Human Genome news

Sponsored by the National Institutes of Health and the U.S. Department of Energy

ISSN: 1050-6101

Vol. 6, No. 3, September 1994

Managing Genome Sequence Data

Collaborative Databases Serve Worldwide Research Community

Historical Perspective, Current Status, Future Challenges Examined

As technology improves and information accumulates exponentially, continued progress in the Human Genome Project will depend increasingly on the development of sophisticated computational tools and resources to manage and interpret data. Various systems now manage data relevant to genome research; these systems range from highly specialized databases supporting local research projects to general databases serving the entire international community both as repositories and analysis resources that guide ongoing research.

Public databases containing the nucleotide sequences of the complete human genome and of selected model organism genomes will be a major product of the Human Genome Project. The ease with which researchers can retrieve and use the data from these and other related databases will provide one measure of the project's success.

Although much progress has been made in database development and operation, many challenges remain in collecting, organizing, storing, and distributing data. As maps and sequences accumulate and the focus shifts from data generation to analysis, new challenges will arise. Some feel that a key task will be to link the various biological databases into a loosely coupled distributed alliance so researchers around the world can explore all relevant facets of a particular topic. Research and development for these interoperable databases demand the close interaction of biologists with mathematicians, software engineers, and programmers to develop the needed software, database tools, operational infrastructure, and algorithms.

Four major nucleotide sequence databases now store almost 200 million bp representing human and more than 8000 other species. The four are GenBank® and Genome Sequence Data Base (GSDB) in the United States, European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database, and the DNA Data Bank of Japan (DDBJ). Each group collects a portion of the total sequence data reported worldwide, often processing submissions and update requests within 48 hours. Because they exchange new and updated sequences frequently—usually daily—the databases contain the same sequences, each in its own format. All four of these evolving databases are working to improve data design and quality.

Database growth is accelerating rapidly, with more than half the sequences having been added in the last 2 years. This number is expected to rise dramatically within the next decade to about 10 billion bp. As more genes are identified and sequenced and understanding of sequence data improves, databases will play an increasing role in capturing new knowledge and making it accessible.

Sequence Database History

The Los Alamos Sequence Library was established in 1979 at the DOE Los Alamos National Laboratory (LANL) to store DNA sequence data in electronic form. At about the same time, database activities were beginning at EMBL, and

In This Issue. . .

Page	Genome News
1	Managing Genome Sequence Data
2	Breast Cancer Gene Found
5	Whitehead, Généthon Groups Present New Linkage Map
7	OHER Launches Microbial Genome Initiative
7	Evans Moves NIH Genome Center to Dallas
8	DOE Biomarker Workshop Meets in Santa Fe
10	DOE ELSI Program Enters Fifth Year
11	*GDB Forum:* Medline Citations; Graphical Interfaces; Sprintnet Access; OMIM on WWW; User Support, Registration, Training
12	Chromosome 16 Mappers Review Status
13	Chromosome Editors (Nomenclature Committee)
	For Your Information
3	*Resources (see also p. 7)*
6	*Publications (see also p. 13)*
14	Calendars of Genome-Related Events, Training
15	Funding Announcements, Guidelines for U.S. Genome Research
16	Subscription/Document Request; Selected Acronyms

Genome News

discussions began in 1982 on collaborations between the two institutions. From these early days, EMBL and LANL agreed that any data submitted to or entered by one group would be forwarded immediately to the other, thus avoiding duplication of effort. As other sequence databases joined the collaboration, data sharing was extended to them.

Breast Cancer Gene Found

A team of some 45 researchers led by Mark Skolnick (University of Utah Medical Center and Myriad Genetics, Inc.) and Roger Wiseman (NIH National Institute of Environmental Health Sciences, North Carolina) reported on September 14 the isolation of the BRCA1 gene. Defective forms of BRCA1 are thought to cause predisposition to certain inherited forms of breast and ovarian cancer. Scientists have been searching for BRCA1 in a 600-kb region since 1990, when Mary-Claire King's group (University of California, Berkeley) demonstrated a pattern of genetic markers in families where breast cancer occurred unusually early and frequently.

Two papers describing the discovery appear in the October 7 issue of *Science*, and a separate study on the approximate location of another gene associated with breast cancer—BRCA2—was published in the September 30 issue. The 100-kb BRCA1 gene is composed of 21 coding exons midway between markers D17S1321 and D17S1325 on the chromosome 17 long arm. Isolation of this gene could lead to significant clues about the risk of developing cancer, not only of the breast and ovaries, but also of the colon and prostate gland.

Although these discoveries are considered very important for studying the disease and eventually developing new diagnostic tools and treatments, scientists warned that much work lies ahead. NIH Director Harold Varmus said at a press conference that breast and ovarian cancers are extremely complicated diseases that are probably affected by various genetic and environmental factors. Several NIH entities, led by the National Center for Human Genome Research, have recently awarded research groups more than $2.5 million to study breast cancer testing and the social, psychological, and economic implications of such tests.

More than 45,000 U.S. women and 300 men died of breast cancer last year. BRCA1 is linked to about 5% of the 180,000 breast cancer cases diagnosed annually in the United States and to 25% of those in women under the age of 30. About 600,000 U.S. women and millions around the world may carry BRCA1 mutations.

Tumor Suppressor

BRCA1 is believed to act as a tumor suppressor regulating cell growth and division. If suppressor genes are lost or damaged by mutation, uncontrolled cell growth can occur, resulting in cancer. Researchers reported finding several different mutations in all family members who inherited the faulty BRCA1 gene and in those who developed breast cancer at an early age; they also observed the mutation in many women with both breast and ovarian cancers. Before diagnostic tests can be developed to detect the aberrant gene, each specific mutation must first be identified.

Although inheriting a mutated BRCA1 gene was found to increase dramatically the chance of developing breast cancer, scientists do not yet understand why only 15% of such women escape cancer, even in extreme old age. This suggests that even susceptible women may be influenced by crucial genetic or environmental factors. Identifying these factors will be critical in developing prevention strategies.◊

The LANL data library evolved into the database GenBank when in 1982 Bolt, Beranek, and Newman (BBN) became the primary contractor for distribution of data and user support. LANL became a subcontractor to BBN, providing data collection and design. Sequence data activities at BBN and LANL were funded by the NIH National Institute of General Medical Sciences (NIGMS) with support from DOE and other agencies. At the end of the first 5-year contract, IntelliGenetics became the primary contractor, with LANL again as subcontractor charged with designing and building the database.

With the conclusion of the second 5-year contract in October 1992, NIGMS transferred its responsibility for the GenBank project to the National Center for Biotechnology Information (NCBI) at the National Library of Medicine. NCBI, directed by David Lipman, had been created in November 1988 to develop automated information systems for supporting biotechnology and molecular biology and to conduct basic research in computational molecular biology. LANL continued to handle all direct GenBank submissions and updates through a DOE-NIH interagency agreement.

In August 1993 the data resources at LANL and NCBI became independent of each other, with both providing collection and distribution services and continuing to enhance access and usability of the databases. The GenBank database remained at NCBI, and the LANL service led by Michael Cinkosky took the new name of Genome Sequence Data Base (GSDB). GSDB recently moved to the National Center for Genome Resources (NCGR) in Santa Fe, New Mexico, which focuses on the development of resource projects to support public and private genome research.

The EMBL Nucleotide Sequence Database was established in 1982 as the EMBL Data Library in Heidelberg, Germany. Directed by Graham Cameron, the database is now maintained and distributed by the European BioinformaticsInstitute (EBI), a new EMBL outstation at Hinxton Hall near Cambridge, U.K. The Sanger Centre and the Medical Research Council's Human Genome Mapping Program Resource Centre are also located at the site. In addition to the Nucleotide Sequence Database, EBI maintains and distributes the SWISS-PROT Protein Sequence Database in collaboration with Amos Bairoch (University of Geneva) as well as more than 30 other specialty databases.

DDBJ was created in 1984 and began operating independently in 1986 with the sponsorship of the Japanese Ministry of Education, Science, and Culture and representative Japanese molecular biologists. Directed by Yoshio Tateno, DDBJ is accumulating nucleotide sequence data, mostly submitted from Japanese researchers. In addition to the sequence database, DDBJ makes available 15 other databases through its Gopher system.

Data Sources and Submission

During the first few years of database operation, all data were collected by scanning published articles for DNA or RNA sequence data, which were then typed into a computer and distributed in both electronic and printed form. The increase in data, however, soon began to overwhelm data processors, causing a delay between publication and appearance in the database. (See graph below.) Also, some researchers became concerned that much sequence data would never be published because journals began limiting the amount they would print, and authors left out the sequences they considered less important.

To alleviate these delays and problems, workers at EMBL, LANL, and IntelliGenetics developed an electronic data-publishing approach and encouraged authors to deposit sequences directly into databases before submitting their results for journal publication. Most journal editors now require such prior submissions, although an author may request that the data not be released until the article appears in print.

Nearly all data are now acquired through direct submissions to one of the four databases, where they are received, processed, and shared with the others. Groups generating large volumes of data can arrange a procedure with the databases to simplify submissions. A small amount of data still enters the databases via journal scanning, done at NCBI, and this is also shared immediately. Under a contract with the European Patent Office, EMBL draws on sequence data reported in the patent literature since 1960. NCBI captures corresponding data from U.S. patents, and DDBJ Release 18 (July 1994) contains 4551 entries processed by the Japan patent office.

The most commonly used direct-submission mechanism is the Authorin automatic-processing program. Developed by IntelliGenetics, Authorin guides users through the process of entering sequence and providing biological and bibliographical annotation. Authorin software can be obtained on diskettes from NCBI (see box, p. 4) or by ftp from the databases and many online repositories of biological software. Early next year, NCBI will offer an alternative to Authorin with a new point-and-click-style program called Sequin, which will have interface improvements for both stand-alone and network users. Network users will have live access to GenBank and MEDLINE for more complete annotation of their sequences.

GSDB uses the Annotator's WorkBench (AWB) software, which allows offsite users with Internet access to have full and immediate control over data submission, annotation, and release, thus offering an advance over batch-mode submission. Offsite users who are unable to run the graphical version of AWB are issued accounts on one of GSDB's machines, from which they can run AWB 2.x. Also, centers with in-house Sybase expertise can write special applications that perform updates directly on the master GSDB database using client-server access.

Important sources of direct submissions to NCBI include numerous expressed sequence tags (ESTs), which are partially sequenced cDNAs that are stored in the dbEST database. To facilitate access to and simplify comparisons of sequence tagged sites (STSs) with sequences in other divisions, NCBI recently created a separate database (dbSTS) that provides detailed information about STS map locations and polymerase chain reaction conditions.

EMBL has established submission accounts for groups producing large volumes of nucleotide sequence data over an extended period, a procedure that has proven flexible and efficient both for database staff and a number of genome research groups. DDBJ recently developed a test version of a relational database system on Sybase for large data submitters such as human and rice EST projects.

LBL WWW Software Tools

Members of the Data Management Research and Development Group at Lawrence Berkeley Laboratory (LBL) have developed a suite of software tools for defining, browsing, and manipulating data in genomic databases. These tools are based on the Object-Protocol Model (OPM) of I-Min A. Chen and Victor M. Markowitz. OPM allows users to model objects and protocols specific to genomic database applications, specify protocols in terms of alternative sequences of steps, and define different database views. OPM and the OPM tools are being used to develop version 6 of the Genome Data Base and are extended with inter-database cross-reference and data-versioning facilities.

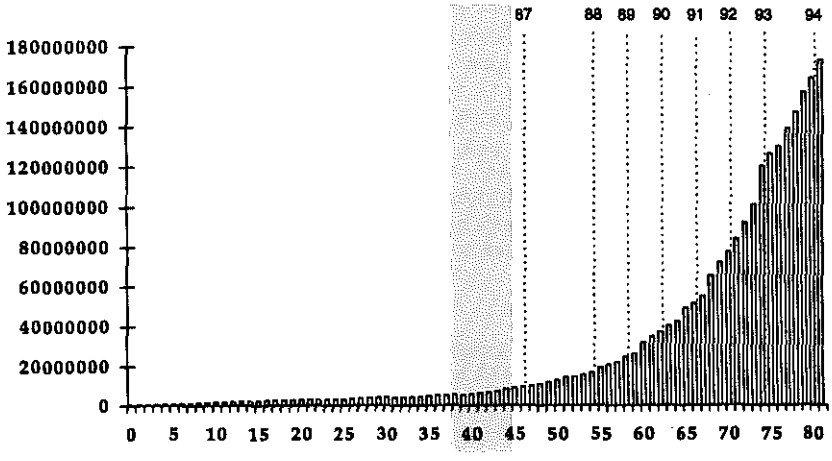
The OPM data-management tools, which target relational database management systems such as Sybase 10 and Oracle 7, include:

- a graphical editor for specifying OPM schemas,
- a translator of OPM schemas into relational database schemas and SQL queries,
- a graphical data-browsing and data-entry tool, and
- a tool for converting definitions of existing relational database schemas into OPM schema definitions.

Tools, documents, and examples are available via WWW using the URL *ftp://gizmo.lbl.gov/pub/DM_TOOLS/OPM/opm.html*. To be notified of future releases, contact *vmmarkowitz@lbl.gov*.◊

Database Use and Access

At present, users of sequence databases typically want to retrieve records based on sequence



Growth in the world's collection of nucleotide sequence data is shown as the number of bases contained in every release of GenBank from 1 through 82. The numbers at the tops of the dotted lines show years (which do not necessarily coincide with a particular number of releases). The shaded bar in the middle represents the period in the mid-1980s when the data volume was, for a time, more than the databases could handle.

Genome News

similarity—which can offer clues to gene sequence functions—or on keywords. For sequence similarity searching, computer programs are used to compare a query sequence with a subset of the database and find statistically meaningful alignments. To retrieve by other criteria such as keywords, gene name, or gene product function, users can search text descriptions. Databases are increasingly important for facilitating gene searches and comparing annotated information to detect sequence relationships that have not been determined experimentally.

Each sequence database record corresponds to a continuous piece of DNA, the largest of which is about 685 kb from the human T-cell receptor. Typical database entries contain in flat-file format a concise sequence description; the taxonomic description of the source organism; bibliographic information; a table of features listing the locations of biologically significant elements such as protein-coding regions, transcription units, mutations, or modifications; and protein translations of coding regions. Each entry is curated by database staff, who check for biological consistency (e.g., coding sequences should not contain "stop" codons). When appropriate, entries may also be cross-referenced to other databases; for example, EMBL has established cross references for SWISS-PROT, Eukaryotic Promoter Database, Transcription Factor Database, and FlyBase.

A variety of methods are used to distribute and access these databases, including magnetic tapes, CD-ROMs, e-mail, and Internet. Data is now accessible through information services such as WWW, Gopher, and WAIS (Wide Area Information Server).

EMBL Data Library

The EMBL sequence database is available via network services and European Molecular Biology Network nodes (EMBnet; 19 sites). EMBL, SWISS-PROT, and a number of other databases distributed by EBI are accessible via EBI network servers and included in quarterly CD-ROMs. For querying the sequence databases, EMBL-Search for Macintosh or CD-SEQ for DOS is supplied with the CD-ROMs. Sequence databases are provided in the format for use with software such as FastA on Macintosh or MS-DOS systems, and EMBLScan is supplied for rapid searching for very similar sequences.

EBI Network Fileserver enables access via e-mail to the full collection of databases, public-domain software, and documentation maintained by EBI (see box, p. 4, for address). For

Database Distribution and Access Details

DDBJ

[National Institute of Genetics; Yata, Mishima, Japan 411 (+81-559/75-0771, Fax: -6040).]

- General inquiries: *ddbj@ddbj.nig.ac.jp*
- Data submission: *ddbjsub@ddbj.nig.ac.jp*
- Data updates: *ddbjupdt@ddbj.nig.ac.jp*
- FastA e-mail server: *fasta@nig.ac.jp*
- FastA e-mail server trouble report: *trouble@nig.ac.jp*
- Ftp server: *ftp.nig.ac.jp or 133.39.16.66*
- Help: README on */directory* and README.dna on */dna directory*

[Newsletter: A substantial part of *DDBJ News Letter* is written in English. Subscription contact: Naruya Saitou, Editor (Internet: *nsaitou@ genes.nig.ac.jp*).]

EMBL Database

[European Bioinformatics Institute; Hinxton Hall; Hinxton, Cambridge CB10 IRQ, U.K. (+44-1223/494-400, Fax: -468).]

- Data submissions: *datasubs@ebi.ac.uk*
- Entry corrections: *update@ebi.ac.uk*
- General inquiries: *datalib@ebi.ac.uk*
- E-mail file server: *netserv@ebi.ac.uk*
- Network server help: *nethelp@ebi.ac.uk*
- Ftp: *ftp.ebi.ac.uk*
- Gopher: *gopher.ebi.ac.uk*
- WWW: *http://www.ebi.ac.uk*
- MPsrch protein sequence search server: *blitz@ebi.ac.uk*
- FastA sequence search server: *fasta@ebi.ac.uk*
- Quicksearch sequence search server: *quick@ebi.ac.uk*

GenBank Database

[NCBI; National Library of Medicine; Bldg. 38A; 8600 Rockville Pike; Bethesda, MD 20894 (User Services: 301/496-2475, Fax: /480-9241, Internet: *info@ncbi.nlm.nih.gov*).]

- Sequence Submissions: *gb-sub@ncbi.nlm.nih.gov*
- Sequence Updates: *update@ncbi.nlm.nih.gov*
- Authorin assistance: *authorin@ncbi.nlm.nih.gov*
- Retrieve e-mail server: Queries: *retrieve@ncbi.nlm.nih.gov* Help: *retrieve-help@ncbi.nlm.nih.gov*
- BLAST sequence similarity e-mail server: Queries: *blast@ncbi.nlm.nih.gov* Help: *blast-help@ncbi.nlm.nih.gov*
- Bulk EST or STS submission information: *info@ncbi.nlm.nih.gov*
- Searching dbEST or dbSTS: Help: *retrieve-help@ncbi.nlm.nih.gov*
- Network *Entrez* information: *net-info@ncbi.nlm.nih.gov*
- Network BLAST information: *blast-help@ncbi.nlm.nih.gov*
- WWW-NCBI Home Page: *http://www.ncbi.nlm.nih.gov*
- Anonymous ftp: *ncbi.nlm.nih.gov* or *130.14.25.1* (log in, *anonymous*; password, user e-mail address)
- CD ROM subscription info, order forms: *info@ncbi.nlm.nih.gov* *Entrez* CD-ROM help: *entrez@ncbi.nlm.nih.gov*

[Newsletter: *NCBI News* (see NCBI contact information at left under GenBank).]

GSDB

[National Center for Genome Resources; 1800 Old Pecos Trail; Santa Fe, NM 87505 (505/982-7840, Fax: -7690).]

- Relational satellites: *gsdb@gsdb.ncgr.org*
- WWW server: *http://www.ncgr.org/gsdb*
- General information: *gsdb@gsdb.ncgr.org*
- Sequence submissions: *datasubs@gsdb.ncgr.org*
- Updates and corrections: *update@gsdb.ncgr.org*
- Off-site user accounts: *offsite@gsdb.ncgr.org*
- Obtaining Authorin: *ftp ftp.ncgr.org*
- Authorin information: *authorin@gsdb.ncgr.org*
- Ftp server: *ftp.ncgr.org*

GSDB software and documentation, including the complete relational schema manual: ftp (*ftp.ncgr.org*) or *http://www.ncgr.org/gsdb*

molecular biology databanks, the EBI WWW server will soon offer the SRS network browser. SRS will allow interactive querying of the EMBL Nucleotide Sequence, SWISS-PROT, PIR, and NRL3D databases, with hyperlinks to cross-referenced entries in several specialized molecular biology databases distributed by EBI.

GSDB

GSDB emphasizes online, networked data access, offering a WWW server for individual users and a fully functional relational database for other developers. Entries in its WWW server are hyperlinked to an array of external sources, including Genome Data Base, SWISS-PROT, and the Enzyme Commission catalog maintained on the WWW server at Johns Hopkins University (JHU). An online relational server continuing the GSDB database is available at NCGR and many satellite sites around the world. Anyone with a Sybase front-end license may access a read-only copy of GSDB at NCGR using either generic database-access tools or special-purpose programs. GSDB may also be searched using the GenQuest system (p. 6).

(see Databases, p. 6)

Whitehead, Généthon Groups Present New Linkage Maps

Groups at Généthon and Whitehead Institute–Massachusetts Institute of Technology (MIT) Center for Genome Research (CGR) recently reported significant progress in constructing genetic linkage maps of the human and mouse genomes [*Nature Genetics*, Special Issue **7**, 217–27, 246–339 (June 1994)]. Citing the tremendous value of automating repetitive steps for map construction, researchers foresee completion of both maps by the end of the year. Increased marker density in the new maps is expected to make gene searches more efficient.

The Généthon team, headed by Jean Weissenbach, presented a human genetic linkage map featuring over 2000 new microsatellite markers spaced on average 2.9 cM apart and integrated into their 1992 map. Another 1000 markers were recently submitted to the Genome Data Base. Généthon researchers plan to produce a map with 5000 markers in spring 1995.

Eric Lander's group at CGR has placed over 5000 simple sequence length polymorphisms (SSLPs) on the mouse genetic map, with an average spacing of 0.30 cM. This represents a 13-fold increase in marker density over CGR's 1992 mouse genetic map, making this the most dense SSLP map constructed for any organism. Because many genes are conserved in both mouse and human, mouse genome maps are considered extremely valuable resources in elucidating the human genome.

After completing a 6000-marker mouse genetic map, CGR will begin constructing a physical map of the mouse genome based on the sequence tagged site (STS) content of mouse yeast artificial chromosome (YAC) libraries. A total of 10,000 STSs will be used, consisting of the 6000 SSLPs from the genetic map plus 4000 random STSs. CGR is also constructing a human physical map based on screening 10,000 STSs on YACs from the CEPH mega-YAC library.

[Généthon contact: Jean Weissenbach; Institut Pasteur; CNRS URA-1445; 25 rue du Docteur Roux; F-75724 Paris Cedex 15, France (Fax: +33-1/6077-8698, Internet: *jsbach@genethon.fr*). CGR contact: Eric Lander; Whitehead Institute/MIT; One Kendall Square, Bldg. 300; Cambridge, MA 02139-1561 (617/252-1906, Fax: -1933, Internet: *lander@genome.wi.mit.edu*).]

Resource Distribution

CGR

Mapping Data. CGR human physical mapping and mouse genetic mapping data are released quarterly, usually within 2 weeks of the quarter's close, and announced by electronic messages posted to the newsgroups *bionet.genome.chromosomes* and *bionet.announce*. Data releases are accessible in the following ways:

- Ftp to *genome.wi.mit.edu* (use *anonymous* for user name and user's e-mail address for the password). Data files are stored in */distribution/mouse_sslp_releases* and */distribution/ human_STS_ releases*.
- E-mail to *genomedatabase@ genome.wi. mit.edu* with *help* as the first word on the subject line or body text. Only the mouse genetic information is currently available via this route.
- WWW (see URL for newsletter).

[Help with database services: Lincoln Stein (617/252-1916, Internet: *lstein@genome.wi.mit.edu*.]

Biological Reagents. To ensure broad and immediate access to mouse and human STSs, in 1990 CGR agreed that Research Genetics, Inc., would retain a portion of all PCR primer pairs synthesized for CGR use and sell aliquots to the scientific community at discount prices. This arrangement now extends to PCR primers from other genome centers. CGR also distributed copies of its mouse YAC library and the CEPH mega-YAC library to both Research Genetics and Genome Systems, Inc.

[Contacts: Research Genetics, Inc.; 2130 Memorial Parkway SW; Huntsville, AL 35801 (800/533-4363, Fax: 205/536-9016). Genome Systems, Inc.; 7166 Manchester Rd.; St. Louis, MO 63143 (800/248-7609, Fax: 314/647-4134).]

Newsletter. In July, CGR began publishing a quarterly newsletter to provide up-to-date information about center progress and resources. The *CGR Newsletter* is available on WWW (URL *http://www-genome. wi. mit. edu/*) and in hard copy. (Subscriptions: Newsletter Editor; Whitehead Institute/MIT; One Kendall Square, Bldg. 300; Cambridge, MA 02139-1561 (Internet: *newsletter@ genome.mit.edu*).] ◊

1993–94 Généthon Map

The 1993–94 Généthon human genetic linkage map described in the first paragraph is available electronically:

- Anonymous ftp: *ftp-genethon.fr.*
- Map directory: *pub/Gmap/ Nature-Genetics-1994.*
- Genotype directory: *pub/Gmap/genotypes.*

[Contact: *fizames@genethon.fr*] ◊

1993 CEPH-Généthon Physical Map

The 1993 CEPH-Généthon physical map announced in the December 16, 1993, issue of *Nature* is also available electronically:

- Anonymous ftp: *ceph-genethon-map.cephb.fr* in directory */pub/ ceph-genethon-map*
- WWW: *http://www.cephb.fr:/bio/ ceph-genethon-map.html* or *http://cartagene.cephb.fr:/bio/ ceph-genethon-map.html*
- E-mail: *ceph-genethon-map@ cephb.fr* or *ceph-genethon-quickmap@ cephb.fr*

[Contacts: *ceph-genethon-map@ cephb.fr* or *denis@ceph.cephb.fr*] ◊

Genome News

NCBI

NCBI has introduced several new services to facilitate public use of GenBank, including the retrieve and BLAST e-mail servers and WWW access to sequence and bibliographic data. Since 1992, NCBI has offered access to gene and protein sequences and related MEDLINE bibliographic citations via a graphical user interface. A combination of the three integrated databases and retrieval software—*Entrez*—is available on CD-ROM, as an Internet client-server application, and via WWW (see box, p. 4). The power of *Entrez* is in the precomputed links among the constituent databases; these links allow users to retrieve a DNA sequence by searching for text terms, author name, or accession number and to look up associated protein and MEDLINE citations by clicking a button. BLAST sequence similarities have also been precomputed for every DNA and protein sequence.

GenQuest Sequence Analysis Service

GenQuest is a sequence analysis service that acts as "middleware" connecting the user with many databases and integrating networked resources into one easy-to-use system. With a Mosaic interface created through a collaboration between ORNL and JHU, the GenQuest graphical client sends a properly formatted request to the Oak Ridge National Laboratory (ORNL) online server. The server uses FastA, BLAST, or full Smith-Waterman to analyze several databases (e.g., GSDB, SWISS-PROT, and PDB), and search results are returned quickly as a standard Mosaic page with hot links established to all referenced data objects in the report.

This integrated resource is possible because the online ORNL GenQuest server can return analyses quickly enough to support a real-time interface; also, all pertinent databases provide data with standard network-accessible protocols. The GenQuest service is an example of how distributed information resources can be combined easily and economically into important tools for the community. *[Anne Adamson and Denise Casey, HGMIS]* ◊

Future Plans for Databases

Integrating sequence data with mapping, structural, and other biological information will require the development of "virtual databases," with components originating at multiple sites. Some informatics scientists envision creating such a virtual system on top of interlocking community databases that form a loosely coupled information infrastructure. Other research efforts are aimed at developing a local collection of primary databases that act as a single virtual database. Everyone agrees that tools should allow users with only minimal computer knowledge to access many resources and that linkages among primary databases should provide a "one-stop-shopping" capability that eliminates the need for separate queries to each database.

An example of such a virtual database may be found at the WWW site maintained by the Baylor Human Genome Center (*http://gc.bcm.tmc.edu:8088/*). Choosing the "Biologist's Control Panel" from Baylor's home page produces a list of more than 20 molecular biology databases. Selecting "BLAST search of GenBank" connects users to NCBI's service in Bethesda, Maryland. Similar hyperlinks are made to GDB, SWISS-PROT, and many other sites.

NCBI's future plans focus on improving GenBank data quality and continuing to provide easy-to-use yet powerful methods for data access. To accomplish this goal, a new GenBank fellowship program has recruited five molecular biologists, who are working with informatics specialists on specific tasks.

GSDB is focusing on online database access by offsite users and direct client-server updates, especially for genome centers. GSDB and EBI also favor establishing connectivity for a federation of interoperable molecular biology databases communicating across computer networks. Because this plan may not require the same level of data standardization as monolithic approaches, it would allow greater autonomy for participating databases. To facilitate such communication, software packages such as EMBL-Search and SRS use cross references built into EBI-distributed databases to allow retrieval of related data.

GSDB anticipates that, although databases are now providing rapid processing of batch submissions, these methods will require too much manual effort to meet the expected future data flow. Capitalizing on the rapid expansion of network availability, GSDB systems are being redesigned to be interactive and network based. To keep up with raw sequence data and the rapidly expanding understanding of sequence function, maintaining database entries will not be limited to original authors but, through support for third-party annotations, will be open to anyone interested. Much future database work may be performed online by the scientific community using client-server tools.

¶ Publications

Gene Wars

The Gene Wars: Science, Politics, and the Human Genome by Robert Cook-Deegan (National Academy of Sciences) is a detailed, comprehensive, firsthand account of events, politics, and personalities involved in developing and implementing the Human Genome Project. The book describes the technological advances that led to the project and includes interviews with many of the participants. An extensive section is devoted to social, ethical, and legal issues. 1994, 416 pp. [Available in bookstores or from the publisher: Norton & Company, Inc.; 500 Fifth Ave.; New York, NY 10110 (212/354-5500, Fax: /869-0856).] ◊

New Frontiers

On the New Frontiers of Genetics and Religion by J. Robert Nelson (Texas Medical Center, Houston) draws on the work of 260 scientific, medical, and religious professionals who met in 1990 and 1992 under the auspices of the Human Genome Project to discuss genetic research and related issues. The author explores such topics as genetic counseling, prenatal diagnosis, treatment of inherited disease, special genetic concerns of women, and the temptations to seek eugenic improvement of human nature and capabilities. Religious critiques by leading experts from Jewish, Christian, and other traditions explain the possibilities for good and the dangers of abuse in human genetic science. Paper, 215 pp. 1994. [Wm. B. Eerdmans Publishing Co.; 255 Jefferson S.E.; Grand Rapids, MI 49504 (800/253-7521, Fax: 616/459-6540).] ◊

Reference Library, Database

Günther Zehetner, Hans Lehrach, and colleagues at the Genome Analysis Laboratory of the U.K. Imperial Cancer Research Fund (ICRF) have developed the Reference Library System (RLS) and its associated Reference Library DataBase (RLDB2). RLS is a high-density genome-mapping system based on the use of common reference libraries that provide simplified access to clones and allow efficient integration of information created in different types of experiments by many scientists.

The distribution service of RLDB sends out high-density library filters to participating laboratories, where investigators hybridize them to their own unique or complex probes and identify the clones containing the probe sequences. Results on probes and positive clones are returned to RLDB, where they are entered into an object-based database (RLDB2), which is accessible online via Internet. The laboratories receive the identified clones for further analysis, giving quick access to the clones and any information that might already be available about them.

The cosmid, P1, YAC, and cDNA libraries, which are freely available to other laboratories for noncommercial use, are distributed on filter grids carrying up to 20,000 clones per 22- by 22-cm nylon membrane. A fee is charged to laboratories outside the European Community to recover filter production and mailing costs.

RLDB2

RLDB2 can store many different experimental (libraries, clones, filters) and informational (contigs, maps, images) objects and their relationship to each other. A prototype information server on WWW allows users to query the database and returns the retrieved data as HTML documents, which can contain text and images as well as hyperlinks to files, other RLDB2 objects, and information in external databases such as GDB, OMIM, and GenBank®. Users can also view RLDB files, request RLS items, return hybridization results, or connect to other biological servers. The system permits nonregistered users to view public data and registered users to access and update their own private data as well.

Access to RLDB

Via Anonymous Ftp

From the NCBI data repository at *ncbi.nlm.nih.gov (130.14.20.1)* in the following directories:

/repository/RLDB (general information)
/repository/RLDB/RLDBlist (probe lists)
/repository/RLDB/RLDBirx (IRX files)

Via WWW

As clients, the RLDB server requires WWW browsers that can use forms (Mosaic, MacWeb, Lynx). To access RLDB2, point the client to the URL *http://gea.lif.icnet.uk/.*

Lists of all used probes, their origin, chromosome location, and number of identified potential or confirmed cosmid, P1, or YAC clones are generated once a week and automatically downloaded to the National Center for Biotechnology Information data repository in Bethesda, Maryland; they can be accessed by anonymous ftp. The public RLDB data are also made available as an IRX (Information Retrieval Experimental Workbench) database from the data repository (see box for addresses). [Contact: Reference Library DataBase; ICRF; 44 Lincoln's Inn Fields, RCS; London WC2A 3PX, U.K. (Fax: +44-71/269-3645, Internet: *genome@icrf.icnet.uk*).] [Reported in *Nature* 367(6462), 489–91 (February 3, 1994).] ◊

OHER Launches Microbial Genome Initiative

In a spin-off from the Human Genome Project, the DOE Office of Health and Environmental Research (OHER) is launching the Microbial Genome Initiative (MGI) to provide genome sequence and mapping data on selected microorganisms. MGI will focus on industrially important microbes and those that live under extreme conditions, including the deep subsurface, geothermal environments, and toxic waste sites. OHER expects that information gained through this project will further the understanding of microbial phylogeny, physiology, and structural biology and help to exploit such industrial opportunities as the cleanup of process and environmental waste.

In MGI's first year, investigative groups will sequence the following organisms:

- *Pyrococcus furiosus*, a marine hyperthermophile (optimum growth temperature 100°C) with an A+T–rich genome of about 2 Mb [Robert Weiss (University of Utah)].
- *Methanococcus jannaschii*, an extreme thermophile and marine barophilic methanogen with a 2-Mb A+T–rich genome [Craig Venter (The Institute for Genomic Research) and Carl Woese (University of Illinois)].
- *Methanobacterium thermoautotrophicum*, a sewage sludge archaeon that grows optimally at 65°C and has a genome of about 1.7 Mb and a G+C content of 50%. Much of the bioconversion biochemistry of CO_2 to CH_4 is based on this archaeon [Doug Smith (Genome Therapeutics Corp.) and John Reeve (Ohio State University)].◊

Evans Moves NIH Genome Center to Dallas

Glen Evans, formerly at the Salk Institute for Biological Studies, has moved his human genome work to the University of Texas Southwestern Medical Center at Dallas [5323 Harry Hines Blvd.; Dallas, TX 75235-8591 (214/648-1660, Fax: -1666, Internet: *gevans@ mcdermott.swmed.edu*)]. More information about his project, other NIH GESTECs, and DOE genome centers will be published in the November issue of *HGN*.◊

Mouse Mapping Data Searchable on WWW

The Mouse Genome Database at Jackson Laboratory, Bar Harbor, Maine, is available in searchable form on WWW. The database includes mouse locus information; data on genetic mapping, mammalian homology, probes and clones, and PCR primers; and over 20,000 references. URL: *http://www.informatics.jax.org/.*

[Mouse Genome Informatics User Support (207/288-3371, ext. 1900, Fax: -2516, Internet: *mgi-help@informatics.jax.org*).] ◊

Genome News

Law Mandates Survey of Workers Who May Have Been Exposed

DOE Biomarker Workshop Meets in Santa Fe

The DOE Second International Workshop on the Development and Application of Biomarkers was held on April 26–28 in Santa Fe, New Mexico. Over 100 U.S. and foreign scientists, occupational physicians, radiation biologists, and government officials attended the meeting, which was supported by the DOE offices of Energy Research; Environment, Safety, and Health; and Environmental Management. The stimulus for this workshop was Public Law 102-484 (1992), which mandates that DOE survey its past and present workers who may have been exposed to chemicals or radiation.

New technologies developed by the Human Genome Project, as well as societal issues raised by such a large-scale health monitoring program, were among the issues discussed at the DOE workshop. Selected presentation highlights follow.

Detecting Exposure and Its Effects

Anthony Carrano (Lawrence Livermore National Laboratory) reviewed the study of human biomarkers, characterizing them as both exciting and frustrating. He commented that a variety of approaches have been explored for detecting biological effects of exposure to external agents; these include somatic cell cytogenetics, sperm damage, gene mutation endpoints, DNA adducts, studies of mechanisms and population, animal models, radiation, and chemical damage measurements.

Mortimer Mendelsohn (Radiation Effects Research Foundation, Japan) reviewed his extensive biomarker research into the effects of radiation on atomic bomb survivors, in whom some 10,000 cancers have been documented. Using chromosome aberration studies, Mendelsohn has shown only a very slight correlation of aberrations with cancers, an indication that the studied lymphocytes may have "forgotten" about the initial radiation exposure and that compensatory mechanisms work remarkably well. Mendelsohn noted that about half of all carcinogens are not mutagens and that validation of novel biomarkers is likely to be a serious stumbling block for research.

Charles Cantor (Boston University) said the Human Genome Project should be a rich source for new methods, discoveries of relevant human genes, identification of DNA regions particularly sensitive to various kinds of damage, integration of major databases, and spinoff technologies in unexpected areas. Advantages to using DNA as a biomarker include its uniqueness and relative stability, usefulness as a label for other molecules, and the ready detectibility of single molecules. Biomarkers could be used in such applications as surveys of organisms at toxic sites, unique identifier tags for released organisms in environmental modification or waste cleanup, and indicators of external damaging agents.

Paul Schulte (National Institute for Occupational Safety and Health) pointed out that markers of exposure may not be equivalent to markers of effect. In a major goal of occupational-disease prevention—reducing exposure—biomarkers would have no prominent role but could be useful in reducing the effects of exposure through medical monitoring. Schulte agreed that, before any medical screening program is started, ethical safeguards should be established. Also, relevant diseases should be significant and treatable, and medical tests should not be inordinately invasive, painful, or costly. He also felt that good consultation and counseling should be available, and tests should be targeted to specific risks. Test validity should be demonstrated first, both at the laboratory and population levels, and the underlying prevalence of the condition established. Although genetic screening is controversial, it could be the most powerful tool in conducting such tests.

Larry Clevenger (Sandia National Laboratory) noted that exposure to an environmental agent does not necessarily imply an effect. This is due to individual and immunological variation, different decay rates for exposure effects, and a very wide continuum for health and disease among people—all of which factors are important in discussions of genetic effects. Thus any correlations between biomarker results and future diseases may be only partial at best and include an irreducible chance component.

Assessing Disease Risk

Richard Albertini (University of Vermont Cancer Center) discussed three types of biomarkers for individual disease risk (see box). Markers in these categories include somatic mutations in reporter genes, chromosomal aberrations, gpa variant frequencies, hprt mutations (mutational spectra in placental blood), and hprt mutational spectra for "specificity." Albertini stated that the major difficulty in using these as biomarkers for an individual's risk is that many current risk estimates are based on population studies, from which an average is determined.

Paul Brandt-Rauf (Columbia University) discussed the usefulness of p21, a product of the K-*ras*-2 oncogene, as a biomarker. A human version of p21 is located on human chromosome 12p11.1. Alterations or mutations in K-*ras*,

Biomarkers

Researchers use biological indicators, called biomarkers, to detect events in biological systems that may be associated with exposure to environmental agents. Examples of biomarkers include changes in genetic material; cell death; and discovery of the environmental substance itself or its metabolites in urine, blood, or expelled air. For ascertaining individual disease risk, biomarkers can be grouped into three broad categories—exposure, biological effect, and susceptibility. The biological events they detect can represent variation in the number, structure, or function of cellular or biochemical components.

Biomarkers and other resources and technologies developed in the Human Genome Project will have a major impact on the study of environmental risk factors. The basic aim of scientists exploring these issues is to determine the nature and consequence of genetic change or variation, with the ultimate purpose of predicting or preventing disease. Most current tests for human exposure to environmental mutagens are only indicators of genetic damage, however. Genetic toxicology and mutation research will focus on advancing beyond this measurement stage to the study of genes and genetic variation. The ability to directly and quickly sequence DNA will revolutionize mutation research and lead to insights into the relationship between exposure and disease development. Genetic toxicology data from these studies could be coupled with medical information to diagnose disease onset and develop therapeutic strategies.

which appears to be involved in a signal transduction pathway, could disrupt the pathway and lead to carcinogenesis. Some 80% of patients with liver angiosarcoma, associated with the known carcinogen vinyl chloride, have the aspartic acid–associated GAC codon, which is present in none of the controls. In one patient, semiquantitation of serum p21 served as a crude biomarker, with some predictive value, until the patient died.

Promising Technologies

James Jett (Los Alamos National Laboratory) characterized flow cytometry (FCM) as rapid, capable of handling large numbers of samples, highly flexible, and very sensitive when compared to conventional microscopy. FCM can measure fluorescence; light-scattered, cell, or particle volume; and other optical properties. FCM probes exist for various targets including DNA, RNA, proteins, [Ca], pH, membrane fluidity, and surface and intracellular antibodies. Up to 32 measurements/cell can be realized at a flow rate of 100,000 cells/s, and various analysis levels and a number of applications could be useful for biomarker measurements. FCM's extensive capabilities could also include probes with wider fluorescent emissions and even FCM on a chip.

In reviewing molecular cytogenetics and biomarker development, Joe Gray (University of California, San Francisco) stated that an important approach is fluorescent in situ hybridization, called chromosome painting when used on entire chromosomes. With this technology, 48,000 to 50,000 metaphases can be scored each day, allowing chromosomal translocations to be directly visualized and counted. The greater the radiation dose, the fewer metaphases are required for detection of abnormalities. Fibroblasts are a convenient and successful target for studies to assess external agent damage. One approach in finding early lesions for cancer cell detection is whole genome subtraction, in which tumor DNA is "subtracted" from whole genomic DNA, identifying differences that may be associated with the tumor phenotype. DNA changes can be mapped to within 10 Mb. An expected outcome of the Human Genome Project is the identification of chromosomal changes associated with some cancers.

Chronic Beryllium Disease

A number of speakers discussed the health impact of beryllium exposure, which in susceptible individuals can lead to an autoimmune-like pulmonary disease. Beryllium is used industrially in the manufacture of light fixtures and certain ceramics. Milton Rossman (University of Pennsylvania) reviewed the immunology of chronic beryllium disease (CBD), first described in 1946 as a "delayed chemical pneumonitis." CBD also causes granulomas in the skin, liver, lymph nodes, and conjunctiva. Steroids are effective if CBD is diagnosed early, but if treatment is delayed, collagen deposits lead to scarring.

Cesare Saltini (University of Modena, Italy) reviewed his research showing that CBD appears to be mediated by beryllium-specific C4 positive T-cells, the same T-cells affected by the HIV virus in AIDS patients. The population frequency of the genetic marker associated with CBD susceptibility is about 30%, while in CDB patients this frequency is nearly 100%. However, CBD incidence in beryllium-exposed populations ranges from 0.4 to 4.9%. This pattern is typical of many genetically influenced (particularly autoimmune) diseases in that a large proportion of CBD patients have a given marker but few with the marker have the disease.

Lee Newman (National Jewish Center for Immunology and Respiratory Medicine, Denver) described CDB detection through the lymphocyte transformation test (LTT). In the presence of beryllium, lymphocytes from sensitized individuals can transform and incorporate an isotope, allowing the degree of sensitization to be measured. However, some people with clearly abnormal LTT blood results have no lung disease. Because sensitization is thought to precede disease, LTT might assist in accurate and early diagnosis.

Emerging Issues

A number of important issues emerged from this workshop. Researchers felt that a considerable gap exists between the demand for detailed and comprehensive data about medical implications of workplace hazards and the current ability to compile relevant biomarker-based information. Existing biomarkers are far from ideal, and their actual significance remains to be worked out. To make biomarkers more useful, attendees saw the urgent need for better population measurements; data on dose reconstruction and low doses, particularly the shape of dose-response curves; and studies of at least 755 little-known industrial chemicals.

More information is also needed on the distribution of risks in individuals other than the "average" working male and on what constitutes an "acceptable" risk. How are the risk-benefit analyses to be accomplished? What sort of communication and education campaign would support appropriate and reasonable research?

Many speakers recognized that ethical, legal, and social concerns are important in accomplishing any research agenda and that these sensitivities should be incorporated at the outset. Although new biomarkers offer many opportunities for increased accuracy, sensitivity, and predictability, the ability to measure will outrun the understanding of medical and health implications. Privacy of biomarker information is a general concern related to more-general issues of privacy of medical records and controls on their access.

[Daniel Drell, DOE OHER] ◊

Book Summarizes Meetings

The first biomarker research workshop was held at the National Academy of Sciences (NAS) on September 24, 1993. A book summarizing that meeting and the one reported here will be published this fall by the Joseph Henry Press (the new arm of NAS Press). For more information, contact John Peeters (301/903-5902, Fax: -5072, Internet: *john.peeters@hq.doe.gov*).

Information on Outreach Efforts Sought

For a future newsletter article, *HGN* staff is gathering information about the efforts of U.S. Human Genome Program investigators to educate the public about genetics. Researchers are asked to send a description of their outreach programs—no matter how modest or extensive—to Betty Mansfield at the address on page 12 by December 15.◊

Genome News

OHER Focuses on Education, Privacy, Workplace

DOE ELSI Program Enters Fifth Year

The DOE Ethical, Legal, and Social Issues (ELSI) Program, administered by the Office of Health and Environmental Research (OHER), aims to anticipate and study how individuals and society will be affected by the large amounts of genetic data being generated through the Human Genome Project. Three years ago, OHER narrowed its ELSI focus to concentrate on genetic education, privacy and confidentiality of personal genetic information, and genetics and the workplace [see *HGN* 4(2), 1–2 (July 1992) and 5(2), 3–4 (July 1993)].

Now entering its fifth year, the DOE ELSI Program added three new projects and two continuing ones to its portfolio of sponsored activities in FY 1994. To avoid unnecessary duplication of effort, OHER collaborates closely on program oversight with the ELSI Branch of the NIH National Center for Human Genome Research (NCHGR).

In concert with the NCHGR ELSI Branch, the DOE program supported a recently released study by the Institute of Medicine on a range of ELSI issues, with recommendations for informed policies. Studies were also initiated on the implications of large DNA-based databanks and accumulations of data, including those under development by the Federal Bureau of Investigation, the U.S. Army, certain commercial companies, and academic research centers. Exhibits on genetics are being partially supported by DOE at both the San Francisco Exploratorium and the Smithsonian's Museum of American History.

New Projects

At the University of Michigan Law School, Rebecca Eisenberg is studying the role of patents in transferring technology generated by the Human Genome Project to society at large Eisenberg will review available literature; query industry, government, and university sources about technology transfer; and explore several specific cases to see what works best for rapidly moving new technologies into the marketplace. The results of this study could affect DOE policy far beyond the genome program.

Lee Hood, Valerie Logan, and Maynard Olson (University of Washington, Seattle) have begun an innovative program in which local high school students determine the sequence of STSs (sequence tagged sites) from cloned human genomic DNA. In addition to learning about human genetics, experiencing science firsthand, and contributing to the Human Genome Project by submitting their checked sequences to a DNA sequence database, the students will also explore ethical, legal, and social implications of the project. Insights gained through this experience may encourage some of them to consider the possibility of a scientific career.

At California State University in Los Angeles, Margaret Jefferson and Mary Ann Sesma are translating into Spanish the Biological Sciences Curriculum Study module, "Mapping and Sequencing the Human Genome: Science, Ethics, and Public Policy." They will also introduce it to students in selected Los Angeles high schools. A key element of this approach is to involve parents so that cultural and family sensitivities and values can be incorporated into the study of genetics. In a project that may serve as a pilot for future curriculum development in other subject areas, knowledge about the genome project is being made available to a community not directly addressed by current educational outreach efforts.

Continuing Projects

Troy Duster's "Pathways to Genetic Screening: Patient Knowledge—Patient Practices" is being renewed for a 2-year term. This project contrasts Caucasian understandings about cystic fibrosis with those of African-Americans about sickle cell disease. Early results suggest that communicating genetic information and understanding immediate health implications vary with factors that include social class, gender, and educational level. Duster also reports that detailed information is best obtained through personal contact and discussion in a familiar environment such as the home, rather than through an impersonal survey or doctor's office visit.

The Cold Spring Harbor DNA Learning Center, under director Jan Witkowski, will continue for another year to hold workshops for opinion leaders and public policymakers on genomics and its implications for society. These workshops are aimed at educating individuals who could assist in introducing Human Genome Project information to society. Workshop attendees have included representatives from the media, genetic support groups, law and the courts, Congressional staff, state legislatures, government and private agencies, policy analysis programs, labor unions, and other organizations.

Potential Benefits vs Challenges

The simple, persistent importance underlying ELSI studies is the recognition that each person has a unique genome that both identifies the individual and has predictive implications for future health. An "ideal" or "perfect" genome does not exist, even if such a concept could be defined. All genomes contain polymorphisms that could severely and adversely affect health under different circumstances or if not influenced or masked by other genes; this information about the individual has value to other people and groups who may have their own agendas. Potential benefits of human genome research for the enhanced health and well-being of humankind are very great, but the challenge is to manage this effort wisely and carefully and, if possible, avoid some of the foreseeable problems. *[Daniel Drell, DOE OHER]* ◊

Retrieve New Medline Citations from GDB Via FTP

Each month MEDLINE citations relevant to human gene mapping are loaded into the Genome Data Base (GDB). MEDLINE scanning is a joint project of GDB and Sue Povey's group at University College London.

Three types of files summarizing citation information are available from the ftp server *ftp.gdb.org* in the */litaware* directory. Due to space considerations, only the most recent 12 months of files are kept on this server; please contact *data@gdb.org* if earlier files are needed. The date is given in YYMM format, as follows (filenames are case sensitive):

- <Date>*back.out* and <Date>*mapp.out* files, respectively, contain background and mapping references that are relevant to human chromosome loci as indexed by MEDLINE in the specified month.
- <Date>*list.out* files contain numerical references for citations listed in the back.out and mapp.out files, grouped by chromosome.◊

Search GDB with Graphical Interfaces

Two graphical interfaces to GDB are available on WWW. The original version, developed as part of the Genome Machine project by David Adler and other investigators at the University of Washington, can be used with any graphical WWW client and supports searches for genes at a specified cytogenetic location (URL *http://www.pathology.washington.edu/* under cytogenetics/genome).

The enhanced version, developed by GDB staff, provides buttons to modify the query for greater searching flexibility. This interface requires a WWW client such as XMosaic that supports image mapping within HTML forms (URL *http://gdbwww.gdb.org/* under Ideogram-based Searching).◊

GDB SprintNet/DataPac Access Ends November 30

Since GDB and OMIM became publicly available, North American users have been able to access the databases via SprintNet/DataPac, usually with a local phone call. Connect charges for SprintNet/DataPac have been paid by the GDB/OMIM project.

Given the limited number of such connections relative to their cost, GDB has been asked to end SprintNet/DataPac service on November 30. Numerous commercial services now offer Internet access to individuals at a reasonable cost.

SprintNet/DataPac users should check with computer support staff at their institution to determine Internet availability. Those without institutional Internet access should contact GDB User Support for a list of service providers in their area and other access methods.◊

GDB USER SUPPORT, REGISTRATION

To become a registered user of GDB and OMIM, contact one of the User Support offices listed below (a user may register to access both Baltimore and a remote node). Questions, problems, or user-registration requests may be sent by telephone, fax, or e-mail. User-registration requests should include name, institutional affiliation, and title (if applicable), street address (no P.O. box numbers), telephone and fax numbers, and e-mail address.

The Help Line in Baltimore is staffed from 9 a.m. to 5 p.m. EST for information on accounts and training courses, technical support, and data questions. Calls received after hours will be forwarded to the appropriate voice mail and returned as soon as possible. To obtain a user's local SprintNet (Telenet) number for locations within the United States: 800/736-1130.

GDB, OMIM Training Schedule

A "GDB/OMIM and Genomic Data on the Internet" class will be held in Baltimore on Nov. 14–15. This course offers thorough coverage of the structure, content, and roles of GDB and OMIM; discusses the strengths and weaknesses of various interfaces for searching the data; and explores related genomic resources available worldwide on the Internet. In addition to using GDB and OMIM application software, participants will learn how to retrieve phenotype, mapping, and sequence data with tools such as ftp, e-mail, Gopher, and the WWW hypertext browser NCSA Mosaic. Contact the U.S. GDB User Support Office.

User Support Offices

UNITED STATES
GDB User Support
Genome Data Base
Johns Hopkins University
2024 E. Monument Street
Baltimore, MD 21205-2100
410/955-9705
Fax: /614-0434
Internet: *help@gdb.org*

AUSTRALIA
Alex Reisner
ANGIS
Electrical Eng. Bldg. J03
University of Sydney
Sydney, N.S.W. 2006
Australia
+ 61/2-692-2948
Fax: -3847
Internet: *reisner@angis.su.oz.au*

FRANCE
Philippe Dessen
Service de Bioinformatique
CNRS-INSERM
7 rue Guy Moquet - BP8
94801 Villejuif Cedex
France
+33/14559-5241
Fax: -5250
Internet: *gdb@genome.vjf.inserm.fr*

GERMANY
Otto Ritter
Molecular Biophysics Dept.
German Cancer Research Center
Im Neuenheimer Feld 280
D-6900 Heidelberg
Germany
+ 49/6221-42-2372
Fax: -2333
Internet: *dok261@cvx12.dkfz-heidelberg.de*

ISRAEL
Jaime Prilusky
Bioinformatics Unit
Weizmann Institute of Science
76100 Rehovot, Israel
+972/8-343456
Fax: -344113
Internet: *lsprilus@weizmann.weizmann.ac.il*

JAPAN
Mika Hirakawa
JICST GDB Center
Numajiri Sangyo Building
783-12, Enokido
Tsukuba City, Ibaraki 305,
Japan
+81/298-38-2965
Fax: -2956
Internet: *mika@gdb.gdbnet.ad.jp*

NETHERLANDS
GDB User Support
CAOS/CAMM Center
Faculty of Science
University of Nijmegen
P.O. Box 9010
6500 GL NIJMEGEN
Netherlands
+ 31/80-653391
Fax: -652977
Internet: *post@caos.caos.kun.nl*

SWEDEN
GDB User Support
Biomedical Center
Box 570
S-751 23 Uppsala
Sweden
+ 46/18-174057
Fax: -524869
Internet: *help@gdb.embnet.se*

UNITED KINGDOM
Administration
HGMP Resource Centre
Hinxton, Cambridge
CB10 1RQ
United Kingdom
+ 44/1223-494511
Fax: -494512
Internet: *admin@hgmp.mrc.ac.uk*

OMIM on WWW, Other Media

Information on clinical phenotypes (descriptions of genetic conditions) can be found in the printed or online version of Victor McKusick's *Mendelian Inheritance in Man* (MIM), 11th edition. OMIM is available from Johns Hopkins University (JHU) by WWW (*http://gdbwww.gdb.org*), Gopher (*gopher.gdb.org*), e-mail (*mailserv@gdb.org*), GDB/Accessor for Macintosh, and IRX via login to JHU. Contact GDB User Support for information. The printed version of MIM can be obtained from bookstores and JHU Press; Hampden Station; Baltimore, MD 21211-2190 (800/537-5487 or 410/516-6956, Fax: -6998). JHU Press also sells MIM on CD-ROM.◊

Genome News

Human Genome news

This newsletter is intended to facilitate communication among genome researchers and to inform persons interested in genome research. Suggestions are invited.

Human Genome Management Information System

Managing Editor
Betty K. Mansfield

Editors/Writers
Anne E. Adamson
Denise K. Casey
Kathleen H. Mavoumin

Production Manager/Editor
Judy M. Wyrick

Production Assistants
K. Alicia Davidson
Larry W. Davis
Sheryl A. Martin
Laura N. Yust

HGMIS Address
Betty K. Mansfield
Oak Ridge National Laboratory
1060 Commerce Park
MS 6480
Oak Ridge,TN 37830
615/576-6669
Fax: /574-9888
Internet: *bkq@ornl.gov*

Sponsor Contacts

Daniel W. Drell
DOE Program Office
Health Effects and Life Sciences Division
Germantown, MD
301/903-6488, Fax: -8521
Internet: *daniel.drell@mailgw.er.doe.gov*

Leslie Fink
NIH National Center for Human Genome Research
Bethesda, MD
301/402-0911, Fax: -4570
Internet: *lsf@cu.nih.gov*

National Center for Human Genome Research

Chromosome 16 Mappers Review Status

The Third International Workshop on Human Chromosome 16 was held May 7–9 at the Pittsburgh Supercomputing Center at Carnegie Mellon University. Workshop goals were to review the status of physical, genetic, and comparative mapping on chromosome 16, construct consensus physical and genetic maps, consolidate data on disease loci, review and deposit data into the Genome Data Base (GDB), and facilitate data sharing and collaborations. Some 29 participants from Australia, Great Britain, Netherlands, and the United States presented 19 abstracts. The workshop was sponsored by DOE, and travel for international participants was funded by the Human Genome Organisation.

Contact:
Norman A. Doggett
LANL Life Sciences Division, MS M88
Los Alamos, NM 87545
505/665-4007, Fax: -3024
Internet: *doggett@gnome.lanl.gov*

Genetic Maps

Helen Kozman [Adelaide Women's and Children's Hospital (AWCH), Australia] and Anne Black [University of Iowa, Cooperative Human Linkage Center (CHLC)] reported the merger of chromosome 16 genetic linkage maps from the CEPH consortium and CHLC into a high-quality consensus linkage map. Loci chosen for the consensus map were highly informative polymorphisms detectable by the polymerase chain reaction (PCR). These polymorphisms included dinucleotide, trinucleotide, and tetranucleotide repeats spaced about 5 to 10 cM apart with odds of an alternative locus placement at 1000:1 or greater. The map extends from the hypervariable locus D16S85 at 16pter to D16S303 at qter. Genetic linkage information on loci generated by the CEPH consortium is available from the CEPH database, and information on loci generated by CHLC can be obtained via ftp (*ftp.chlc.org*).

Comparative Maps

Michael Siciliano and Zuoming Deng (University of Texas M.D. Anderson Cancer Center) summarized all available data on conserved homology, synteny, and linkage of human chromosome 16 with the mouse. The most significant new findings were the identification of regions of (1) conserved linkage involving five loci from p13.12 to p13.3 and mouse chromosome 16 and (2) conserved synteny involving two loci between p12.3 and p12.2 and mouse chromosome 11.

Physical Maps

Three working groups were responsible for summarizing physical mapping data for pter to p13.3, p13.2 to the centromere, and centromere to 16qter. Peter Harris (John Radcliffe Hospital, U.K.) reported on the terminal short-arm band 16p13.3, which contains four genetic-disease loci: *PKD1* (polycystic kidney disease), *TSC2* (tuberous sclerosis), *MEF* (familial Mediterranean fever), and *RSTS* (Rubenstein-Taybi syndrome). Efforts to map and clone these genes have made this one of the best characterized regions of chromosome 16. Isolation of the *TSC2* gene by a positional-cloning approach was reported in December 1993 by the European Chromosome 16 TSC consortium. Positional cloning of the *PKD1* gene was reported soon after the workshop by the European polycystic kidney disease consortium. Harris's working group presented a consensus physical map, assembled at the workshop, that contained 35 ordered markers spanning 9 cytogenetic breakpoints in 16p13.3.

Sara Mole (University College London) summarized physical-mapping data for 16p13.2 to the centromere. This region is divided into 30 intervals by somatic cell hybrids and fragile sites. A yeast artificial chromosome (YAC) contig was constructed across the folate-sensitive fragile site FRA16A at p13.11. This fragile site was later cloned and found to be the expansion of a CCG repeat (Nancarrow et al., 1994). The Batten disease gene *CLN3* location has been substantially refined in p11.2 with new markers that display allelic association. Several candidate genes and sequences identified by exon amplification are being characterized for this region.

David Callen (AWCH) summarized physical-mapping data for the centromere to 16qter, a region divided into 39 intervals by hybrid breakpoints and fragile sites. A number of new genes, expressed sequence tags (ESTs), and DNA markers have been mapped to these intervals and entered in GDB. Bardet-Biedl and Townes-Brocks syndromes have been mapped to the long arm. Bardet-Biedl was reported to be heterogeneous, with relatively few families showing linkage to this chromosome. Genetic mapping is narrowing the search area for this gene; in general, gene density is increased in the 16q22.1 and 16q24.3 bands. Construction of high-density cosmid maps for the 16q24.3 region is progressing as researchers probe high-density cosmid grid filters with *Alu*-PCR products from somatic cell hybrids that contain only restricted regions of this chromosome.

Norman Doggett [Los Alamos National Laboratory (LANL)] reported construction of an integrated physical-genetic-cytogenetic map of human chromosome 16 based on the high-resolution cytogenetic breakpoint map. The physical map consists of both a low-resolution YAC contig map and a high-resolution cosmid contig map. The YAC contig map is composed

of 450 CEPH mega-YACs and 200 flow-sorted, chromosome 16–specific YACs that are anchored to the breakpoint map with ESTs and 300 sequence tagged sites (STSs) from cosmid contigs, genetic markers, and genes. This YAC map provides nearly complete coverage of the euchromatic arms of the chromosome.

LANL has produced a high-resolution, "sequence-ready" map consisting of 4000 fingerprinted cosmids assembled into contigs covering 60% of the chromosome. This map was integrated with the YAC and cytogenetic breakpoint maps by mapping STSs from cosmid contigs and by detecting hybridizations between YACs and cosmids. A highly informative microsatellite-based genetic map of PCR-typable markers was integrated with the cytogenetic and physical map by placing markers on the breakpoint map and screening against the YAC and cosmid maps. All these data were assembled into an integrated map using SIGMA software developed at LANL.

Discussions were held on coordinating efforts to complete the cosmid contig map, develop an EST map, and begin large-scale sequencing. Martijn Breuning (Leiden University) agreed to host the next Chromosome 16 Workshop in November 1995. *[Daniel Drell (DOE), Norman Doggett (LANL), and David Callen (AWCH)]* ◊

DIMACS Special Year 1994–95

Mathematical Support for Molecular Biology

A Special Year on Mathematical Support for Molecular Biology was inaugurated in August by the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS). This consortium of Rutgers and Princeton universities, AT&T Bell Laboratories, and Bellcore designed the Special Year to acquaint discrete mathematicians and theoretical computer scientists with molecular biology problems related to their fields and to foster collaborations and interdisciplinary research.

Joachim Messing and Fred Roberts (Rutgers University) are chairing the Special Year, with cochairs Lawrence Shepp (AT&T Bell Laboratories) and Michael Waterman (University of Southern California). A steering committee composed of leading experts from a number of collaborating institutions and industries is assisting in planning and implementing activities. See current and future *HGN* calendars for selected events (p. 14).

Special Year Activities

- **Workshop series** on such topics as sequence alignment, phylogeny, and HIV sequence analysis. Every workshop has two main organizers, one each from the mathematical and biological sciences.
- **Distinguished Lecture Series** consisting of 11 lectures by outstanding researchers.
- **Seminars** weekly, except when a workshop, miniworkshop, or Distinguished Lecture Series address is scheduled.
- **Miniworkshops** (1-day) on such topics as combinatorial structures in molecular biology, DNA topology and regulation, gross and fine structure of DNA, global minimization of nonconvex energy functions, sequence-based methods for protein folding, antibody sequence and structure, and geometrical methods for conformational modeling.
- **Algorithm Implementation Challenges**, culminating in a September 1995 workshop, to challenge researchers to develop algorithmic methods for a series of benchmark problems dealing with DNA sequence determination from shotgun sequence data.
- **Visitor program** allowing researchers to spend time at the center. Prominent researchers in biocomputing and biomathematics are visiting, some for as long as a year.

DIMACS Center, Rutgers University
P.O. Box 1179
Piscataway, NJ 08855-1179
908/445-5928, Fax: -5932
Internet: *special@dimacs.rutgers.edu*
Gopher: *dimacs.rutgers.edu*
WWW: *http://dimacs.rutgers.edu/*
Telnet: *telnet dimacs.rutgers.edu*
(log in as *info*)

In addition to the many technical reports, journal articles, and conference volumes that usually result from a DIMACS special year, sponsors hope to produce a volume of papers, many of them expository, by leaders in the field. For dates and places of scheduled events and further information, contact DIMACS Center.◊

¶ Publication: Genome Landmarks Chart

"Landmarks of the Human Genome" is an up-to-date chromosome-by-chromosome chart (27 by 39 in.) of genetic research markers, including genes, bordered by an alphabetical index of nearly 1000 health disorders and their chromosomal locations. Third edition, January 1994. Fee charged. [Genome Poster; The Journal of NIH Research; 1444 I Street NW, Ste. 1000; Washington, DC 20005 (202/785-5333, ext. 10, Fax: /872-7738).] ◊

Errors in *HGN*?

Please contact *Human Genome News* staff so we may correct them for our readers. (Fax: 615/574-9888, Internet: *bkq@ornl.gov*). ◊

✔ Correction

Chromosome Editors*

Committee	Editors	Location	Fax	E-Mail
Nomenclature	**PHYLLIS J. McALPINE**	Univ. of Manitoba, CN	204/786-8712	*mcal@genmap.hgen.umanitoba.ca*
	Claude Boucheix	INSERM, U268, FR	+33-14/958-1085	*jasmin@citi2.fr*
	Benjamin Carritt	Univ. Coll. of London	+44-71/387-3496	*b.carritt@crc.ac.uk*
	Margaret A. Pericak-Vance	Duke Univ. Med. Ctr.	919/684-6514	*mpv@dnadoc.mc.duke.edu*
	Sue Povey	Univ. Coll. of London	+44-71/387-3496	*m.povey@hgmp.mrc.ac.uk*
	Thomas B. Shows	Roswell Park Cancer Inst.	716/845-8449	*tbs@shows.med.buffalo.edu*

*Chromosome editors are appointed by the Human Genome Organisation to review information submitted for inclusion in the Genome Data Base. The names above were inadvertently omitted from the listing in the July *HGN* (pp. 8–9).◊

Calendar of Genome-Related Events* (acronym list, p. 16)

October

18–22. 44th ASHG Annu. Meet.; Montreal [M. Ryan, 301/571-1825, Fax: /530-7079]

20. DIMACS Spec. Yr.: Distinguished Lect. Ser.—Samuel Karlin (also offered Oct. 21) [For specific location, contact M. Farach, 908/445-4580, Fax: -5932, *special@dimac.rutgers.edu*]

20–23. 8th Annu. N. Am. CF Conf.; Orlando, FL [CFF, C. McPherson, 301/951-4422, Fax: -6378]

24–28. 3rd Intl. Conf. on Nanometer-Scale Sci. and Technol. in conj. with 41st AVS Natl. Symp.; Denver [C. Marrian, 202/767-3150, Fax: -4998]

25. Stephen Fodor: DNA Sequence Anal. Using Oligonucleotide Arrays; Bethesda, MD [NCHGR Lect. Ser., E. Feingold, 301/496-7531, Fax: /480-2770, *fey@cu.nih.gov*]

28–31. ASBMB Fall Symposia: Oligonucleotide Sel. and Mol. Diversity; Lake Tahoe, CA [ASBMB, 301/530-7010, Fax: -7014]

31. J. Craig Venter: DNA, Genet., and Biotechnol.; Gaithersburg, MD [TIGR/NIST Distinguished Speaker Ser., D. Hawkins, 301/869-9056, Fax: -9423]

November

3. DIMACS Spec. Yr.: Distinguished Lect. Ser.—Temple Smith; contact for specific location [see contact: Oct. 20]

4. **DIMACS Spec. Yr.: Combinatorial Struct. in Mol. Biol. Mini-Workshop; New Brunswick, NJ [see contact: Oct. 20]

4–6. Nordic Genome Initiative Workshop: Cloning of Large DNA Fragments; Oslo [H. Prydz, +47-22/958-754, Fax: /694-130]

4–8. 3rd Intl. *E. coli* Genome Meet.; Woods Hole, MA [MBL, M. Riley, 508/540-6055, Fax: /540-0155, *confserv@mbl.edu*]

5–9. Symp. on Comput. Appl. in Med. Care; Washington, DC [AMIA, 301/657-1291, Fax: -1296, *amia@camis. stanford.edu*]

6–10. 2nd S.-N. Hum. Genome Conf.; Beijing, CH [S. Matsui, Fax: +33-1/4567-2639 or Z. Chen, Fax: +86-1/250-1844]

6–10. 8th Intl. Mouse Genome Conf.; London [S. Brown, +44-71/723-1252 ext. 5848, Fax: /706-3272, *s.brown@sm.ic.ac.uk*]

7–9. Impact of Nucleic Acid-Based Technol.: Revolution in Clinical Diagnosis, Appl., & Res.; Amsterdam [CHI, B. Keddy, 617/487-7989, Fax: -7937]

9–11. 5th Intl. Workshop on Chromosome 21; Tsukuba-city, JP [N. Shimizu, Tel/Fax: +81-3/3351-2370, *shimizu@dmb.med.keio.ac.jp*]

10–12. **DIMACS Spec. Yr.: Sequence Alignment Workshop; Princeton, NJ [see contact: Oct. 20]

13–17. **4th DOE Genome Contractor-Grantee Workshop; Santa Fe, NM [S. Spengler, 510/486-4879, Fax: -5717, *sylviaj@ ux5.lbl.gov*]

14–15. Gene Therapy; Washington, DC [IBC Conf., 508/481-6400, Fax: /481-7911]

14–16. Computational Approaches in the Anal. and Eng. of Proteins; Madrid [CIMB, A. González, +34-1/435-4240, Fax: /576-3420]

14–18. Supercomput. 94: Conf. on High Perf. Comput. and Commun.; Washington, DC (poster deadline: Aug. 1) [Supercomput. 94, 515/294-0673, Fax: -0888, *info@sc94.ameslab.gov*]

17. Eric Green: Towards a Complete Physical Map of Hum. Chromosome 7—A Front-Line View of the HGP; Bethesda, MD [see contact: Oct. 25]

17–19. Genet. Revolution; San Diego [AACC, 202/857-0717, Fax: /833-4576]

17–20. 1994 Miami Bio/Technol. Eur. Symp. on Adv. in Gene Technol.: Mol. Biol. and Hum. Genet. Dis.; Monaco [N. Forrest, +44-71/386-6633, Fax: /379-5417 or H. Jackson, /872-0104, Fax: /240-2408]

18–19. Preparing Schools for the Genet. Revolution; Lincoln, NE [Ctr. on Children, Families, and Law, 402/472-3479, Fax: -8412, *gwright@unlinfo.unl.edu*]

18–21. 2nd Meet. of the Eur. Working Group on Hum. Gene Transfer and Therapy Conf.; London [P. McIntyre, +44-42/352-8494, Fax: /350-0685]

20–22. Intl. Cong. on Genomic Imprinting; Florence, IT [M. Uzielli, +39-55/566-2942, Fax: -2916]

28. W. French Anderson: DNA, Genet., and Biotechnol.; Gaithersburg, MD [see contact: Oct. 31]

28–30. Intl. Symp. on Gene Therapy; Valencia, SP [FVEA, S. Grisolia, +34-6/392-0604, Fax: /391-1549]

29–30. Technol. Adv. for Gene Therapy; CHI, Washington, DC (poster deadline: Sept. 30) [see contact: Nov. 7–9]

29–Dec. 4. Translational Res. in Cancer: New Opportunities for Progress; Asheville, NC [AACR, 215/440-9300, Fax: -9313]

December

9. **DIMACS Spec. Yr.: DNA Topology and Struct. Mini-Workshop; New Brunswick, NJ [see contact: Oct. 20]

12–13. Rational Drug Design: Struct.-Based Design & Dev.; IBC, San Diego [see contact: Nov. 14–15]

14–16. Combinatorial Libraries for Mol. Diversity; IBC, San Diego [see contact: Nov. 14–15]

15. Rebecca Eisenberg: Role of Patents in Technol. Transfer; Bethesda, MD [see contact: Oct. 25]

15–16. 5th Generation Comput. Syst. Workshop: Fusion of Mol. Biol. and Knowledge Processing; Tokyo (abs. deadline: Oct. 20) [H. Tanaka, +81-3/3456-3192, Fax: -1618, *htanaka@icot.or.jp*]

19–20. Genome Informatics Workshop 1994; Nish-ku, JP (abs. deadline: Oct. 1) [T. Takagi, +81-3/5449-5614, Fax: -5434, *workshop @ims.u-tokyo.ac.jp*]

January 1995

3–6. Biotechnol. Comput. Track HICSS-28; Maui, HI (paper deadline: June 15) [L. Hunter, 301/496-9300, Fax: -0673, *hunter@ work.nlm.nih.gov* or T. Takagi, at Dec. 19–20 contact above]

5–11. Oncogenes: 20 Yrs. Later; Keystone, CO (reg. deadline: Sept. 7) [Keystone Symp., 303/262-1230, Fax: -1525]

9–12. BioEast '95; Washington, DC [BioConf. Intl., W. Small, 301/652-3072, Fax: -4951]

15–19. Plant Genome III; San Diego (abs. deadline: Nov. 1) (PG-I and PG-II abs. available on WWW, *http://probe.nalusda.gov: 8000/plant/index.html)* [Scherago Intl., 212/643-1750, Fax: -1758, *scherago@biotechnet.com*] ◊

An extended list of genome events is available from HGMIS. See p. 12 for contact information.

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

For Your Information

Training Calendar*

October

31–Nov. 2. Charge-Coupled Devices, Camera, and Appl.; Los Angeles [UCLA Short Courses, 310/825-1047, Fax: /206-2815]

31–Nov. 4. PCR Methodol.; Columbia, MD [Exon-Intron, 800/407-6546, Fax: 410/730-3983]

31–Nov. 4. Recombinant DNA: Tech. & Appl.; Rockville, MD [ATCC, 301/231-5566, Fax: /770-1805]

November

1–14. Mol. Genet., Cell Biol., & Cell Cycle of Fission Yeast; Cold Spring Harbor, NY (appl. deadline: July 15) [CSHL, 516/367-8345, Fax: -8845, *meetings@cshl.org*]

2–7. Computational Genomics; CSHL (appl. deadline: July 15) [see contact: Nov. 1–14 above]

4–5. DNA Databanks & Repositories; St. Paul [AFIP/ARP, 301/427-5231, Fax: -5001, *lowther@email.afip.osd.mil*]

7–11. In Situ Hybridization & rDNA Technol.; Exon-Intron, Inc., Columbia, MD [see contact: Oct. 31–Nov. 4]

7–11. PCR Tech.; Germantown, MD [LTI, 800/952-9166, Fax: 301/258-8212]

7–11. Photometry and Colorimetry in Electronic Imagery and Industry; UCLA Short Courses, Los Angeles [see contact: Oct. 31–Nov. 2]

8–11. PCR Appl./Cycle DNA Sequencing; ATCC, Rockville, MD [see contact: Oct. 31–Nov. 4]

8–21. Mol. Markers for Plant Genet. & Plant Breeding; CSHL (appl. deadline: July 15) [see contact: Nov. 1–14]

10–11. DNA Sequencing Without Radioact.; West Haven, CT [BTP, S. Chance, 800/821-4861, Fax: 603/267-1993]

10–11. Metaphase & Interphase Chromosomes; Gaithersburg, MD [Oncor, Inc., 800/556-6267, Fax: 301/926-6129]

14. Intro. to PCR; BTP, West Haven, CT [see contact: Nov. 10–11]

14–15. GDB/OMIM & Genomic Data on the Internet; Baltimore [GDB User Support, 410/955-9705, Fax: /614-0434, *help@gdb.org*, see p.11]

14–18. Recombinant Baculovirus Tech.; LTI, Germantown, MD [see contact: Nov. 7–11]

15–16. Quantitative RNA-PCR; BTP, West Haven, CT [see contact: Nov. 10–11]

17. Rapid Hybridization of Metaphase & Interphase Chromosomes; Oncor, Inc., Montreal (also offered Nov. 18) [see contact: Nov. 10–11]

17–18. Basic Cloning & Hybridization Tech.; BTP, West Haven, CT [see contact: Nov. 10–11]

28–Dec. 2. Cell Culture Tech.; LTI, Germantown, MD [see contact: Nov. 7–11]

December

12–16. Recombinant DNA Tech. I; LTI, Germantown, MD [see contact: Nov. 7–11]

January 1995

9–13. **Adv. Linkage Course; New York (appl. deadline: Nov. 10) [K. Montague, 212/960-2507, Fax: /568-2750, *jurg.ott@columbia.edu*] ◊

U.S. Genome Research Funding Guidelines

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

NIH National Center for Human Genome Research (NCHGR)

Application receipt dates:

- R01, P01, R21, R29, P30, P50, K01,* and R13 grants – February 1, June 1, and October 1.
- Individual postdoctoral fellowships – April 5, August 5, and December 5.
- Institutional training grants – January 10, May 10, and September 10.
- Small Business Innovation Research Grants (SBIR: firms with 500 or fewer employees) – April 15, August 15, and December 15.
- Research supplements for underrepresented minorities – applications are accepted on a continuing basis.
- Requests for Applications (RFAs) – receipt dates are independent of the above dates. Notices will appear in *HGN* and other publications.

*Expedited review possible. Check with NCHGR during application development phases.

Program announcements are listed in the weekly *NIH Guide for Grants and Contracts*,* which is available electronically through one of the following methods.

- Gopher (*gopher.nih.gov*).
- Institutional Hubs. A designee receives automatic updates and distributes them locally to researchers. Send a message naming the responsible person to BITNET: *q2c@nihcu* or Internet: *q2c@cu.nih.gov*.
- NIH Grant Line (also known as DRGLINE): Electronic bulletin board updated weekly. Connection is through a modem (301/402-2221), and files can be transmitted rapidly via BITNET or Internet. The Grant Line is also accessible by Telnet to *wylbur.cu.nih.gov*. When connection is open, type *VT100*. At the INITIALS prompt, type *BB5* and at the ACCOUNT prompt, type *CCS2*. For more information, contact John James (301/594-7270, Fax: -7384).

Full text of RFAs listed in the NIH grants guide may also be obtained from NIH NCHGR in Bethesda, Maryland (301/496-0844).

DOE Human Genome Program

For funding information or general inquiries, contact the program office via

- 301/903-6488 or Internet (*genome@er.doe.gov*). Relevant documents are available by ftp (*oerhp01.er.doe.gov* in directory */genome*).

SBIR Grants

DOE and NIH invite small business firms to submit grant applications addressing the human genome topic of SBIR programs, which are designed to strengthen innovative firms in research and development and contribute to the growth and strength of the nation's economy. For more information on human genome SBIR grants, contact

- Kay Etzler; c/o SBIR Program Manager, ER-16; DOE; Washington, DC 20585 (301/903-5867, Fax: -5488).
- Bettie Graham; Bldg. 38A, Rm. 610; NIH; 9000 Rockville Pike; Bethesda, MD 20892 (301/496-7531, Fax: /480-2770).

National SBIR conferences: San Jose, CA (November 14–16); Chicago, IL (April 26–28, 1995). Conference Hotline: 407/791-0720.◊

Fellowships Offered by University of Iowa Program in Biomedical Ethics

The University of Iowa Program in Biomedical Ethics invites applications for its visiting fellowships in molecular and clinical genetics. This project is part of the ethical, legal, and social implications (ELSI) core of the Cooperative Human Linkage Center, one of the genome centers funded by the NIH National Center for Human Genome Research. Work in laboratory and clinical settings is central to the 2- to 4-month fellowships, which include a monthly stipend of $3500. The fellowships are intended for philosophers, historians, attorneys, journalists, nurses, and other professionals who are not biological scientists but have demonstrated a strong interest in the ELSI aspects of human genetics. Application deadline: December 30. [Inquiries and requests for applications: Jay Horton; Program in Biomedical Ethics; University of Iowa; 1-112 MEB; Iowa City, IA 52242-1000 (319/335-9631, Fax: -8515, Internet: *jay-horton@uiowa.edu*).]◊

This newsletter is prepared at the request of the DOE Office of Health and Environmental Research and the NIH National Center for Human Genome Research by the Biomedical and Environmental Information Analysis Section of the Health Sciences Research Division at Oak Ridge National Laboratory, which is managed by Martin Marietta Energy Systems, Inc., for the U.S. Department of Energy, under Contract DE-AC05-84OR21400.◊

Human Genome Management Information System Subscription/Document Request* (Vol. 6, 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 _____

1. ___ *Human Genome News* ___New Subscriber ___Change of Name/Affiliation/Address (circle all that apply) ___Drop Subscription
2. ___ Reprint of "A New Five-Year Plan for the U.S. Human Genome Project" *(Science*, October 1, 1993) by Francis Collins and David Galas
3. ___ DOE *Human Genome 1993 Program Report* ___DOE *Primer on Molecular Genetics*
4. ___ *Meeting Report: DOE Informatics Summit*—DRAFT (April 26–27, 1993, Baltimore, Maryland)

*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. 12.

SELECTED ACRONYMS

AACC Am. Assoc. for Clin. Chem.
AACR Am. Assoc. for Cancer Res.
AMIA Am. Med. Informatics Assoc.
AFIP/ARP Armed Forces Inst. of Pathol./Am. Registry of Pathol.
ASHG Am. Soc. of Hum. Genet.
ASBMB Am. Soc. for Biochem. & Mol. Biol.
ATCC Am. *Type Culture* Collection
AVS Am. Vaccum Soc.
BTP Biotechnol. Train. Programs
CEPH Centre d'Etude du Polymorphisme Humain
CFF Cystic Fibrosis Found.
CHI Cambridge Healthtech Inst.
CIMB Ctr. for Intl. Meet. on Biol.
CSHL Cold Spring Harbor Lab.
DIMACS Discrete Math. & Comp. Sci.
GDB/OMIM Genome Data Base/Online *Mendelian Inheritance in Man*
HGP Hum. Genome Proj.
HICSS Hawaii Intl. Conf. on Syst. Sci.
IBC Intl. Bus. Comm.
LTI Life Technologies, Inc.
MBL Marine Biological Lab.
NCHGR Natl. Ctr. for Human Genome Res.
TIGR/NIST The Inst. for Genomic Res./Natl. Inst. of Standards and Technol.
UCLA Univ. of Cal. Los Angeles
WWW World Wide Web

OAK RIDGE NATIONAL LABORATORY • MANAGED BY MARTIN MARIETTA ENERGY SYSTEMS, INC. • 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