searchable on the HGMIS Web site (*www.ornl.gov/hgmis/publicat/publications.html#hgn*). Articles and other text prepared for *HGN* after 1995 are indexed according to subject.◊

In the News

HUGO Ethics Committee Statement Addresses Sample Collection, Sharing

In its February *Statement on DNA Sampling: Control and Access*, the international Human Genome Organisation's Ethics Committee addressed several ethical issues pertinent to sample collection and sharing in genetic research. The committee, which is made up of scientists, ethicists, and lawyers from ten countries, also confirmed its commitment to the principles of its March 1996 *Statement on the Principled Conduct of Genetic Research* (*hugo.gdb.org/conduct.htm*).

Chaired by Bartha Knoppers (University of Montreal), the ethics group made the following recommendations regarding DNA sampling:

- Choices offered in the consent process should reflect the potential uses of the DNA sample and its information.
- Routine samples obtained during medical care may be used for research if there is general notification of such a policy, the patient has not objected, and the sample has been coded or anonymized.
- Research samples obtained with consent may be used for other research if the conditions in the statement above are met.
- Security mechanisms must be initiated to ensure respect for the choices made and the desired level of confidentiality.
- Special considerations should be made for immediate relatives, who should have access if there is a high risk of having or transmitting a serious disorder and if prevention or treatment is available.
- Stored samples may be destroyed at the request of the person if immediate relatives do not need access.
- Except as authorized by law, no disclosure of research participation or results should be made to institutional third parties without appropriate consent.
- International standardization of ethical requirements for control and access of DNA samples and information is essential.◊

Sickle Cell Mice May Lead to New Treatments

Genetically engineered mice that mimic all the symptoms of human sickle cell disease were developed by a team led by Edward Rubin and Chris Paszty at LBNL (*Science*, October 31, 1997). This new mouse strain, which carries human hemoglobin with no counteracting mouse genes, provides a means for effective testing of experimental treatments. Each year, sickle cell disease afflicts about 100,000 babies, primarily of African descent, who endure the painful debilitating condition caused by a mutant hemoglobin gene. Other members of the team included Catherine Brion, Mary Stevens, and Mohandas Narla.◊

Microbial Genome News

TIGR Sequencing Six More Microbes

The DOE Microbial Genome Program (MGP) aims to determine the sequence of bacteria having potential usefulness in energy, environmental, and evolutionary research (*www.er.doe.gov/production/ober/HELSRD_top.html*). With the support of MGP, The Institute for Genomic Research (TIGR) and its collaborators are sequencing the genomes of an additional six microbes.

- *Pseudomonas putida* (5.0 Mb)
- *Thiobacillus ferroxidans* (2.9 Mb)
- *Desulfovibrio vulgaris* (1.7 Mb)
- *Caulobacter crescentus* (3.8 Mb) with Lucille Shapiro (Stanford University) and Bert Ely (University of South Carolina)
- *Chlorobium tepidum* (2.1 Mb)
- *Dehalococcoides ethenogenes* ◊

Tuberculosis Microbe Sequenced

New Drugs, Vaccines May Result

In June researchers reported obtaining the DNA sequence of the complete 4.4-Mb genome of *Mycobacterium tuberculosis*, the organism that causes tuberculosis. The sequence is the first completed at The Wellcome Trust Pathogen Genome Unit at the Sanger Centre, U.K.

An estimated 2.9 million people died from this chronic infectious disease in 1997, and concern is growing over new antibiotic-resistant strains that have emerged in recent years. According to a *Nature* online special report on the global tuberculosis epidemic (*www.nature.com*), about one in every three people in the world is infected with *M. tuberculosis*, and each has an estimated 10% lifetime risk of progressing to clinical disease. Scientists hope that knowledge of the DNA sequence will provide clues to designing more effective therapeutic agents and vaccines.

The sequence, reported in the June 11 issue of *Nature* (**393**, 537–44), is accessible from the Sanger Centre Web site (*www.sanger.ac.uk/Projects/M_tuberculosis*).

Searching TB Genome

A tool is available through the South African National Bioinformatics Institute for searching and extracting genome sequence and open reading frames from the genome of *M. tuberculosis* (*ziggy.sanbi.ac.za/tbsearch.html*). Searches also can be performed against incomplete *M. leprae* data.◊

C. Elegans Sequencing Project Nears Finish

At the 1997 Santa Fe meeting, NIH-funded researcher Stephanie Chissoe [Washington University, St. Louis (WUSTL)] provided an overview of the final, closure phases of the project to sequence the 100-Mb genome of the roundworm *Caenorhabditis elegans*. Working in equal collaboration with the Sanger Centre (Hinxton, U.K.), researchers expect completion by the end of this year, marking another major achievement in the Human Genome Project.

A clone-based sequence-ready map provided the majority of sequencing substrates, including cosmids and YACs. Analysis and annotation of the finished sequence include identification of potential exons by similarity to EST data and known protein sequences and by gene-prediction programs. Before submission to GenBank, the generated data sets are read into the ACeDB database and reconciled manually with each other and with ancillary *C. elegans* map data. Thus far the teams have identified 13,747 annotated genes in 71.4 Mb of annotated sequence from the February 1998 ACeDB release. Some 30% match a *C. elegans* EST, and 55% have some similarity. About half of *C. elegans* genes lack significant database hits that are likely to provide clues to function, Chissoe noted, so WUSTL investigators are generating *C. briggsae* comparative sequencing data.◊

cDNA Cloning Workshop Identifies Critical Issues

Full-length cDNA Cloning: A Workshop on Problems and Solutions was held at the Banbury Center, Cold Spring Harbor Laboratory, on March 23–25. It was sponsored by Merck Genome Research Institute, NIH National Cancer Institute, and Research Genetics, Inc., and organized by M. Bento Soares (University of Iowa) and Piero Carninci (Tsukuba Life Science Center, Japan). A complete report of the meeting, including an extensive section on strategies for constructing libraries enriched for full-length cDNAs, is on the Web (*www.mgri.org,* click on "Special Reports").

Critical issues pertaining to synthesis and cloning of full-length cDNAs were identified and discussed throughout the meeting. Following are some topics on which attendees reached general consensus and made recommendations.

Starting RNA. In constructing full-length libraries, every effort should be made to isolate cytoplasmic mRNA from cells in culture or fresh (soft) tissues only, and the mRNA should be tested for contaminating nuclear RNA and DNA. Testing can be done in different ways; for example, PCR reactions can test for the presence of introns of a ubiquitously expressed gene. The use of total cellular (rather than cytoplasmic) RNA is not desirable due to the expected contamination with nuclear transcripts that will be particularly significant in the upper molecular-weight range.

Full-Length cDNA Library. Although a really full length cDNA should encompass all sequences from the 5′ cap to poly (A) addition sites, a cDNA comprising the entire protein-coding sequence should be considered worthy of full-length sequencing at high accuracy. However, every effort should be made to obtain truly full length cDNAs so sequence information can be obtained from both 5′ and 3′ noncoding regions as well.

Quality Assessment of Full-Length Libraries. Applying a set of common criteria to every new full-length library generated will become increasingly important. As part of the characterization of every new library, a common set of probes representing mRNAs of 2 kb, 4 kb, 6 kb, and 8 kb should be used to hybridize Southern blots of library DNA, endonuclease restricted to release insert from cloning vector. Jim Hudson (Research Genetics) volunteered to identify a putative list of probes corresponding to mRNAs of different sizes and abundance levels that eventually could be made available through Research Genetics. Libraries that are enriched for full-length cDNAs should be accessible for sequencing even if Southern

(See Cloning, p. 18)

Instrumentation (from p. 3)

To complete the human genome sequence within the available budget and time, substantial improvements in existing sequencing methods would be advantageous. Further, OBER places a strong emphasis on research directed to completely new approaches to genomic analysis. After 2005 the ongoing need will be for fast and cost-effective determination of DNA sequence to compare sequences among individual humans and also to determine the genomes of numerous organisms of biomedical and commercial interest. Additionally, with the continuing acquisition of this remarkable base of biological data, high-throughput experimental tools will be required to assist in a practical and useful understanding of gene function.

Specific Instrumentation Goals

- Approaches to determining DNA sequence more rapidly, accurately, and economically, particularly to increase current maximum read lengths at least 2.5 times to 2000 to 2500 bases.
- Instrumentation that integrates and more thoroughly automates DNA sequence determination (e.g., sample preparation, loading, and analysis) and data analysis. Priority will be placed on approaches that emphasize miniaturization and microfabrication.
- Approaches that (1) verify a previously determined DNA sequence's accuracy without having to redetermine its entire sequence and (2) provide economical error-checking and proofreading of newly determined DNA sequence.
- Tools that enable efficient comparison of a known DNA sequence with a related but previously undetermined DNA sequence.
- Techniques for determining the functions of large numbers of genes in parallel, particularly those that match the speed and volume of DNA sequence determination.

This solicitation was intended to stimulate research that tests the applicability of concepts unrelated to the standard instrumentation for gene sequencing. The emphasis is on basic science that will enable genomic studies in the next century, when genomic data will be widely available and the appetite for new data will be undiminished. Robust tools for using this information within new quantitative, mechanistic, and predictive biology will be paramount. Work supported within the redirected program will hasten the arrival of an epoch when ideas and experiments requiring genome-scale data are within the scope of investigator-initiated, hypothesis-driven science. *[Charles G. Edmonds (301/903-0042, Fax: -0567,* charles.edmonds@oer.doe.gov*)]* ◊

The solicitation was the topic of an editorial in the May 1 issue of *Analytical Chemistry* [**70**(9), 292A], in which Royce Murray (University of North Carolina, Chapel Hill) urged readers to think creatively about the role analytical chemists might play in fulfilling OBER goals.

For a complete description of DOE's genome instrumentation goals, see the Web site [Program Notice 98-16 and companion Notice LAB98-16 (may be listed under Closed Solicitation Notices): *www.er.doe.gov/production/grants/grants.html*].

Resources

¶ Book Focuses on Biomarker Implications

Biomarkers: Medical and Workplace Applications (Joseph Henry Press, National Academy of Sciences, May 1998) is the outgrowth of an international meeting called "Biomarkers, the Genome and the Individual: Workplace and Medical Implications of a Rapidly Evolving Technology." Held in May 1997 in Charleston, South Carolina, the DOE-sponsored meeting was organized by the Medical University of South Carolina (MUSC).

The book, also supported by a DOE grant, offers a comprehensive review of the biomarker field through a sampling of 33 talks from the Charleston meeting. It focuses on the use of biomarkers to estimate prior exposures, identify genomic changes, and evaluate underlying susceptibilities in humans. The book was edited by Mort Mendelsohn (Lawrence Livermore National Laboratory), John Peeters (DOE Office of Occupational Medicine), and Lawrence Mohr (MUSC). [Contact for book: National Academy Press (800/624-6242 or 202/334-3313, Fax: -2451)]

The meeting's main focus was the broad and challenging frontier of newly developing and increasingly sophisticated biomarkers. Until recently, biomarkers involved conventional applications of biochemistry and molecular biology to medical and toxicological situations. The new paradigm beginning to attract attention is the enormous effect of biomarker information expected to emerge from the Human Genome Project as less expensive, more accurate methods to assess DNA changes become available.

> Biomarkers are specific features or traits of an organism that can be used to detect events in biological systems. Examples of biomarkers include changes in DNA; cell death; and discovery of an environmental substance or its metabolites in urine, blood, or expelled air.

Opened with a keynote address by geneticist and Nobel laureate Joshua Lederberg (Rockefeller University), the meeting addressed the following broad subjects: (1) biomarkers for assessing prior exposures; (2) sentinel applications; (3) ethical issues and benefits of worker testing; and (4) genomics, including new technologies and biomarkers of genetic susceptibility.

Leading the speakers on genomics and new technologies, Leroy Hood (University of Washington) stated that investigators were facing a sea change in biomarker use and that the deluge of knowledge and new technologies had started already. He predicted that scientists soon will be capable of using biomarkers on a wide scale, identifying for example pharmaceutical or occupational variabilities and creating an entirely new approach to preventive medicine. David Cox (Stanford University School of Medicine) discussed the magnitude of human polymorphisms and gave an optimistic prediction that the number of markers needed to characterize people will prove surprisingly finite. Technology development was emphasized by Robert Lipshutz (Affymetrix), who spoke about the high-density DNA chip; and J. Michael Ramsey (Oak Ridge National Laboratory), who summarized the "lab-on-a-chip" approach.

Most presentations focused on current developments in biomarker research or testing, especially new biomarkers and those with recently improved validation. A new biomarker with great promise is the Clara cell protein biomarker (CC16) for pulmonary toxicity, presented by Alfred Bernard (National Fund for Scientific Research) and his colleagues from Belgium. Two presentations on electron spin resonance (ESR) technology for measuring prior exposure to radiation showed remarkable improvements and the encouraging strong correlation between tooth enamel ESR and stable chromosome aberrations in individual Japanese atomic bomb survivors.

Extensive discussions centered on the ethical and social issues of using biomarker technology to test for genetic susceptibility. A highlight was a second presentation by Cox in which he summarized the ethical aspects of genomics as accentuated by the Human Genome Project. Another highlight was a discussion of the courts' perspective on these issues as presented by two sitting judges (one federal and one state). The discussion was moderated by Franklin Zweig (Einstein Institute for Science, Health, and the Courts). No attempt was made to reach closure on the ethical issues; rather, the goal was to sensitize the audience to their importance. This field is moving so rapidly that ethical and technological issues must be reconciled soon.

Despite the huge potential of the technology, attendees felt that several highly tangible demonstrations of biomarkers' predictive power and potential value will be needed to motivate society to create a truly constructive approach for applying these methods. A current example may be beryllium toxicity, for which a hypersensitivity genetic marker has been identified. Rather than test directly in the workplace in such situations, a reasonable alternative is to offer workers a chance to be tested by a third party so they can make informed decisions about medical risk without endangering employment.

More than 400 people attended the Charleston biomarkers meeting, which brought together geneticists and other scientists, physicians, ethicists,

Access to Biomarkers Meeting Information (including abstracts)

- *www.ehap.musc.edu/agenda.html* (for more information, contact any author or speaker listed)

Cloning *(from p. 17)*

hybridization indicates suboptimal complexity. Those constructed according to some cap-selection procedures might not be very complex but could be extremely useful if significantly enriched for full-length cDNAs.

Sequencing of Random Primed Libraries to Generate Full-Length Sequence Information. There was very little overall enthusiasm for this idea because the goal is to generate full-length sequence and produce full-length clones that should be available without restrictions to academic and industrial communities.

Cloning Vector. Despite the advantages of certain lambda vectors to preferentially clone longer cDNAs, plasmids are considered advantageous given the ease of subsequent manipulation, sequence generation, and high cloning efficiencies that can be achieved via electroporation. En masse excision protocols from lambda libraries generally are not desirable because clone representation and frequencies may be altered significantly, and most participants seemed to favor cloning into plasmid vectors. Waclaw Szybalski (University of Wisconsin) argued that the use of single-copy pBAC-like vectors should be considered as far as cloned cDNA stability is concerned. The conditionally amplifiable pBAC is preferred.

The next cDNA meeting is planned for March 1999 in Japan. [*Bento Soares, bento-soares@uiowa.edu*] ◊

industrialists, government workers, attorneys, and members of the public to discuss the science and societal implications associated with the use of this emerging technology. Meeting organizers made a significant effort to involve the genomics community and to enhance connections with the occupational medicine and worker communities.

A report summarizing the state of biomarker technology and its possible applications for DOE workers is expected later in 1998. Another book, *Biomarkers and Occupational Health: Progress and Perspectives* (1995), edited by Mendelsohn, Peeters, and Mary Janet Normandy (DOE Office of Occupational Medicine), resulted from a previous DOE-sponsored biomarker meeting held in 1994 in Santa Fe, New Mexico. [See *HGN* **6**(3), 8–9 for 1994 meeting report.] *[Mort Mendelsohn* (mendelsohn2@llnl.gov), *John P. Peeters* (john.peeters@eh.doe.gov), *and Betty Mansfield* (bkq@ornl.gov)] ◊

¶ Report on Functional Consequences of Gene Expression

The final report of the workshop, "Functional Consequences of Gene Expression in Health and Disease," held March 31 to April 3, 1997, in San Antonio, Texas, has been published by DOE (see box on Report Access, below right).

At the workshop were experts representing genetics, biochemistry, molecular and cellular biology, physiology, oncology, radiology, and nuclear medicine. They discussed with DOE representatives the expectations and possibilities for helping clinical investigators and physicians use the vast new knowledge coming from the Human Genome Project. A workshop goal was to identify (1) functional units in terms of biochemical circuits within such complex adaptive systems as the human body that can be observed in vivo and described as a consequence of interacting substrates in response to specific gene expression; (2) useful, practical, and economical tools for in vivo observations of metabolic and functional circuits in response to gene expression in individuals; and (3) promising applications of these concepts and tools for medical research and practice. Specific models, radiopharmaceuticals, measurement techniques, instrumentation, and methods for linking recognized phenotypic molecular expressions to individual genotypes are crucial to the task.

Broad discussions clarified individual perceptions of concepts, tools, and applications. These sessions were followed by presentations of experimental data on various aspects of signal transduction and pathways in cellular metabolism. Also covered were the technology of studying the relationship between genes and particular metabolic reactions and phenomena in specific organs, and respective modeling and data interpretation. In this context, genotype and phenotype linkage as active in both directions was reemphasized. Immediate research opportunities exist for studying various specific metabolic reaction circuits in recognized linkage to gene expression in neurology, cardiology, oncology, and gerontology.

The workshop concluded with a plea to exploit new opportunities created by the genome program's success, integrate diverse efforts, and optimize resources.◊

Workshop Report Access

- **Web:** *www.er.doe.gov/production/ober/mab/MABRD_top.html*
- **Phone:** 301/903-3123
- **E-mail:** *roland.hirsch@oer.doe.gov*
- **Mail (fee charged):** Office of Scientific and Technical Information; P.O. Box 62; Oak Ridge, TN 34831-8062

☛ Mouse Genome Informatics (MGI) Release 2.0

The MGI Web site (*www.informatics.jax.org*) provides integrated access to various information resources on the genetics and biology of the laboratory mouse. The site includes the Mouse Genome Database, Gene Expression Database (GXD), and Encyclopedia of the Mouse Genome.

- **New Gene-Expression Data and Query Form**: Data set of RT-PCR assays produced by Tom Freeman's laboratory represents analyses of 517 genes in 46 mouse tissues from animals 6 to 8 weeks old and studies of tissues from 15-day-old mouse embryos. Additional gene-expression data sets will be publicly available soon.
- **Gene-Expression Data Query Form**: Users can search for expression data related to specified genes, anatomical structures, or developmental stages. Query results can be returned as listing of assays or assay results.
- **Reorganized and Simplified Linkage Maps Form**: Users can generate a linkage map display of markers retrieved from the database. They can view a mouse chromosome or chromosomal region and may choose to include homologies from another selected species, syntenic markers, and DNA segments or add new marker information.
- **AXB-BXA Mapping Data Sets in EXCEL Format via Ftp**: Produced by Beverly Paigen (Jackson Laboratory), new and revised AXB and BXA typings represented in RI mapping experimental data records will be incorporated into MGD's composite RI data sets.
- **Revised Data-Submission Guidelines**: Facilitate e-mail MGD data submission.
- **1998 Chromosome Committee Reports Online**

[MGI User Support: 207/288-6445, Fax: -6132, *mgi-help@informatics.jax.org*] ◊

¶ Genome Analysis Protocol Handbook

The *ICRF Handbook of Genome Analysis*, edited by Nigel Spurr (SmithKline Beecham Pharmaceuticals) and others, is a combination protocol manual and information resource drawn from expert contributors at a number of research centers. It describes and evaluates a wide range of techniques, providing step-by-step protocols for genetic and physical mapping and DNA sequencing. The two volumes cover human and several model organism genomes. 1998. [Contact: Blackwell Science (800/759-6102 or 781/388-8250, Fax: -8255)] ◊

☛ HGMIS Seeks "Teachers of Teachers"

HGMIS is compiling a list of names and contact information for teachers to serve as mentors, advisors, and sources of material for others who are developing teacher-education programs in genetics, genomics, and Human Genome Project issues. The list will be included in resource material furnished to biology teachers and may be posted on the HGMIS Web site.

If you are willing to list your name or if you know of someone who might be interested, please contact HGMIS at the address on p. 10.◊

Resources

¶ After the Genome Project: Understanding the Data

Survey Identifies Growing Need for Synchrotron Analyses

Structural Biology and Synchrotron Radiation: Evaluation of Resources and Needs (1997) is a report on the current status of biological uses and demands of synchrotron radiation in the United States (see box at right). For this report, staff at the synchrotron radiation facilities and their user communities were surveyed, and a group of experienced structural biologists analyzed the data.

In evaluating what synchrotron facilities and support operations are needed and in anticipating future requirements for sustaining the exciting progress in structural biology, the BioSync Committee noted the expanded impact of structural biology in recent years. This expansion has led to an increase in the size and complexity of macromolecular structures being determined and in the difficulty of experiments being pursued. Structural biology is having a widening effect on such diverse fields as immunology, neurobiology, cell biology, virology, physiology, molecular biology, medicine, and biotechnology.

Recent advances in structural biology can be attributed to (1) methodological improvements that allow a vast array of cellular proteins to be cloned and expressed in quantities sufficient for structural studies, (2) use of cryocrystallography to prepare extremely stable crystals, and (3) availability of and technological innovations at synchrotron radiation facilities (see box, Envisioning the Proteome, at right). These factors have brought many more projects of high biological significance into the realm of structural biology. Without synchrotron sources, many of these new research projects could not have been undertaken.

The BioSync Committee reached the following main conclusions:

- Structural biology research is producing results of high biological impact that have a direct bearing on human health issues.
- Synchrotron radiation, combined with multiwavelength anomalous diffraction phasing, has revolutionized the discovery of new macromolecular structures;
- Noncrystallographic applications to structural biology continue to expand.
- Demand for structural information and synchrotron time is growing very rapidly in all molecular fields of biology.
- Regional facilities will increase in importance.
- The most cost-effective way to improve throughput at synchrotron facilities is to upgrade existing beamlines.
- More cooperation is highly recommended among organizations funding synchrotron facilities and basic research.
- A BioSync report published in 1991 concluded that structural biology, especially crystallography, was a very rapidly expanding field with a growing impact on basic and applied biology and that synchrotron radiation facilities available at the time were insufficient for the community's needs. Construction of additional beamlines and improved support for existing beamlines were recommended. Much of this increase has been realized with new facilities at Argonne and Berkeley and additional beamlines for biological use at Brookhaven, Stanford, and Cornell.◊

1997 BioSync Report

- Electronic version: *www.ornl.gov/hgmis/biosync*
- Print copies: J. Hollister; Department of Biological Sciences; Purdue University; West Lafayette, IN 47907

Related Information

- *www.er.doe.gov/production/ober/HELSRD_top.html* (click on "Structural Biology")

Envisioning the "Proteome"

Translating the increasing stores of genome data into practical knowledge about biological function—a rapidly growing field known as functional genomics—will be one of the biggest challenges facing modern biology. One promising method is to look for clues by visualizing the 3-D structure of the proteins (the human "proteome") encoded by the human genome's estimated 80,000 genes. Because biological structures have been shaped by evolution to serve their functions, they could reveal important patterns that suggest common functional mechanisms.

To enable explorations into structural biology, the DOE Office of Biological and Environmental Research supports research at synchrotron radiation sources that focus X-ray beams on tiny protein crystals and produce a diffraction pattern to reveal the protein's intricate structure. Users of these DOE facilities, which often cost hundreds of millions of dollars, include scientists from universities, medical schools, government laboratories, and pharmaceutical companies. Although synchrotron radiation sources were once the sole province of physicists, biologists now account for about a third of all users.◊

¶ Book on Tuskegee Conference

Plain Talk About the Human Genome Project is a 292-page paperback compilation of talks presented during a 3-day DOE-sponsored conference at Tuskegee University in September 1996. The talks were updated by presenters in November 1997. Distinguished leaders, scientists, ethicists, and educators spoke on wide-ranging topics related to the Human Genome Project's promise and perils, matters of race and diversity, and education about the project and its implications. For lay and academic readers, the book appends the DOE *Primer on Molecular Genetics* and lists useful Web sites. [Contact: Tuskegee University Publications Office (334/727-8035, Fax: /724-4451; for a list of contributors and other information, access *agriculture.tusk.edu/caens/genome/genome.html*] ◊

☛ Expressed Human Genome Sequences Database

The National Center for Genome Resources and the South African National Bioinformatics Institute are collaborating on the Sequence Tag Alignment and Consensus Knowledgebase, or STACK. This public database of expressed sequences provides a unified view of human genes.

STACK features expressed gene sequences organized according to tissue and provides a comprehensive representation of each gene with alignments of its expressed fragments. Algorithms used to generate the database include efficient error-compensation methods that can create longer, more accurate consensus sequences. [STACK access: *www.ncgr.org/gsdb* and *www.sanbi.ac.za/stack*] ◊

Web Site Provides HGP Access to Scientific and Public Audiences

The Human Genome Management Information System (HGMIS) was established in 1989 by the DOE Human Genome Program Task Group to inform scientists, policymakers, and the public about the program's research. To make Human Genome Project (HGP) data, technologies, and implications more accessible, HGMIS produces *Human Genome News* (*HGN*) and a number of other information resources. [More information on HGMIS can be found at *www.ornl.gov/hgmis/about/hlights.html*]

Sites for Different Audiences

HGMIS started its Web site in 1994 to house the electronic version of *HGN*. As interest in the genome project and Web use expanded, HGMIS added other publications, information, and links to make the site a comprehensive, text-based Web server on project-related topics. Gradually, in a shift that reflects Web trends in general, public users became predominant.

To address the growing interests of a population beginning to realize the societal impact of genetic research while continuing to meet the needs of the broad scientific community, HGMIS divided its Web site into two sections in January. One section is targeted to the general public and the other to researchers, but all users still have easy access to both sections.

The decision to split the Web site was based on an evaluation of user needs. HGMIS considered a user survey; the types of domains accessing the site; and user questions, a large number of which were from students seeking basic facts. Content and ease of access for the public and researchers were crucial considerations in the site's renovation. In addition to providing informative text, this comprehensive site about HGP also serves as a launch pad to myriad related sites.

The HGMIS site, which contains 1800 text files, has received numerous awards and has been listed by several exclusive Web rating systems. Some 2400 outside sites link to the site, whose files are accessed by about 17,000 host computers each month for about 2 million file transfers yearly. HGMIS has adapted its design to accommodate as many users as possible.

Public Section

www.ornl.gov/hgmis

The public section, *Human Genome Project Information*, emphasizes materials to facilitate public knowledge about genetics. The site includes project history; recent developments and discoveries; ethical, legal, and social issues; information for people with genetic conditions; educational materials; and resources for teaching high school and college classes. HGMIS reviewed questions and comments over the previous year for commonalities to develop two new files: Frequently Asked Questions (FAQs) and QuickFact Finder. These pages, which answer many routine questions once fielded individually by HGMIS staff, are proving very popular with students and newcomers to the project.

To contribute to its user-friendly stance, the public-oriented section takes a less formal tone and avoids jargon where possible. Much of the material has been reorganized to target such specific groups as teachers and students.

Research Section

www.ornl.gov/hgmis/research.html

The backbone of the technical section, *Human Genome Project Information: Research in Progress*, comprises the pages that categorize files by subject areas: Mapping; Sequencing; Informatics; Instrumentation; and Ethical, Legal, and Social Issues. These pages, developed in a previous remodeling, provide general information and link directly to current research abstracts, *HGN* articles, and Web resources related to the subject area. In addition, the research section contains HGP publications, conference and workshop listings, funding and training opportunities, and links to the larger genome community.

Print Document Access

Most HGMIS print documents, along with publications from other sources, are provided electronically through the Web site. These resources include *HGN*; the DOE *Primer on Molecular Genetics*; *To Know Ourselves*; and reports on the DOE Human Genome Program, Santa Fe contractor-grantee workshops, and other related topics. In addition, *Your Genes, Your Choices*, produced by the American Association for the Advancement of Science with DOE support, is accessible via the HGMIS site.

Future Enhancements

Every good Web site adapts to changes in purpose, technology, and audience needs. As interest and requirements expand, HGMIS plans to add other useful features, including a resource page for medical professionals and an expanded listing of FAQs. The *Human Genome Project Research* and *Human Genome Project Information* sites will continue to evolve as user needs and the project progress.◊

NCGR Announcements

The National Center for Genome Resources (NCGR) in Santa Fe, New Mexico (*www.ncgr.org*), made the following announcements in its July newsletter:

- NCGR welcomes it new chief scientific officer, Michael M. Harpold, who most recently served as vice-president of research at SIBIA Neurosciences, Inc. Harpold will fulfill a critical role in NCGR's efforts to enhance its bioinformatics and computational biology initiatives related to Genome Sequence DataBase (GSDB) and Agricultural Genomics.
- A 64-processor Sun HPC 10000 server, popularly known as the Starfire, has been added to serve the center's databases and computing needs. The powerful Starfire is particularly well suited to NCGR because it can be partitioned into Dynamic System Domains, which are separate, smaller computers with the ability to allocate resources between them on the fly.
- GSDB is adding the DeCypher ES-1920 server to provide users with access to Smith-Waterman and other powerful search algorithms. The new hardware will be beta tested and fully implemented before its public debut, expected in September.◊

New System Identifies Polymorphisms

Polymorphism finder (*POMPOUS. swmed.edu*): A suite of programs that detects tandem repeats ranging from dinucleotides up to 250-mers, scores them according to predicted level of polymorphism, and designs appropriate flanking primers for PCR amplification.◊

DOE Supports Web Site for 1997 AAAS Symposium Talks

With the partial support of DOE, a Web site has been created to provide electronic audio access to presentations by nine distinguished speakers at the February 1997 symposium, "The Human Genome Project: What's the Public Got To Do With It?" In addition to these presentations, which were made at the annual meeting of the American Association for the Advancement of Science in Seattle, Washington, the site includes the transcript of a lively and informative presymposium dialogue conducted via the Internet [*www.aaas.org/spp/dspp/sfrl/projects/hgdp/about.htm*].◊

Calendar of Genome and Biotechnology Meetings*

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

September 1998

10. TIGR/NRC/DOE Distinguished Speaker Series: Edward O. Wilson (Harvard Univ.); Washington, DC [D. Hawkins, 301/838-3501, Fax: -0209; dhawkins@tigr.org; *www.tigr.org*]

11. Exploring Parasite Genetics; London [D. Johnston, +44-171/938-9297, Fax: -9294; daj@nhm.ac.uk; *www.umds.ac.uk/bsp/as98info.htm*]

11–13. Partnerships and Strategies: Forging Understanding and Collaboration among Genetic Scientists, Policy Makers, and Consumers. 1998 AGSG Meeting; Washington, DC [M. Wilson, 800/336-4363]

13–17. Macromolecular Organization and Cell Function; Oxford, U.K. [GRC, 401/783-4011, Fax: -7644; grc@grcmail.grc.uri.edu; *www.grc.uri.edu*]

14–15. NHGRI Advisory Council Meeting; Bethesda, MD [K. Malone, 301/402-2205, Fax: -0837; kimberly@od.nhgri.nih.gov]

14–15. Bioinformatics: Tools for Genomic Research; Seattle [IBC, 508/481-6400, Fax: -7911; reg@ibcusa.com; *www.ibcusa.com*]

14–16. Genes, Proteins, and Computers V; York, U.K. [M. Faller, +44-1925/603-492, Fax: -100; m.g.faller@dl.ac.uk; *wserv1.dl.ac.uk/CCP/CCP11/conferences/gpc_v*]

16–17. Natl. Bioethics Advisory Commission Meeting; Washington, DC [NBAC, 301/402-4242, Fax: /480-6900; *www.bioethics.gov/bioethics/meetings.html*]

17–18. Pharmacogenetics/Pharmacogenomic Business Strategies: CRO-Pharmaceutical Partnerships; Baltimore [NMHCC, 941/373-1290, Fax: -1638; biotech@nmhcc.com; *www.biotech.nmhcc.org*]

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

20–24. 6th Small Genomes Conf.; Arrowhead, CA [ASM, 202/942-9356, Fax: -9340; meetingsinfo@asmusa.org; *www.asmusa.org/mtgsrc/genomeg.htm*]

22. 11th Annu. Colorado Biotechnol. Symp.; Fort Collins, CO [CIRB, 970/491-1791, Fax: -7369; vince@engr.colostate.edu]

23–27. Gene Therapy; Cold Spring Harbor, NY [CSHL, 516/367-8346, Fax: -8845; meetings@cshl.org; *www.cshl.org*]

24–25. 16th Annu. ATCC Biotech Patent Forum; Washington, DC [ATCC, 703/365-2700, Fax:-2701; workshops@atcc.org; *www.atcc.org/workshops/workshop.html*]

30–Oct. 3. 12th Intl. Mouse Genome Conf.; Garmisch-Partenkirchen, Germany [D. Miller, 716/845-4390, Fax: -8169; dmiller@mcbio.med.buffalo.edu; *www.gsf.de/isg/imgc98.html*]

October 1998

4–7. Structure-Based Functional Genomics; Avalon, NJ [R. Watson, 732/235-5321; *www.cabm.rutgers.edu/bioinformatics_meeting*]

6–7. Transcription Regulation of Clinically Relevant Genes; La Jolla, CA [see IBC contact: Sept. 14–15]

8–10. 5th Intl. Conf. on Automation in DNA Mapping and Sequencing; St. Louis [WUSTL School of Medicine, ams98@watson.wustl.edu; *www.genome.wustl.edu/gsc/ams98.html*]

8–10. Genomic Imprinting and Environmental Disease Susceptibility; Durham, NC [F. Tyson, 919/541-0176; tyson2@niehs.nih.gov; *www.niehs.nih.gov/envgenom/events.htm*]

14–18. Gene Regulation and Cancer; Hot Springs, VA [AACR, 215/440-9300, Fax: -9313; webmaster@aacr.org; *www.aacr.org*]

17. Individual Genetic Variation: Implications of the Coming Transformation of Medicine; Palo Alto, CA [H. Silverberg; heather.silverberg@stanford.edu; *www-leland.stanford.edu/dept/scbe/individu.htm*]

17–18. Genetics, Ethics, Medicine and Society; San Francisco [Imamia, 701/293-7833, Fax: /232-4626; imiwaiting@juno.com]

17–21. Behavioral Science and Cancer Genetics: New Roles, New Partners; Washington, DC [J. Linder-Crow, 800/374-2721, Fax: 202/336-6151; jlinder-crow@apa.org; *www.apa.org/ce/genetics.html*]

24–26. Beyond the Identification of Transcribed Sequences; Washington, DC [peggy@eri.uchsc.edu; *www.ornl.gov/TechResources/meetings/iwits8/announce.html*]

24–27. 1998 Annu. ISONG Educ. Conf.; Denver [J. Jenkins, 301/496-0921, Fax: -0047; jenkinsj@navmed.nci.nih.gov; *nursing.creighton.edu/special/isong*]

24–27. Back to the Future: Genetic Counseling in the 21st Century. 17th Annu. NSGC Educ. Conf.; Denver [B. Leopold, 610/872-7608, Fax: -1192; nsgc@aol.com; *members.aol.com/nsgcweb/ec17.htm*]

26–28. Protein Analysis and Characterization: Proteomics; Baltimore [see contact: Sept. 17–18]

27. HUGO Mutation Database Meeting; Denver [R. Horaitis, horaitis@ariel.ucs.unimelb.edu.au; *ariel.ucs.unimelb.edu.au:80/~cotton /denver.htm*]

27. 5th Intl. Chromosome 15 Workshop; Denver [R. Schultz, 214/648-1681, Fax: -1666; schultz@ryburn.swmed.edu]

27–31. ASHG; Denver [M. Ryan, 301/571-1825, Fax: /530-7079; *www.faseb.org/genetics/ashg/ann-meet/ashgmeet.htm*]

31–Nov. 3. Computational Genomics Conf. II; Washington, DC [TIGR, 301/838-3515, Fax: -0229; bglauro@tigr.org; *www.tigr.org*]

31–Nov. 7. 4th Tsukuba Intl. Bioethics Roundtable and World Congress on Bioethics; Tsukuba Science City, Japan [D. Macer, +81-298/53-4662, Fax: -6614; macer@sakura.cc.tsukuba.ac.jp; *www.biol.tsukuba.ac.jp/~macer/TRT4.html*]

November 1998

4–7. NABT Natl. Convention; Reno, NV [NABT, 703/471-1134, Fax: /435-5582; nabtconv@aol.com; *www.nabt.org/convention.html*]

9–12. 1998 Hanson Symp.: From Genes to Therapeutics; Adelaide, Australia [Hanson Centre for Cancer Research, plevin@camtech.net.au; *192.133.41.49:80/symposia/symposium.html*]

11–15. Endogenous Sources of Mutations; Ft. Myers, FL [see contact: Oct. 14–18]

17–18. Natl. Bioethics Advisory Commission Meeting; Miami [see contact: Sept. 16–17]

19. TIGR/NRC/DOE Distinguished Speaker Series: Norman Pace (UCB); Washington, DC [see contact: Sept. 10]

29–Dec. 1. GeneCom '98: Gene Technology and the Community; Adelaide, Australia [Secretariat, +618/8364-6688, Fax: /8333-0944, genecom@gryphen.com.au]

December 1998

10–11. 9th Workshop on Genome Informatics; Tokyo [S. Miyano, +813/5449-5615, Fax: -5442; giw@ims.u-tokyo.ac.jp; *giw.ims.u-tokyo.ac.jp/giw*]

January 1999

TBA. 7th DOE HGP Contractor-Grantee Meeting; California [K. Poe, 510/486-4879, Fax: -5717; kjpoe@lbl.gov; *www.lbl.gov/Conferences/DOE_HGP*]

4–9. Pacific Symp. on Biocomputing; Mauni Lani, Big Island, Hawaii [L. Hunter, 301/496-9303, Fax: -0673; hunter@nlm.nih.gov; *www.cgl.ucsf.edu/psb*]

17–21. Plant and Animal Genome VII; San Diego [D. Scherago, 212/643-1750, Fax: -1758; pag@scherago.com; *www.intl-pag.org/pag/pag7.html*]

20. TIGR/NRC/DOE Distinguished Speaker Series: Michael Brown (UTSWMC); Washington, DC [see contact: Sept. 10]

23–24. Genome Seminar II as part of the 1999 AAAS Annu. Meeting; Anaheim, CA [AAAS Meeting Office, 202/326-6450, Fax: /289-4021; confinfo@aaas.org; *www.aaas.org/meetings/meetings.htm*]

29–Feb. 1. Conf. on Microbial Genomes III: Sequencing, Functional Analysis and Comparative Genomics; Chantilly, VA [see contact: Oct. 31–Nov. 3

February 1999

TBA. 2nd HUGO Pacific Genome Meeting; Bali, Indonesia [Secretariat, +62-21/391-71313, Fax: /314-7982; yoko-ono@ims.u-tokyo.ac.jp]

22–23. NHGRI Advisory Council Meeting; Bethesda, MD [see contact: Sept. 14–15]

27–28. 2nd Annu. Genomic Opportunities: Emerging and Early State Partners; San Francisco [CHI, 617/630-1300, Fax: -1325; chi@healthtech.com; *www.healthtech.com/conferences*]

March 1999

1–3. 6th Annu. Human Genome Project: Commercial Implications; San Francisco [see contact: Feb. 27–28]

4–5. 3rd Annu. Gene Functional Analysis; San Francisco [see contact: Feb. 27–28]

27–30. HGM '99; Brisbane, Australia [A. Day, +617/3365-4447, Fax: -4430; a.day@cmcb.uq.edu.au] ◊

*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.

DOE

HGP Ethical, Legal, and Social Implications (ELSI)

Notice 98-19

www.er.doe.gov/production/grants/fr98_19.html

Topic: Research that addresses, analyzes, or anticipates ELSI issues associated with the use of information and knowledge emanating from the Human Genome Project. DOE particularly encourages research into the privacy of genetic information in the workplace; access to and protection of stored genetic information; development and dissemination of educational materials to enhance understanding of ELSI issues related to genomics; and ELSI implications of advances in the scientific understanding of complex or multigenic conditions, gene-environment interactions, and human polymorphisms.

- **Application due date**: September 17, 1998 (Potential applicants are strongly urged to contact DOE program staff to discuss their projects before submitting applications.)
- **Contact**: Daniel W. Drell (301/903-6488, *daniel.drell@oer.doe.gov*) ◊

Training Events*

October 1998

4–15. Genetic Approaches in Complex Heart, Lung, and Blood Diseases; Bar Harbor, ME [Jackson Laboratory, 207/288-6262; education@jax.org; *www.jax.org*]

14–27. Positional Cloning: Contig to Candidate Gene; Cold Spring Harbor, NY [CSHL, 516/367-8346, Fax: -8845; meetings@cshl.org; *www.cshl.org*]

November 1998

7–12. Computational Genomics; Cold Spring Harbor, NY [see contact: Oct. 14–27]

9–13. Advanced Linkage Course; New York [K. Montague, 212/327-7979, Fax: -7996; montagk@rockefeller.edu, *linkage.rockefeller.edu*]

16–27. Practical Course: Gene Expression in Mammalian Systems; Trieste, Italy [G. Chatterjee, +91-11/616-7356, Fax: /616-2316; chatterk@icgebnd.ernet.in; *www.icgeb.trieste.it*]

19–21. Ethics and Genetics: Advanced European Bioethics Course; Nijmegen, Netherlands [B. Gordijn, +31-24/361-5320; b.gordijn@efg.kun.nl; *www.azn.nl/fmw/onderwys/ukgene.htm*]

December 1998

2–12. Methods in Genome Sequencing and Analysis; Heidelberg [W. Ansorge, + 49-6221/387-355, Fax: -306; ansorge@embl-heidelberg.de; *www.embl-heidelberg.de/~saffrich/DNASEQ98/DNASEQ98.html*] ◊

NIH NHGRI

The following three awards are designed to foster the career development of persons with expertise in scientific disciplines that would further technological developments critical to the Human Genome Project and to understanding the genetic basis of diseases. Individuals must have degrees in computer science, mathematics, chemistry, engineering, physics, or such closely related scientific disciplines as bioinformatics, computational biology, statistics, biomathematics, and bioengineering.

- Individual Mentored Research Scientist Development Award in Genomic Research and Analysis ***PAR-98-96*** (*www.nih.gov/grants/guide/pa-files/PAR-98-061.html*)
- Institutional Mentored Research Scientist Development Award in Genomic Research and Analysis ***PAR-98-062*** (*www.nih.gov/grants/guide/pa-files/PAR-98-062.html*)
- Curriculum Development Award in Genomic Research and Analysis ***PAR-98-063*** (*www.nih.gov/grants/guide/pa-files/PAR-98-063.html*)

Application receipt dates for all three topics: February 1, June 1, and October 1

Contact for all three topics: Bettie Graham (301/496-7531, Fax: /480-2770, *bettie_graham@nih.gov*) ◊

NHGRI Initiates Mailing List

NIH NHGRI has initiated a mailing list for disseminating funding information to a wider audience. The list will be used infrequently to (1) send a brief summary of Program Announcements and Requests for Applications, along with a Web page pointer to the full text and (2) notify the community about new policies specifically for NHGRI grantees. Announcements about the ELSI Research Program are sent from a separate list (*ELSI@nhgri.nih.gov*). To subscribe to the mailing list, send an e-mail message to *listserv@list.nih.gov* and leave the subject line blank. Into the message, insert SUBSCRIBE NHGRIBULLETIN-L YOURNAME (type your name in place of YOURNAME). The listserv will obtain your e-mail address from your message. Do not include anything else in the body of the message.◊

Meetings with ASHG

Several October conferences and workshops will be held concurrently with the annual meeting of the American Society of Human Genetics in Denver, Colorado. See the calendar on p. 22.◊

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: *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, www.orau.gov/ober/hollaend.htm*)

Computational Molecular Biology Postdoctoral Fellowships

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

- Next deadline: January 18, 1999
- Contact: Christine Trance; Alfred P. Sloan Foundation; 630 Fifth Ave., Ste.2550; New York, NY 10111 (212/649-1649, Fax: /757-5117, *trance@sloan.org*)

NIH National Human Genome Research Institute

- NHGRI program: 301/496-7531, Fax: /480-2770, *www.nhgri.nih.gov/About_NHGRI*
- Program announcements: *www.nhgri.nih.gov/Grant_info*
- ELSI: 301/402-4997

Small Business Innovation Research Grants

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

Contacts:

- DOE SBIR/STTR Office: 301/903-1414 or -0569, Fax: -5488, *sbir-sttr@oer.doe.gov*. DOE solicitations released in early October for STTR and early December for SBIR. SBIR: *sbir.er.doe.gov/sbir*; STTR: *sttr.er.doe.gov/sttr*
- Bettie Graham (see contact, NHGRI). NIH SBIR due April 15, August 15, and December 15. STTR, April 1, August 1, and December 1

National SBIR/STTR conference: Nov. 3–5, 1998, Boston, MA (360/683-5742, *www.zyn.com/sbir*). For regional conferences, see Web site. ◊

Human Genome Management Information System Subscription/Document Request* (Vol. 9, No. 3)

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. ___ *Human Genome Program Report: Part 1, Overview and Progress; Part 2, 1996 Research Abstracts* (*www.ornl.gov/hgmis/publicat/97pr*)
3. ___ *To Know Ourselves* (DOE) ___DOE *Primer on Molecular Genetics* (also see *www.ornl.gov/hgmis/publicat/publications.html*)
4. ___ Reprint of "What Can the New Gene Tests Tell Us?" by Denise Casey (*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

AAAS Am. Assoc. for Advancement of Science
AACR Am. Assoc. for Cancer Res.
AGSG Alliance of Genet. Support Groups
ASHG Am. Soc. for Hum. Genet.
ASM Am. Soc. for Microbiol.
ATCC Am. Type Culture Collection
BAC bacterial artificial chromosome
bp base pair
CHI Cambridge Healthtech Inst.
CIRB Colorado Inst. for Res. in Biotechnol.
cM centimorgan
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
FISH fluorescence in situ hybridization
GRC Gordon Res. Conf.
HGM Hum. Genome Mtg.
HGMIS Hum. Genome Mgmt. Information Syst.
HGP Hum. Genome Proj.
HUGO Hum. Genome Org.
IBC Intl. Bus. Commun.
ISONG Intl. Soc. of Nurses in Genet.
LANL Los Alamos Natl. Lab.
LBNL Lawrence Berkeley Natl. Lab.
LLNL Lawrence Livermore Natl. Lab.
NABT National Assoc. of Biology Teachers
NBAC Natl. Bioethics Advisory Commission
NHGRI Natl. Hum. Genome Res. Inst.
NIH Natl. Inst. of Health
NMHCC Natl. Managed Health Care Cong.
NRC Natl. Res. Council
NSGC Natl. Soc. of Genet. Counselors
PAC P1 artificial chromosome
STS sequence tagged site
TIGR The Inst. for Genomic Res.
UCB Univ. of Calif., Berkeley
UTSWMC Univ. of Texas SW Med. Cntr.
WUSTL Washington Univ., St. Louis
WWW World Wide Web
YAC yeast artificial chromosome

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

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

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

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

Recycled Recyclable