Human Genome news

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

ISSN: 1050-6101

Vol. 4, No. 4, November 1992

Researchers Report Mapping Progress

Data Published on Chromosomes Y and 21, Leukemia Gene, Genetic Linkage Map

Mapping Techniques Put 5-Year Goals in Sight

International progress in physical and genetic mapping has recently been reported as the U.S. Human Genome Project moves into its third year:

- Physical maps covering the entire functional portions of human chromosomes Y and 21 have been completed. Yeast artificial chromosome (YAC) clones now cover 98% of the Y chromosome euchromatic region. The YAC contig for chromosome 21 spans the pericentromeric through subtelomeric loci of 21q and is estimated to contain 42 Mb.
- A gene that appears to be associated with some forms of childhood leukemia has been identified during construction of a physical map of human chromosome 11. Findings suggest that chromosomal breakage during translocation disrupts the function of the gene, whose product may be involved in regulating other genes active in white blood cells.
- The most comprehensive and up-to-date genetic linkage map of the human genome has been published. The new map contains 1676 markers spaced an average of about 5 cM apart, covering 90% of the genome.

■ Y CHROMOSOME

In two articles appearing in the October 2 issue of *Science*, David Page and his colleagues at the Whitehead Institute and Massachusetts Institute of Technology presented their physical map of the entire functional (euchromatic) portion of the human Y or "male" chromosome, which at 60 Mb is one of the smallest. The new map, consisting of 196 overlapping YAC clones, is expected to facilitate positional cloning of genes for a given phenotype, speed gene identification, and provide material for large-scale sequencing. The work was supported by the NIH National Center for Human Genome Research (NCHGR) to further the interim (5-year) Human Genome Project goal of constructing physical maps of all human chromosomes with markers spaced at 100-kb intervals (see box, p. 4).

The euchromatic region includes the short arm, centromere, and proximal long arm. It contains all known genes on the

In This Issue. . .

Genome News

STS-Content Mapping Strategy Used For Y and 21 Physical Maps:

- **STS probe sets assembled and ordered.**
- **Total human YAC library screened for STS landmarks.**
- **YAC clones ordered by determining STS overlaps.**

Y chromosome and appears relatively constant in size. By contrast, the heterochromatic region has no known function; consists of short, repetitive sequences; and varies tremendously among males (undetectable in some and twice the size of the euchromatic region in others).

Because most of the Y chromosome does not undergo meiotic recombination, a genetic linkage map is precluded, and Y-linked gene identification has been based on physical mapping of naturally occurring deletions. Deletion mapping, in which DNA loci are ordered along the chromosome, is a practical approach because Y chromosome deletions occur fairly frequently in the population.

STS-Based Deletion Map

In the first of two *Science* articles,[1] Page and colleagues described deletion mapping based on sequence tagged site (STS) detection; an STS is a short DNA sequence that can be detected by the polymerase chain reaction (PCR)—a fast, sensitive assay. An STS can be mapped to a specific point that becomes a chromosomal landmark easily disseminated through publication in journals and computer databases.

The researchers assembled a set of STS probes to screen DNA from 96 patients who had either visible Y chromosome abnormalities or sex chromosome constitutions (XX males and XY females) that did not correlate with their outward sexual appearance. To establish the relative STS order, investigators compared sets of positive points (indicating the presence of an STS) and negative points for each patient.

The deletion map is expected to be useful in identifying Y chromosomal genes, studying the origin of chromosomal disorders, and tracing the evolution of the Y chromosome.

Physical (Ordered Clone) Map

In their second *Science* article,[2] the researchers describe how ordered STS probes, generated during deletion mapping, provided the framework for rapid construction of the chromosomal physical map. Investigators began by preparing a library of YAC clones from the genomic DNA of a human XYYYY male to obtain a Y chromosome representation fourfold greater than could have been achieved with a comparably sized library for an XY male.

Use of a total human YAC library eliminated the labor-intensive step of constructing chromosome-specific YAC libraries. The YAC library, containing 10,368 clones with an average insert size of 650 kb, was then analyzed by "STS-content mapping" and screened for landmarks using 160 deletion-mapped STSs. YAC clones were ordered by determining overlap (common STSs). Based on their presence or absence in the YACs, 207 Y-DNA probes were assigned to 127 ordered intervals with an average spacing of about 220 kb. The authors estimate that the physical map spans at least 28 Mb; previous estimates based on cytological observations were in the 30- to 40-Mb range.

This mapping strategy enabled investigators to isolate specific YACs from a complex library, order and arrange them in larger units to form "contigs," and place the STSs along the chromosome. These ordered points on the physical map will provide landmarks for all future investigations of the Y chromosome, the authors stated. They also noted that this method can be applied to other human chromosomes if large numbers of STSs can be generated for them.

■ CHROMOSOME 21 MAP

In the October 1 issue of *Nature*,[3] French researcher Daniel Cohen of Genethon and 35 others from 12 institutions [including American researchers Stylianos Antonarakis (Johns Hopkins School of Medicine); David Cox (University of California, San Francisco); Harold Riethman (Wistar Institute); and Katheleen Gardiner and David Patterson (Eleanor Roosevelt Institute)] report a complete physical map of the long arm of human chromosome 21. The smallest and one of the best studied of the human chromosomes, chromosome 21 has been associated with several genetic diseases, including Down's syndrome, some forms of Alzheimer's disease, amyotrophic lateral sclerosis (Lou Gehrig's disease), and progressive myoclonus epilepsy. The new physical clone map will speed identification of the chromosomal regions involved in these and other diseases after the regions are further localized by genetic linkage or cytogenetic analysis.

The map, containing 191 STSs spaced at about 220 kb, was constructed from human YAC libraries having insert sizes of 400 kb to 2 Mb. The new map contains an average of five clones in each interval between two

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 and Safety 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.◊

STSs, an indication of the "robustness" of the contig.

The work was conducted at laboratories in France, the United States, Spain, and Japan. It was supported by the French Ministry of Research and Technology and the Association Francaise Contre les Myopathies through the Genethon program and by grants from NCHGR and other NIH institutes.

■ CHILDHOOD LEUKEMIA GENE

Researchers constructing a human chromosome 11 physical map consisting of ordered DNA clones have determined the base-pair sequence of fragments spanning the chromosomal region 11q23. The area is frequently broken in patients with some forms of childhood leukemia; this breakage may affect a gene that normally plays a role in regulating cell division in white blood cells. Leukemias arise when immature white blood cells become abnormal and grow and divide rapidly.

This work, supported by NCHGR and DOE, was reported in the October issue of *Nature Genetics*[4] by Glen Evans and his coworkers at the Salk Institute in La Jolla, California, and Mark Bower and Bryan Young at the Imperial Cancer Research Fund in the United Kingdom. Evans and his collaborators studied chromosomes of patients with acute leukemias, which account for almost all leukemia cases among children and young adults.

Investigators compared the sequence of the 11q23 fragment to others stored in computer databases and found it to be similar to that of the fruit fly gene *trithorax* (*trx*), which instructs cells to produce a protein having a "zinc finger" region (so called because it loops around a molecule of zinc). The *trx* protein regulates the transcription of developmental genes in the fruit fly.

The findings suggest that chromosome 11 breakage during translocation disrupts the function of the human *trx*-like gene and makes it unable to produce a normal protein.

■ GENETIC LINKAGE MAP

The genetic linkage map of the entire human genome, which appears in the October 2 issue of *Science*,[5] results from the efforts of researchers at more than 70 laboratories worldwide; the work was sponsored largely by NCHGR and the Centre d'Etude du Polymorphisme Humain. Genetic mapping data were assembled into a single publication with chromosomes represented in a common format. Collated by Washington University (St. Louis) geneticist Helen Donis-Keller and her staff, the linkage map contains restriction fragment length polymorphism (RFLP) and microsatellite DNA markers.

Markers Used for Genetic Mapping

According to the *Science* article, the major difference between a previously published genome map and the new one lies in the number and type of markers used. This map incorporates over 4 times more markers, some 300 of which are variable repeats of short (2 to 4 bp) DNA units called microsatellites.

The first human genetic linkage maps were based on protein polymorphisms (assayed by a variety of methods including serologic testing, gel electrophoresis, and enzymatic assay). However, because such markers

➡

Ninth NIH Genome Center Established at University of Iowa

The University of Iowa (UI) Department of Pediatrics has been awarded a grant to become the ninth genome center funded by the NIH National Center for Human Genome Research. Jeffrey Murray (UI) will direct the 4-year, $15-million project. Principal goals are to (1) generate a high-resolution, microsatellite-based genetic linkage map of the entire human genome that will help locate disease-causing genes; (2) address the ethical, legal, and social issues (ELSI) raised by genetic research; and (3) train high school science teachers in genetics and the genome project.

The project involves research teams headed by Val Sheffield (UI); Ken Buetow (Fox Chase Cancer Center, Philadelphia); James Weber (Marshfield Medical Research Foundation, Marshfield, Wisconsin); and Geoffrey Duyk (Harvard University). Sheffield will study DNA variation in markers with an emphasis on cDNAs. Murray's team and the Marshfield investigators will characterize variations in inheritance patterns in a group of about 20 reference families. Harvard researchers will develop microsatellite DNA markers, and the Fox Chase Cancer Center team will use data collected by the other three groups to construct genetic maps and a database of information that can be used by researchers around the world.

Robert Weir and James Hanson (both at UI) will study ELSI concerns, including job discrimination and access to health care insurance for people with genetic diseases, quality of procedures for genetic testing, and confidentiality of test results. The project includes funds for U.S. high school science teachers to attend summer workshops at UI for laboratory work and lectures.

The other NCHGR-funded genome centers are at the Baylor College of Medicine (Houston); University of Michigan (Ann Arbor); Children's Hospital (Philadelphia); University of Utah (Salt Lake City); Whitehead Institute for Biomedical Research (Cambridge, Massachusetts); University of California, San Francisco; University of California, Berkeley; and Washington University (St. Louis).◊

Genome News

Microsatellites Now the Dominant Marker Type for Genetic Mappers

were scarce, genetic mapping lagged before the discovery that restriction enzymes could be used to reveal DNA sequence variation among individuals (polymorphisms) and that the amount of variation was sufficient for construction of a rudimentary linkage map spanning much of the genome.

Although RFLPs have contributed greatly to advances in genetic mapping, most are not usefully informative in the small-pedigree resources typically available for disease-gene mapping. Furthermore, the best RFLPs, variable number tandem repeats (VNTRs), have limited value in disease-gene hunts because they cluster near the ends of chromosomes. RFLPs also suffer the disadvantage of assay by Southern hybridization, now considered a tedious technology that could not be scaled efficiently to high-production levels.

More recently, researchers discovered that microsatellites are abundant, with an estimated 500,000 well distributed across the human genome. Microsatellites are particularly useful for developing markers at gene sequences because a 1-Mb gene-containing segment may have several of these repeat elements along its length. In contrast, testing of candidate genes in linkage studies is severely limited because of the rarity of RFLPs near or within gene sequences.

In addition, microsatellites can be assayed by PCR, making them easily accessible from sequences published or posted in databases and quickly available to the scientific community. Collection and distribution of probes from public repositories often led to long delays, sometimes up to years. Because of their advantages, microsatellites have already become the dominant marker type. The next objective is a genome map of markers that are informative in at least 70% of the population and spaced at intervals of about 10 cM; this map is expected to consist entirely of microsatellites. The original goal of constructing a genetic linkage map with markers spaced an average of 2 to 5 cM apart now appears easily surpassable.

Connecting Genetic and Physical Maps

Because microsatellite markers are STSs, they can be used as anchors to connect the genetic linkage maps with physical maps of DNA clones. Parent clones from which the microsatellites derive will serve as probes for cytogenetic mapping. With the addition of more microsatellite markers to the genetic map, information from linkage and physical maps will converge, the *Science* article states. "It is anticipated that full integration of genetic, cytogenetic, and physical mapping information will be possible, thereby providing a new view of the [human] genome upon which to base future biological studies," the authors conclude.

Physical Maps

A physical map represents actual locations (as compared with relative locations on a genetic map) of identifiable landmarks such as STSs, restriction enzyme cutting sites, DNA repeats, and genes. Distance is measured in base pairs. Different types of physical maps vary in their resolution (level of detail), the coarsest being a map of the banding patterns on stained human chromosomes visible through a light microscope (a cytogenetic map). The contig map provides more detail and depicts the order of overlapping DNA fragments. The new maps for chromosomes Y and 21 described in the accompanying article are contig maps that contain overlapping clones spanning entire chromosomal regions.

Complete physical maps spanning the entire genome will facilitate the correlation of genetic and cytogenetic maps with the underlying DNA, offer access to clones from any region of interest, and provide material for large-scale sequencing. Physical maps for the genomes of several model organisms (such as the bacterium *Escherichia coli*, the yeast *Saccharomyces cerevisiae*, and the nematode *Caenorhabditis elegans)* have been completed or are nearly complete.

Genetic Maps

By correlating the inheritance of DNA markers with the appearance of biological traits in large numbers of related people, scientists can identify the chromosome on which a gene resides. Markers — DNA properties that differ among individuals and are easily identified in the laboratory — serve as signposts along a chromosome and can be genes or segments that have no known coding function but whose inheritance pattern can be followed. Such markers include the DNA microsatellite repeats described in the accompanying article.

Genetic linkage maps, which represent the relative chromosomal order and spacing of DNA markers according to their inheritance patterns, are made by determining how frequently two markers are passed together from parent to child. Closely linked markers are less likely to be separated during normal processes in gamete formation. Distances are measured in centimorgans (cM); two markers are said to be 1 cM apart if they are separated by recombination 1% of the time in humans. A genetic distance of 1 cM is roughly equal to a physical distance of 1 million base pairs (1 Mb).◊

Footnotes

1. "The Human Y Chromosome: A 43-Interval Map Based on Naturally Occurring Deletions." *Science* **258**, 52–59 (October 2, 1992).

2. "The Human Y Chromosome: Overlapping DNA Clones Spanning the Euchromatic Region." *Science* **258**, 60–66 (October 2, 1992).

3. "A Continuum of Overlapping Clones Spanning the Entire Human Chromosome 21q." *Nature* **359**, 380–86 (October 1, 1992).

4. "A Trithorax-Like Gene is Interrupted by Chromosome 11q23 Translocations in Acute Leukaemias." *Nature Genetics* **2**, 113–18 (October 1992).

5. "A Comprehensive Genetic Linkage Map of the Human Genome." *Science* **258**, 67–86 (October 2, 1992).

For additional background information, see also

"A Genetic Linkage Map of the Human Genome," *Cell* **51**, 319–37 (October 23, 1987).◊

Genome News

NIH-DOE Joint Working Group on the Mouse

The third meeting of the NIH-DOE Joint Working Group on the Mouse was held March 5–6 in Washington, D.C., to focus on physical mapping tools and strategies. Attendees included working group members, invited speakers, and federal government representatives. Following are some meeting highlights.

P1 System for Cloning DNA

Nat Sternberg (Du Pont-Merck Pharmaceuticals) reviewed progress in producing mammalian genomic libraries in the P1 cloning system. The average insert in the mouse library is 70 kb, with about 80 to 90% of mouse gene sequences represented. Polymerase chain reaction (PCR) techniques can be used to isolate any one clone from the library in a week or two. The library, which is now available commercially, is arranged in 300 pools of 400 to 500 clones each. Groups of 10 to 50 pools are screened to identify those to be tested further. In the next 6 months Sternberg anticipates expanding the P1 mouse library threefold to sixfold.

Sternberg pointed out the advantages of P1 clones over cosmid clones and yeast artificial chromosomes (YACs) for mapping. He has included in the newest vector (1) rare restriction sites to size the insert and (2) flanking T7 and Sp6 promoters to generate labeled RNA probes (riboprobes) from the insert ends. With cycle sequencing techniques, he can use 200 to 300 bp of sequence from the insert ends to construct a sequence tagged site (STS). Sternberg also described efforts to develop transposon technology for the orderly dissection of cloned DNA and its use in transforming mammalian cells.

ICRF Physical Mapping Project

To make the Imperial Cancer Research Fund (ICRF) mouse and human YAC libraries more widely available, Hans Lehrach (ICRF) has for 2 years been testing the feasibility of transferring filter screening technology to laboratories in the United States and Europe. He plans to include in his arrays the Rad 52 YAC and P1 libraries developed by Steve Brown at St. Mary's Hospital in London. If the transfer is successful, this system can combine information originating in the limited number of mapping laboratories with the large amounts of data generated in laboratories working on mouse genetics or embryology.

Lehrach raised the possibility of mapping the genomes of man, mouse, and perhaps other mammals as a single project. He suggested extending the reference library concept to create a biological information net based on common YAC, P1, cosmid, cDNA, and exon trap clone libraries from many different experiments and laboratories. Such a system would allow the efficient accumulation of many types of biological information, reduce unintended overlaps, and simplify the establishment of pertinent relational databases. ➠

The P1 mouse library manuscript has been published [*Mammalian Genome* 3, 550–58 (1992).]

■ *Education Meetings*

High School Teachers Attend San Francisco Forum

"Winding Your Way Through DNA," a symposium and public forum on molecular biology and its applications, was conducted for San Francisco high school teachers and students on September 19–20. Eminent scientists and biotechnology experts spoke on various aspects of DNA technology, followed by questions from a panel of science writers. The meeting was sponsored by the Exploratorium and the School of Medicine of the University of California, San Francisco (UCSF), and organized by Harold E. Varmus (UCSF).

A capacity audience of 1200 attended the sessions, and another 3000 viewed transmissions by satellite to 27 sites around the country. A set of six edited videotapes of the symposium is available in VHS, PAL, and NTSC formats from Cold Spring Harbor Laboratory Press; 10 Skyline Drive; Plainview, NY 11803-9729 (Continental United States and Canada: 800/843-4388; other locations: 516/349-1930, Fax: -1946.)◊

Upcoming Workshops (see also p. 15)

DOE Sponsoring Kansas City Workshop for Science Teachers

A week-long DOE-funded workshop for secondary science teachers will be held in Kansas City in late June 1993 to explore ethical, legal, social, and technical issues of the Human Genome Project. Topics will include basic human genetics; population screening; biotechnology; linkage analysis; non-Mendelian inheritance; careers in human genetics; science fair projects; genetic resources; and classroom curricular materials. Attendees will participate in hands-on laboratory work and discussions with experts in law, ethics, and medical genetics as well as with panels of individuals with inherited conditions. Funding is available for travel, meals, lodging, stipend, reference books, student materials, and teaching resources that include the new DOE Biological Sciences Curriculum Study module on the Human Genome Project (see p. 9). Grant participation requires a commitment for two consecutive summers. Dates for 1994 and 1995 sessions will be announced later.

For an application, contact Debra Collins; Genetics Education; University of Kansas Medical Center; 3901 Rainbow Boulevard, 4023 Wescoe Bldg.; Kansas City, KS 66160-7318 (Fax: 913/588-3995, BITNET: *collins@ukanvm*).◊

Utah Plans Mapping, Sequencing Workshop

A 2-week workshop for graduate students and postdoctoral fellows on state-of-the-art gene mapping and sequencing skills will be held June 13–26, 1993, at the University of Utah. In addition to providing an overall perspective, the workshop will involve hands-on experience in a variety of sequencing and mapping techniques. Medical applications of genome technology will be discussed thoroughly, and students will use genetic markers and linkage analysis to study a specific disease gene. Faculty and guest lecturers will include well-known genome researchers from the University of Utah, other universities, and national laboratories.

The fee includes tuition, most meals, activities, and hotel housing ($1160 double occupancy; $500 additional for private room). Applications for early acceptance due March 1, 1993. For a descriptive brochure and application, contact Genome Technology Workshop; Eccles Institute of Human Genetics; University of Utah; Bldg. 533, Room 2160; Salt Lake City, UT 84112 (801/585-5606, Fax: 581-7796, Internet: *gtw@corona.med.utah.edu*).◊

Genome News

Mouse Reference Tool Offered

***Encyclopedia of the Mouse Genome II* (1992), a companion reference tool to the journal *Mammalian Genome*, is now available. 294 pp. Free with journal subscription; $49 separately. (Springer-Verlag New York, Inc.; Attn: Dean Smith; 175 Fifth Ave., New York, NY 10010 (212/460-1500 or 800/777-4643).◊**

Physical Mapping Strategies for the Human Genome

Eric Green [Washington University (WU) School of Medicine] discussed his group's early progress in establishing a complete physical map of human chromosome 7. The strategy includes isolating DNA in YACs, assembling YAC contigs based on STS identification in isolated clones, and constructing STS-content maps that reflect spatial relationships deduced from assembled YAC contigs. A detailed summary was published in *PCR Methods and Applications* **1**, 77–90 (1991).

The focus of this project has been to develop high-throughput STS-specific PCR assays from random DNA sequences derived from human chromosome 7. These efforts have produced more than 400 chromosome 7–specific STSs, all of which are sufficiently robust for screening a comprehensive YAC library. Computer simulations by Phil Green (WU) suggested that current STSs probably represent one-third to one-half of those required to construct an STS-content map of human chromosome 7 with 100- to 200-kb resolution.

More than 600 chromosome 7–specific YACs have been isolated from the WU total human DNA library. Characterization of these clones, which together should contain about 15 to 25% of the DNA from chromosome 7, is under way. The new library should have larger cloned inserts and fewer chimeric YACs and reduce the screening required to identify YACs containing particular STSs. Eric Green also discussed the relative merits of PCR vs hybridization-based approaches for identifying physical landmarks in sets of isolated YAC clones.

Mapping Technologies Transferable to Mouse Genome Mapping Project

Pieter de Jong (Lawrence Livermore National Laboratory) discussed important human DNA cloning methods that might be applied to the mouse. He favored using small cloned fragments from chromosome-enriched libraries as anchors in combination with chromosome-enriched YAC or bacterial artificial chromosome arrays to prepare a contig map of the mouse genome. Using cosmid libraries made from flow-sorted chromosomes as an anchor source would require extensive preparation, de Jong felt, and a more realistic approach would be to construct small-fragment PCR libraries for (1) individual mouse chromosomes, using a single-event flow-sorting approach, or (2) individual chromosome regions by microdissection.

Future Directions

The group discussed the need to make available to the mouse community mapping resources such as YAC and cosmid libraries and cDNA arrays. To meet this need, the NCHGR RFA HG-92-002 provides that applicants distribute resources.

Many attendees felt that Human Genome Project goals would be achieved more quickly if mouse genomic analysis were tied closely to functional analysis. Experimental approaches currently allow evaluation on a gene-by-gene basis of the effect of a given gene on the entire phenotype. Given the rapid progress being made in generating genetic and physical maps and accumulating sequence information, new, efficient, cost-effective methods should be developed for functional gene mapping, embryonic stem cell resources, and gene targeting. The mouse offers an appropriate model for these technologies. The working group agreed to devote most of the next meeting to a fuller discussion of these issues and suggested that other NIH components be encouraged to participate.◊

Reported by Bettie Graham
NIH NCHGR
and
Verne Chapman
Roswell Park Memorial Institute

A complete meeting report is available from the Joint Mouse Working Group. Contact:

- **Bettie J. Graham, Chief**
  Research Grants Branch
  NIH NCHGR
  Bldg. 38A, Room 617
  Bethesda, MD 20892
  301/496-7531
  Fax: 301/480-2770

E-Mail Searcher Verifies Protein Sequence Homology

As an aid to detecting and verifying protein sequence homology, Fred Hutchinson Cancer Research Center (FHCRC) has introduced the BLOCKS e-mail searcher. It compares a submitted protein or DNA sequence with a database of protein sequence blocks (short, multiply aligned, ungapped segments corresponding to the most highly conserved regions of proteins). This database was constructed by successively applying the automated PROTOMAT system to individual entries in the PROSITE catalog of protein groups that are keyed to the SWISS-PROT protein sequence databank.

Typically, members of a protein group have more than one region in common; this relationship is represented as a series of blocks separated by the unaligned regions. The concentrated form of information from multiply aligned sequences reduces background and increases sensitivity for detection of distant relationships. If a particular block scores high when a submitted DNA sequence is compared with the database, the sequence may be related to the group represented by the block.

For a detailed help file, send a blank e-mail message to *blocks@howard.fhcrc.org* with *help* in the subject line. A protein or DNA sequence can be transmitted in FASTA, Genepro, GenBank®, EMBL, SWISS-PROT, GCG, or PIR (DNA is automatically translated in all six reading frames for searching). This work was funded by NIH and the Howard Hughes Medical Institute (HHMI). Contact: Steven Henikoff; HHMI-FHCRC; 1124 Columbia St.; Seattle, WA 98104 (206/667-4515, Fax: -5889; Internet: *steveh@sparky.fhcrc.org*).◊

GenBank Transition Provides Continuity

On October 1 the NIH GenBank® Genetic Sequence Database, an international database of known DNA sequences, moved to the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM). GenBank contains over 100 million nucleotides, up from 40 million in 1991 and 1.5 million in 1985.

For the past 10 years, GenBank has been managed by the National Institute for General Medical Sciences (NIGMS), with database production and distribution contracted to Bolt, Baranek, and Newman from 1982 to 1987 and to IntelliGenetics, Inc. (IG) from 1987 to 1992. In 1988, Congress established NCBI to help create a national program for developing information systems for molecular biology. Over the past 2 years, NCBI has been working with NIGMS and IG to ensure a smooth transition.

GenBank data will continue to be compiled from direct submissions and journal scanning, including all journals previously scanned. In addition, specially trained biological indexers in the NLM Library Operations Division will identify and annotate nucleotide and protein sequences from more than 3600 journals in MEDLINE®, augmented by plant and veterinary science journals from the National Agricultural Library. Citations for sequence data are enhanced with abstracts from MEDLINE.

NCBI News, ***a newsletter designed to keep the community informed on database and software distribution services, is distributed three times a year. To be added to the mailing list or to request information on NCBI services, contact NLM NCBI; Bldg. 38A, Room 8N-803; Bethesda, MD 20894 (301/496-2475, Fax: 301/480-9241, Internet:*** **info@ncbi.nlm.nih.gov**).◊

Authors should continue to submit data and Authorin material directly to Los Alamos National Laboratory (LANL) (Internet: *gb-sub@genome.lanl.gov*); under an interagency agreement, LANL processes, maintains, and releases submitted sequence data to NCBI. Other author submissions come from the DNA databases EMBL in Germany and DDBJ in Japan. To facilitate the submission of sequence data, the Authorin program for PC and Macintosh computers is available from IG (800/477-2459). Corrections and sequences missing from published data should be reported to *update@ncbi.nlm.nih.gov.*

NCBI-GenBank data is available on CD-ROM and through e-mail servers and FTP over the Internet (see box at right). The standard flat file format for GenBank data distribution will continue, in addition to an integrated database version that includes retrieval software on CD-ROM for text searching and browsing related bibliographic and sequence entries. See p. 13 for information on the NCBI repository of molecular biology databases.◊

Reported by Barbara Rapp NCBI

ACCESS TO GENBANK

E-mail Servers

The Retrieve e-mail server accepts single or multiple text strings (e.g., locus names, accession numbers, keywords, author names) as queries, runs an IRX search against a single specified database, and returns the matching full records as a mail message. The BLAST e-mail server accepts either a nucleic acid or protein query sequence in FASTA format, runs the search against a single specified database or NCBI's combined nonredundant database; results are returned in an e-mail message.

An NCBI phone number or special account is not needed to use the system, only the ability to send an electronic mail message to NCBI. The query structure for mail messages is nearly identical to the IG server, but NCBI documentation should first be obtained by sending the one-word message (*help*) to the addresses below.

- *retrieve@ncbi.nlm.nih.gov* (Retrieve server).
- *blast@ncbi.nlm.nih.gov* (BLAST server).

Anonymous FTP

NCBI-GenBank is also distributed over the Internet through the FTP program from

- *ncbi.nlm.nih.gov* or *130.14.20.1.*

The full release in flat file format is available as compressed files in the directory *<ncbi-genbank>*. A cumulative update file is contained in the subdirectory *<daily>* and a noncumulative file in the subdirectory *<daily-nc>.*

CD-ROM Distribution

NCBI CD-ROMS are available through annual subscriptions that include a full release every 2 months. The following titles may be ordered through the Government Printing Office (202/783-3238, Fax: /512-2233) or from the Superintendent of Documents; P.O. Box 371954; Pittsburgh, PA 15250-7954:

- *NCBI-GenBank* incorporates all cumulative GenBank DNA data and is intended for users who have been receiving the flat file format directly from IG in the past. No software is included. Users should contact the producer of their commercial software to confirm compatibility. ($47 per year.)
- *Entrez: Sequences* integrates nucleotide sequences, protein sequences, and journal abstracts from GenBank, EMBL, SWISS-PROT, DDBJ, and PIR linked to MEDLINE citations, with text retrieval software for Macintosh and PC-compatible systems. For information on compatibility with commercial software packages, contact software producers. ($57 per year.)
- *NCBI-Sequences* is the integrated sequence data set from the *Entrez: Sequences* CD-ROM in the ASN.1 standard data description format. This is a data distribution disk containing the database only and is intended primarily for software developers. No software is included. ($47 per year.)

A CD-ROM with sequence data organized for similarity searching by FASTA is planned for early 1993. For further information on CD-ROMs, contact NCBI at the newsletter address in the box at left.◊

Genome News

DOE Announcements

Award, Activities, Resources Cited

R&D 100 Award for GRAIL

GRAIL contact:

Ed Uberbacher
615/574-6134
Fax:-7860
Internet:
ube@stc10.ctd.ornl.gov
BITNET:
ube@ornlstc.bitnet

The Gene Recognition and Analysis Internet Link (GRAIL) developed by Oak Ridge National Laboratory (ORNL) researchers has received an R&D 100 award from *R&D Magazine* as one of the 100 top new technologies for 1992. GRAIL, which was developed by Edward Uberbacher, Reinhold Mann, and Richard Mural, is a neural network that uses machine learning to identify exons, the instruction-giving sequence regions of genes. Each year inventors are honored at an awards dinner at the Museum of Science and Industry in Chicago, and the winning technologies are displayed there for a month.

Distinguishing between exons and introns (DNA that has little or unknown useful information) has been a very slow and labor-intensive task. To speed this process, investigators first obtained sequences from a DNA sequence database and assembled sets of them to train the neural network to recognize patterns. Users worldwide can now e-mail DNA sequence files to the system and have the analysis returned automatically by e-mail.◊

GnomeView: Visual Representation of Data

GnomeView, developed by Richard Douthart's team at Pacific Northwest Laboratory (PNL) is undergoing beta testing as a user interface to access information generated by the Human Genome Project. GnomeView provides graphical and textual styles of data presentation and displays color representations of genomic maps for review, analysis, and manipulation. It supports both chromosome and DNA sequence maps with data from Genome Data Base and GenBank, respectively. The standard representation of a chromosome map in GnomeView is a stylized depiction of the chromosome's banding pattern. DNA sequence is represented as a number line with features from GenBank as regions and the base pairs as tick marks or letters. GnomeView also provides density maps, a type of color-coded histogram indicating the distribution of objects over an associated genomic map.

GnomeView addresses the problem of moving among databases by using cross references to achieve smooth transition across databases and among different levels of data. Given a GDB locus, for example, GnomeView can retrieve all the associated GenBank sequences, and vice versa. Another unique feature is the ability to magnify any region and scroll maps interactively in either direction. Overlapping labels become separable as the magnification increases. Regions, tick marks, and letters are color coded so that patterns can be discerned more easily. The GnomeView window-based query interface allows the user to specify complex queries through the use of buttons and menus, with minimal text entry.

ASCII flat files are downloaded from the public databases to a local disk via anonymous FTP over the Internet. The local GnomeView database is then loaded from these flat files. The user interacts with the local GnomeView database via a graphical user interface.

GnomeView will be demonstrated at the February DOE Contractor-Grantee meeting in Santa Fe, New Mexico, and will be available to the general community in 1993. For more information, contact Richard Douthart; Life Sciences Center; PNL; Richland, WA 99352 (509/375-2653, Fax: 509/375-3649, Internet: *dick@gnome.pnl.gov*).◊

Publications About the Human Genome Project

Extensive sections on the Human Genome Project are included in recent publications from three DOE national laboratories:

Lawrence Berkeley Laboratory (LBL). "Unraveling the Genetic Message" describes genome program goals and projects under way at LBL, with special emphasis on informatics and the automation of laboratory procedures, instrumentation, and technology. *LBL Research Review* **17**(1), 2–15 (Spring 1992). [Available from LBL Research Review; LBL; Building 50C; Berkeley, CA 94720 (510/486-5771).]

Lawrence Livermore National Laboratory (LLNL). "The Human Genome Project" discusses basic molecular genetics; advances in informatics and instrumentation; and the application of mapping and sequencing procedures to the genome program, particularly progress in chromosome 19 mapping strategies, techniques, and data analysis at LLNL. Information on the National Laboratory Gene Library Project is also included. *Energy and Technology Review*, 29–62 (April–May 1992). [Available from Leilani Corell; LLNL; Bldg. 361, MS 452; P.O. Box 808; Livermore, CA 94551-9900 (510/423-3841)] or from National Technical Information Service (NTIS); U.S. Department of Commerce; 5285 Port Royal Road; Springfield, VA 22161.]

Oak Ridge National Laboratory (ORNL). "New Technologies for DNA Sequencing" and "Covering All the Bases: ORNL Probes the Human Genome" describe sequencing research and technologies being developed at ORNL. These activities include stable isotope sequencing, luminescent labeling systems, high-speed gel-less sequencing methods based on mass spectroscopy and hybridization, DNA imaging by atomic force and scanning tunneling microscopy, supercomputer analysis of DNA-carcinogen interactions, and Gene Recognition and Analysis Internet Link (GRAIL) software. ORNL investigators are using insertional and targeted mutagenesis and homologous

recombination to study mouse-human homologies; develop physical, functional, and mutation maps of the mouse genome; and predict functions of the corresponding regions of the human genome. *Oak Ridge National Laboratory Review* **25**(1), 18–39 (1992). [Available from ORNL Review; Bldg. 4500S, MS 6144; P.O. Box 2008; Oak Ridge, TN 37831-6144 (615/574-7183 or -6974) or from NTIS at the address above.]

DOE Funds BSCS Instructional Module

The Biological Sciences Curriculum Study (BSCS) is mailing to all U.S. high school biology teachers a 94-page, DOE-funded instructional module entitled "Mapping and Sequencing the Human Genome: Science, Ethics, and Public Policy." The unit contains 31 pages of background information for the teacher about the objectives, science, and technology of the Human Genome Project and the project's ethical and public policy implications. Some 63 pages describe 4 detailed, genome-related instructional activities for the classroom, including copymasters for students, annotated teacher pages, questions for discussion, glossary, and list of references.

The materials are designed for use by students in first-year biology courses (generally taught in the 10th grade), with extensions and elaborations appropriate for advanced students. About 80% of all U.S. high school students, some 2.25 million annually, take introductory biology.

In collaboration with the American Medical Association, a team headed by Joseph McInerney (Colorado College) developed the module in a 16-month BSCS project. For more information contact BSCS; 830 North Tejon, Suite 405; Colorado Springs, CO 80903 (719/578-1136; Fax: -9126).◊

HGMIS Requests Information on Software and Instrumentation

As space permits, *Human Genome News* will publish selected short announcements about software and instrumentation available to the genetics research community. Genome researchers may send such information by mail, e-mail, or fax to Betty Mansfield at the Human Genome Management Information System address on p. 12.◊

Software/Databases and GenBank Submissions at LANL

James Fickett, Michael Cinkosky, and Joe Gatewood [Los Alamos National Laboratory (LANL)] have developed a new e-mail service to help cDNA sequencing laboratories stay aware of overlaps with other laboratories. cDNA-Inform, a collection of public and private sequences and software for automatic searches and comparisons, can be queried by e-mailing a newly determined sequence to the address below. Sequence identifiers and contact information are reported to owners of matching sequences, who are continually informed of new matches. Although matches are reported, privately submitted sequences are seen only by the contributor.

To use cDNA-Inform, send a message to *bioserve@t10.lanl.gov* containing the text *cdna-inform* in the first line. For FASTA format, start the second line with a right angle bracket (>) followed by a sequence identifier (if no identifier is supplied, cDNA-Inform will assign one). On the same line, identify the sequence source. Place the sequence, with no annotation, numbering, or punctuation, on the third and succeeding lines. For GenBank format, the first line should be *cdna-inform -g*, and the remaining message should contain a sequence in GenBank entry format. [Contact: James Fickett; LANL; T-10 MS K710 (505/665-5340, Internet: *jwf@life.lanl.gov*).]◊

Contractor-Grantee Workshop

The DOE 1993 Contractor-Grantee Workshop will be held February 7–10 in Santa Fe, New Mexico. At least one representative from each DOE-funded project, including those dealing with ethical, legal, and social issues (ELSI), is expected to attend. Abstracts are past due to Sylvia Spengler; Human Genome Center; Lawrence Berkeley Laboratory; 1 Cyclotron Road; MS-1-213; Berkeley, CA 94720 (510/486-4943, Fax: -5717, Internet: *sylviaj@violet.berkeley.edu*).◊

Applications for Fellowship Program

Applications for the next cycle of DOE Human Genome Distinguished Postdoctoral Fellowships are now being accepted by the Oak Ridge Institute for Science and Education (ORISE). Application deadline is February 1, 1993.

This fellowship program was created to offer challenging training opportunities for recent doctoral degree recipients to conduct research in support of the DOE Human Genome Program. Up to 2 years are served at university or DOE laboratories having substantial DOE-sponsored genome research. Stipends are $35,000 for the first year and $37,000 for the second. Contact: Linda Holmes; ORISE Science and Engineering Education Division; P.O. Box 117; Oak Ridge, TN 37831-0117 (615/576-4805, Fax: -0202).◊

1991–92 Program Report

The DOE *Human Genome 1991–92 Program Report* is available to requestors without charge. Domestic subscribers to *Human Genome News* should have received their copies of the 246-page, red-covered program report. The document will be sent to foreign subscribers on request.

The report includes the following:

- history and goals of the U.S. Human Genome Project,
- highlights of research progress,
- narratives from the three DOE human genome centers,
- program management infrastructure,
- international coordination,
- abstracts of DOE-funded research, and
- primer on molecular genetics.

The "Primer on Molecular Genetics" is also published as a separate document. Multiple copies may be requested from HGMIS for educational purposes. To request copies of either of these publications, use the Subscription/Document Request Form on p. 16.◊

GDB Forum

GDB Version 5.0 Highlights

New GDB Version Enhances Content, Data Relationship, Software for Retrieval and Display

Genome Data Base (GDB) Version 5.0 contains major new features enhancing database content, data relationships, and software for data retrieval and display. Highlights are summarized below, and a detailed description will be available online in the Release Notes under *News*.

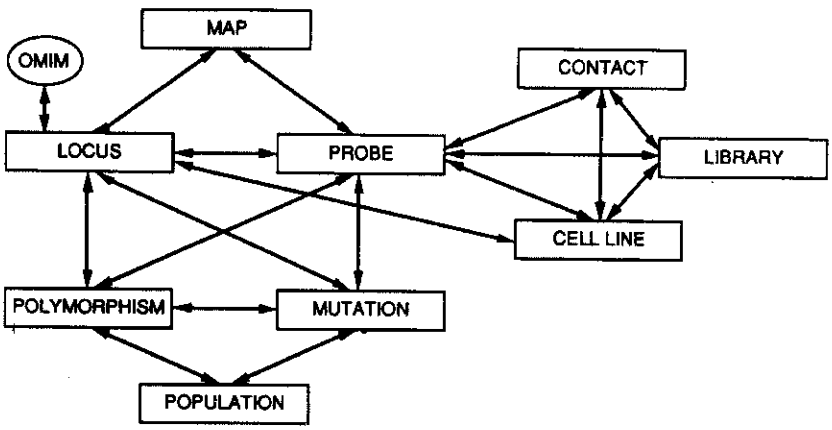
Database Content and Organization

Information in GDB is organized into ten data managers: *Map*, *Locus*, *Probe*, *Polymorphism*, *Mutation*, *Population*, *Library*, *Cell Line*, *Citation* (previously named *Source*), and *Contact*. Five of them are newly added to increase the searchability and types of data that can be retrieved and displayed:

- *Polymorphism* – Data previously available for display only is now directly searchable and has been reorganized to show on one screen all polymorphism information (location, detection method, alleles, frequencies, and population).
- *Population* – Population definitions, also previously available for display only, have been broken down into component elements such as race, geographical region, or ethnic group and reconstituted to facilitate creation of new population definitions. Mutations and polymorphisms specific to particular populations can now be retrieved.
- *Mutation* – Mutation data will include location, sublocalization within the genes, and comparison of wildtype and mutant nucleotide and amino acid sequences.
- *Cell Line* – Definitions can be retrieved independently from their breakpoints.
- *Library* – DNA library descriptions will aid researchers in obtaining clones from the GDB contacts.

Significant amounts of new data have also been added to three existing managers:

- *Map* – Distance information (including degree of overlap) between two map objects is now included. Confidence limits for localization data are also available.
- *Probe* – Amplification conditions for PCR primer sets include buffer concentrations, time/temperature cycles, and the name of the thermal cycler used. These conditions are linked to citations, and multiple sets of amplification conditions can be linked to a single set of primers. Probe-to-probe interactions will make available such data as sequence tagged site screening of yeast artificial chromosomes.
- *Citation* – Citations are ranked (important, supportive, background)



Relationships among GDB data managers and the Online *Mendelian Inheritance in Man* (OMIM). OMIM data are linked to citations in all managers except *Contact*.

GDB 5.0 Data Conversion Aids for Software Developers

When GDB version 5.0 is released, the FTP server flat files and the SQL server database (both generated weekly from the online GDB) will reflect the 5.0 format. To help developers test programs for converting from 4.2 to 5.0 format, GDB plans to set up two new FTP directories containing flat files with equivalent data in two formats:

- GDB 4.2 data at time of conversion and
- GDB 5.0 data just after conversion.

The *README* file in the top-level directory will describe the location of these new test files.

A second SQL server containing 4.2 data at the time of conversion will also be available. Information about this test server will be sent via e-mail to everyone with an SQL account.

The test files and server will be accessible for 3 months. Developers interested in obtaining the 5.0 schema and accompanying data dictionary should

with respect to all linked entries. MEDLINE® citations include cross references to other databases including GenBank®, EC (Enzyme Commission) number, RN (CAS Registry) number, and symbols or abbreviated forms of gene names as they appear in the published citation.

Flexibility has been increased for retrieving linkage relationships among entries in different managers:

- Multiple entries can be selected before another manager is called to retrieve linked entries.
- A list of all linked data for a selected entry is available from the *View* menu.

Retrieving and Displaying Data

A query can retrieve any number of entries. If more than 500 entries are retrieved, data are grouped in sets and the user can move between sets. To provide a variety of ways to search for data, a basic set of fields (*Cytogenetic Location, Locus Symbol, Locus Name, Probe Symbol*) is included on the *retrieve* screens for *loci, probes, polymorphisms, mutations*, and *populations*. The *detail view* screens in each of these managers also include these basic fields.

To make information available as quickly as possible, new entries that are complete in fields and citations but subject to modifications during review will be viewable by general users as *PROPOSED DATA*. However, unpublished material may be withheld from view for up to 6 months if requested by the submitter.

Formats Available for Data Output

Output functions now include additional types and formats of GDB data that can be sent via e-mail. In addition to personal search results, standard reports (see announcement at bottom of page) are also available in ASCII, Tab-delimited, and PostScript formats.

Using GDB

An online system provides help with individual screens and fields and general topics. Help is now included in *Citation* and *Locus* managers and will be added to other managers on an ongoing basis.

Menu choices are selected through <control> *R* followed by unique menu letters that are consistent across all screens and managers. A set of function keys is available for the most commonly used menu and submenu options.◊

GDB USER REGISTRATION

To become a registered user of GDB and OMIM, contact one of the User Support offices listed at right (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.

GDB and OMIM Training Course Schedule

Comprehensive hands-on training courses on the use of GDB and OMIM will have at least one computer workstation for two participants. Registrants will receive at least 3 weeks notice if insufficient registration causes class cancellation.

- The general course for scientific users provides a basic understanding of the databases and relationships among different types of data.
- The course for users with editing privileges includes instructions on adding, modifying, and deleting GDB data.

Class frequency and location will be determined by demand (schedule below). Courses are free, but attendees must pay their own travel and lodging expenses. Hotel information and directions will be mailed with registration materials.

As interest in GDB continues to grow, organizations around the world will offer training that requires access to GDB in Baltimore. Notifying GDB User Support about planned training activities will enable the staff to ensure database availability by scheduling maintenance and repairs at other times.

COURSE REGISTRATION INFORMATION: contact U.S. GDB User Support Office (at upper right).

PLANNED EXHIBITIONS (acronym list, p. 16)

- AAAS, Boston, Feb. 11–16, 1993.
- Experimental Biology '93, New Orleans, March 28–April 1, 1993.
- AAP/AFCR/ASCI, Washington, D.C., April 30–May 3, 1993.

Course	Dates
BALTIMORE	
General User	February 22–23, 1993
General User	April 26–27, 1993
Editing	May 9–11, 1993
General User	June 21–22, 1993

SUPPORT OFFICES

United States
GDB User Support
Welch Medical Library
1830 E. Monument Street, Third Floor
Baltimore, MD 21205
410/955-7058
Fax: 410/614-0434
Internet: *help@welch.jhu.edu*

The Help Line 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.

United Kingdom
Christine Bates
Human Gene Mapping Program Resource Center
CRC, Watford Road
Harrow, Middx HA1 3UJ U.K.
(Int.) 44/81-869-3446
Fax: (Int.) 44/81-869-3807
Internet: *cbates@uk.ac.crc*

Germany
Otto Ritter
Molecular Biophysics Group
German Cancer Research Center
Im Neuenheimer Feld 280
D-6900 Heidelberg 1, FRG
(Int.) 49/6221-42-2372
Fax: (Int.) 49/6221-40-1271
Internet: *dok261@cvx12.dkfz-heidelberg.de*

Australia
Alex Reisner
ANGIS
Electrical Engineering Building, J03
University of Sydney
Sydney, N.S.W. 2006
Australia
(Int.) 61/2-692-2948
Fax: (Int.) 61/2-692-3847
Internet: *reisner@ee.su.oz.au*

GDB Standard Reports Available via FTP

A series of standard reports will be generated regularly for the GDB FTP server (*mendel.welch.jhu.edu* or *128.220.59.42*). The first set is based on the content and style of tables accompanying the HGM 11 chromosome summary reports. Each report will contain a specific extract of GDB data and will be available in these formats: (1) PostScript files (very large) can be sent to the user's local PostScript printer to produce a formatted document of near-typeset quality. (2) Formatted ASCII files (smaller than PostScript files) produce human-readable copy. (3) Tab-delimited ASCII tables are in computer-readable format for input to the user's computer program. The *README* file in the *<gdb/report>* directory will describe the location and filenaming conventions for the report series. For information about FTP access or to suggest report enhancements or subjects for new reports, contact GDB User Support in Baltimore.◊

Meeting Reports

Human Genome news

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

Managing Editor
Betty K. Mansfield

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

Production Manager/Editor
Judy M. Wyrick

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

Correspondence Address:
Betty K. Mansfield
ORNL
P.O. Box 2008
Oak Ridge,TN 37831-6050
Phone: 615/576-6669
Fax: 615/574-9688
BITNET: *bkq@ornlstc*
Internet: *bkq@ornl.gov*

Sponsor Contacts:

Daniel W. Drell
DOE Program Office
Germantown, MD 20545
301/903-4742
Fax: 301/903-5051
Internet:
drell@mailgw.er.doe.gov

Leslie Fink
NIH National Center for Human Genome Research
Bethesda, MD 20892
301/402-0911
Fax: 301/480-2770
Internet: *lsf@cu.nih.gov*

National Center for Human Genome Research

Conference on Ethics and Technology

The topic "How Should Insurance Companies Use Genetic Information?" was featured at the first annual Conference on Ethics and Technology held May 8–9 at Southern Oregon State College (SOSC) in Ashland. According to organizers, the conference received very positive reactions from the public and has led to statewide interest in establishing (1) an intercollegiate faculty advisory board of bioethics to foster informed discussion about emerging ethical issues in technology and health care and (2) a southern Oregon human genome network to distribute information to the professional and lay communities about Human Genome Project goals and objectives.

About 100 people attended the meeting, which was hosted by the newly formed Oregon Honors Consortium comprising Portland State University, the University of Oregon, Western Oregon State College, and SOSC. *The Oregon Scholar,* an annual consortium publication devoted to scholarly work on important social themes, will feature selected student papers resulting from the conference.

Presentations at the meeting were made by Eric Juengst (NIH National Center for Human Genome Research), Jude Payne (Health Insurance Association of America), Paul Billings (Pacific Presbyterian Medical Center), and Michael Garland (Oregon Health Sciences University). Each presentation was followed by student responses and public discussion.

Gregory Fowler, Honors Program Director, said that the SOSC Churchill Scholars Honors Program had prepared the students to take leading roles in the conference and public forum. Having spent the spring term studying the Human Genome Project on their home campuses, participants welcomed the opportunity to exchange ideas and to work closely with the four visiting experts.

The conference was supported by the Ethics and Technology Lectureship Program of the GTE Foundation, the Oregon Council for the Humanities, and the Southern Region office of Blue Cross–Blue Shield of Oregon.◊

Reported by Gregory Fowler
Southern Oregon State College

First South-North Human Genome Conference

The First South-North Human Genome Conference was held May 12–15 in Caxambu, Brazil. About 200 participants from 22 countries assembled to envision strategies that would enable developing nations to participate in the Human Genome Project. Main conference sponsors were the United Nations Educational, Scientific, and Cultural Organization (UNESCO) and the Brazilian Society of Biochemistry and Molecular Biology, whose president, Sergio Pena (Federal University of Minas Gerais), chaired the organizing committee. Other financial support was provided by a number of international and Brazilian agencies.

Major Conference Themes

- The impact of the Human Genome Project on the developing world.
- Human genetic diversity in the context of the Human Genome Project.
- The practice of molecular biology in developing nations.

Meeting Highlights

Georgio Bernardi (Institut Jacques Monod, Paris) spoke on genome organization, introducing the concept of isochores. Cassandra Smith (Boston University) summarized her group's progress with the mapping of chromosome 21. Lap Chee Tsui (Hospital for Sick Children, Toronto) described finding the cystic fibrosis gene. Phyllis McAlpine (University of Manitoba) examined the question of gene nomenclature.

J. Craig Venter (formerly with NIH, now at the Institute for Genomic Research) described his strategy of gene identification by expressed sequence tag/cDNA. His presentation emphasized two points:

- the cost of establishing a competitive laboratory (at a few hundred thousand dollars) is within reach of Third World countries, and
- developing countries could advance technologically and contribute effectively to the overall effort by studying

the genomes of relevant organisms, such as parasites or major crop plants.

Nancy Wexler (Columbia University) reviewed her work with the Venezuelan Lake Maracaibo population that led to mapping the Huntington's chorea gene to chromosome 4. This project represents significant South-North collaboration.

In discussing how Third-World countries could simplify and adapt molecular biology techniques, Pena described his team's development of nonisotopic probes and techniques for in-gel hybridization; he stated that the next frontier for the Third World will be the development of practical nonisotopic DNA sequencing techniques. Sakol Panyim (Mahidol University, Thailand); Onesmo ole-MoiYoi (International Laboratory for Research on Animal Diseases, Kenya); and Pedro Leon (University of Costa Rica) showed how elegant genomic studies can be accomplished in the Third World.

Recent methodological advances now permit the launching of the worldwide Human Genome Diversity Project, expected to be a proving ground for North-South collaboration in gathering information on ancient human populations. Conference attendees recommended that (1) the work begin soon because many key populations are at risk of becoming extinct or losing their genetic distinctiveness, (2) Third-World research groups should be involved fully in the study, and (3) appropriate technology should be transferred to these research groups.

The activities of the Human Genome Organization (HUGO) and UNESCO in the international coordination of the Human Genome Project were summarized, and Walter Bodmer (Imperial Cancer Research Fund) announced plans for establishing a Latin American office of HUGO.

The Second South-North Genome Conference will take place in Bangkok in October 1993.◊

Reported by Sergio Pena
Federal University of Minas Gerais

■*RESOURCES*

HUMAN GENOME PROGRAM NEWSGROUP FACILITATES COMMUNICATION

The Human Genome Program Newsgroup, sponsored by the National Science Foundation and the NIH and DOE human genome programs, operates through the BIOSCI electronic bulletin board network to allow researchers worldwide to communicate, share ideas, and find solutions to problems. Genome-related information is distributed through the newsgroup, including requests for grant applications, reports from recent scientific and advisory meetings, announcements of future events, and listings of free software and services. Genome program staffs participate in the newsgroup and respond to postings when appropriate (address for posting messages: *gnome-pr@net.bio.net*). Nongenome researchers are encouraged to use this forum for discussing genome issues with the research community and funding agencies.

The BIOSCI network also supports about 30 other newsgroups of interest to biologists (see box below for a partial list). For information on these newsgroups and BIOSCI, send a message to the Internet address (*biosci@net.bio.net*), which also serves BITNET users. For help in using these electronic addresses, contact Dave Kristofferson (415/962-7339 or *kristoff@net.bio.net*).

The Human Genome Program Newsgroup may also be accessed on USENET as *bionet.molbio.genome-program*. E-mail subscriptions are not needed by those who have access to USENET NEWS. All interested users are encouraged to obtain USENET NEWS software, which is in the public domain and more convenient than subscribing to newsgroups by e-mail.◊

BIOSCI NEWSGROUPS: PARTIAL LISTING

ARABIDOPSIS: Arabidopsis genome project newsgroup

BIO-MATRIX: Computer applications to biological databases

BIO-SOFTWARE: Biological sciences software

CHROMOSOME-22: Chromosome 22 mapping and sequencing

COMPUTATIONAL-BIOLOGY: Mathematical and computer applications in biology

GDB: Messages to and from Genome Data Base staff

GENBANK-BB: Messages to and from GenBank staff

GENETIC-LINKAGE: Genetic linkage analysis newsgroup

MOLECULAR-EVOLUTION: Molecular evolution research discussions

PROTEIN-ANALYSIS: Protein research discussions

SCIENCE-RESOURCES: Scientific funding agency information

NCBI REPOSITORY OF MOLECULAR BIOLOGY DATABASES

NCBI maintains a repository over 20 molecular biology databases that are freely available for network users through Internet FTP. Some of the databases are listed below with their curators.

- ACEDB: *C. elegans* Genomic Database (Richard Durbin and J. Thierry-Meig)
- EcoSeq and EcoMap: Integrated *E. coli* Sequence and Map Data (Kenn Rudd)
- EPD: Eukaryotic Promoter Database (Philipp Bucher)
- FlyBase: *Drosophila* Genetic Database (Michael Ashburner)
- KABAT: Sequences of Proteins of Immunological Interest (Elvin Kabat)
- PROSITE: Dictionary of Protein Sites and Patterns (Amos Bairoch)
- REBASE: Restriction Enzyme Database (Richard Roberts)
- SWISS-PROT: Protein Sequence Database (Bairoch)
- TFD: Database of Transcription Factors (David Ghosh)

A <*toolbox*> directory contains a set of software and data exchange specifications used by NCBI to produce portable software and includes ASN.1 tools and specifications for molecular sequence data. A <*pub*> directory offers public-domain software such as BLAST (a sequence-similarity search program) and MACAW (a multiple-sequence-alignment program).

All data in this repository can be transferred over Internet by FTP. To connect, users should type *ftp ncbi.nlm.nih.gov* or *ftp 130.14.20.1* and enter *anonymous* for the login name and their e-mail address as the password. Directories should be changed to <*repository*>, <*toolbox*>, or <*pub*> to download databases, ASN.1 tools, and public-domain software, respectively.

In addition to anonymous FTP access, NCBI is now distributing the May Data Repository CD-ROM at no charge on an experimental basis. The next release is scheduled for this month. A subscription service may be set up for future releases that occur at 6-month intervals, depending on demand and database update frequency. Questions, suggestions, requests for copies of the CD-ROM, and proposals for repository additions should be sent to Scott Federhen (301/496-2475, Fax: /480-9241, Internet: *repository@ncbi.nlm.nih.gov*).◊

Calendar of Genome Events* (acronyms, p. 16)

December

7. *DOE Human Genome Coordinating Committee; Bethesda, MD

7–8. DOE-NIH Joint Subcommittee on the Human Genome; NIH Program Advisory Committee on the Human Genome; Bethesda, MD [J. Ades, 301/402-2205, Fax: -2218]

8. WSS. Charles Cantor: Implications of the Human Genome Project; Washington, DC (charge for dinner) [D. McIlwain, 301/565-8861, evenings until 9:00 p.m. EST]

17. NCHGR Lecture Series. Neil Holtzman: Getting Genetic Tests to the Public—Safely and Effectively; Bethesda, MD [C. Dahl, 301/402-0838]

January 1993

5–8. Biotechnology Computing Track of the 26th Hawaiian International Conference on System Sciences; Maui, HI [L. Hunter, 301/496-9300, Fax: -0673, Internet: *hunter@nlm.nih.gov*]

7. NCHGR Lecture Series. Norman Arnheim: Analysis of DNA Sequences in Single Cells; Bethesda, MD [see contact: Dec. 17]

12. WSS. Phillip Sharp: MIT and the Human Genome; Washington, DC [see contact: Dec. 8]

17–22. "Advances in Gene Technology: Protein Engineering and Beyond" at the 1993 Miami Bio/Technology Winter Symposia; Miami Beach, FL [S. Black, 305/547-3597, Fax: /324-5665]

24–25. *National Advisory Council for Human Genome Research; Bethesda, MD [see contact: Dec. 7–8]

February 1993

1–6. Oncogenes and Anti-Oncogenes in Cell Differentiation, Development, and Human Cancer; Big Sky, MT [AACR, 215/440-9300, Fax: -9313]

7–11. *Third DOE Contractor-Grantee Workshop; Santa Fe, NM (abstract deadline: Nov. 16) [S. Spengler, 510/486-4879, Fax: -5717]

9. WSS. J. Craig Venter, Paul Watkins, and Wallace Steinberg: Is This the New Silicon Valley?; Washington, DC [see contact: Dec. 8]

15–19. *15th Annual Conference on the Organization & Expression of the Genome; Lorne, Victoria, Australia [S. Easteal, (Int.) 61/6-249-4719, Fax: -4712]

18. NCHGR Lecture Series. Richard Durbin: ACEDB Genome Database and Data Analysis from the Nematode Sequencing Project; Bethesda, MD [see contact: Dec. 17]

March 1993

6–8. Chromosome 20 Workshop; Paris [C. Smith, 510/643-6376, Fax: -1188]

9. WSS. Eric Juengst, Ruth Hubbard, and LeRoy Walters: Is This The Brave New World?; Washington, DC [see contact: Dec. 8]

11–12. First International Workshop on Human Chromosome 1; Cambridge, MA (abstract deadline: Dec. 15) [N. Dracopoli, 617/253-8575, Fax: /258-8728]

18. NCHGR Lecture Series. Steve Warren: Triplet Repeat Expansion Mutations—Example of the Fragile X; Bethesda, MD [see contact: Dec. 17]

25. NCHGR Lecture Series. Francis Collins: Identification of Human Disease Genes by Positional Cloning; Bethesda, MD [see contact: Dec. 17]

April 1993

12–18. Gene Therapy; Keystone, CO (abstract deadline: Dec. 2) [1993 Keystone Symposia Meeting, 303/262-1230, Fax: -1525]

12–18. Genetically Targeted Research & Therapeutics-Antisense & Gene Therapy; Keystone, CO (abstract deadline: Dec. 2) [see contact: April 12–18, above]

15. NCHGR Lecture Series. Troy Duster: Socio-Historical Context of Genetic Explanations of Behavior; Bethesda, MD [see contact: Dec. 17]

18–21. Third International Workshop on Human Chromosome 5; Irvine, CA [J. Wasmuth, 714/856-7067, Fax: /725-2688]

23–24. Fourth International Human Chromosome 21 Workshop; Paris [J. Delabar, (Int.) 33/1-42-730-960, Fax: -659]

May 1993

9–12. *Fourth Annual X Chromosome Workshop; St. Louis [M. Thomas, 314/362-7259, Fax: -1232]

14–15. Fourth International Workshop on Chromosome 3; Groningen, Netherlands [C. Buys, (Int.) 31/50-632-925, Fax: -947]

16. Chromosome 3 & Cancer; Groningen, Netherlands [see contact: May 14–15, above]

16–17. *National Advisory Council for Human Genome Research; Bethesda, MD [see contact: Dec. 7–8]

19–22. 84th Annual Meeting of AACR; Orlando, FL [see contact: Feb. 1–6]

20. NCHGR Lecture Series. Nancy Wexler: Long Day's Journey into Night—Search for the Huntington's Disease Gene; Bethesda, MD [see contact: Dec. 17]

20–22. First International Workshop on Chromosome 7; Marburg, Germany [K.-H. Grzeschik (Int.) 49/6421-28-4080, Fax: -5630]

June 1993

17. NCHGR Lecture Series. Jasper Rine: Dog Genome Initiative—Towards the Genetics of Morphology and Breed Traits; Bethesda, MD [see contact: Dec. 17]

21–22. DOE/NIH Joint Subcommittee on the Human Genome; NIH Program Advisory Committee on the Human Genome; Bethesda, MD [see contact: Dec. 7–8]

21–23. Second International Workshop on Chromosome 6; Berlin [A. Ziegler, (Int.) 49/30-3035-2617, Fax: -3778]

26–29. Human Gene Therapy; Washington, DC (abstract deadline: Mar. 15) [NY Acad. of Sci., 212/838-0230, Fax: -5640]

Training Calendar**†

December

8–11. DNA Fingerprinting; Rockville, MD [ATCC, 301/231-5566, Fax: /770-1805]

8–11. Molecular Modeling: Methods and Techniques; Athens, GA [ACS, 202/872-4508, Fax: -6336]

14–18. Advanced Recombinant DNA Methodology; ATCC, Rockville, MD [see contact: Dec. 8–11]

15. Introduction to PCR; Gainesville, FL [BTP, 800/821-4861, Fax: 515/232-8306]

16–18. Cloning & Hybridization of PCR Products; BTP, Gainesville, FL [see contact: Dec. 15]

21–22. Clinical Diagnosis Using PCR & Hybridization Analysis; BTP, Gainesville, FL [see contact: Dec. 15]

January 1993

4–7. PCR Methodology; Columbia, MD (also offered at later dates) [Exon-Intron, Inc., 410/730-3984, Fax: -3983]

4–8. Recombinant DNA Methodology; Washington, DC (also offered Mar. 8–12) [CATCMB/CUA, 202/319-6161, Fax: -5721]

9–11. PCR Techniques; CATCMB/CUA, Washington, DC [see contact: Jan. 4–8]

11–15. †Advanced Linkage Courses; New York (application deadline: Nov. 16) [K. Montague, 212/960-2507, Fax: /568-2750]

12–15. Recombinant DNA Techniques; New Brunswick, NJ (early registration by Dec. 31) (also offered Mar. 16–19 with early registration by Mar. 1) [Office of Continuing Professional Education, 908/932-9271, Fax: -8726]

25–29. Biomedical Initiative: Connection Machine Techniques Workshop; Pittsburgh (application deadline: Dec. 15) [N. Blankenstein, 412/268-4960, Internet: *blankens@a.psc.edu*]

February 1993

22–26. Recombinant DNA: Techniques & Applications; ATCC, Rockville, MD [see contact: Dec. 8–11, 1992]

➠

*Attendance at meetings listed with asterisk is either limited or restricted. Dates may change; check with contact person.

**Dates and course status may change, and courses may be offered at other times and places; check with contact person.

†NCHGR-funded event.

For Your Information

Genome-Related Teacher Workshops

Nine 2-week summer workshops on genetics and bioethics for secondary school biology teachers are scheduled in 1993 at various locations around the country. The workshops are part of Project Genethics, a 3-year program directed by Jon Hendrix and Thomas Mertens (Ball State University) and funded by the National Science Foundation. Attendees, limited to 24 at each workshop, are paid a $600 stipend and receive 4 semester hours of graduate credit. Two Saturday follow-up sessions are conducted at each site during the school year. For an application, call the area mentor teacher listed below after 5 p.m. at home.

June 14–25
Winter Park, FL [M. While, 407/678-1340 or C. Urbano, 517/642-2987]

Plymouth, MN [E. Thornton, 612/545-8976 or G. Mendenhall, 317/849-3022]

June 21–July 2
Florence, MT [C. Kuchel, 406/273-0207 or D. Brown, 317/747-9529]

June 28–July 9
Maplewood, NJ [L. Rosenbaum, 201/763-6667 or J. Dayner, 201/891-7541]

July 5–16
San Francisco [K. O'Neil, 206/459-5797 or M. McCaffrey, 215/233-4368]

July 19–30
Belmont, MA [G. Taylor, 617/484-1370 or J. Johnson, 407/359-9993]

Spring, TX [M. Bonetati, 713/353-3223 or J. Bealer, 602/378-6341]

August 2–13
Chicago [R. Yashon, 312/327-5658 or J. Ruhl, 317/463-7012]

Schenectady, NY [D. Ely, 802/862-5224 or B. Carvellas, 802/863-6560]◊

U.S. Genome Research Funding Guidelines

Note: Investigators wishing to apply for NIH and DOE funding are urged to discuss their projects with 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 and 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 staff during application development phases.

Program announcements are listed in the weekly *NIH Guide for Grants and Contracts*,* which is available through

- Hard-copy subscription: call 301/496-7441.
- Electronic version (E-Guide): Access through one of the following methods.
  1. Institutional Hubs. A designee receives automatic updates and distributes them locally to researchers. To use this NIH-preferred method, send a message naming the responsible person to Rebecca Duvall (BITNET: *q2c@nihcu*, Internet: *q2c@cu.nih.gov*).
  2. NIH Grant Line (also known as DRGLINE). User reads electronic bulletin board for weekly updates. Connection is through a modem, and a new feature allows files to be transmitted rapidly via BITNET or Internet. For more information, contact John James (301/496-7554 or BITNET: *zns@nihcu*).

*Full text of RFAs listed in the NIH grants guide may be obtained from either of the two electronic sources or from NIH NCHGR in Bethesda, Maryland (301/496-0844).

DOE Human Genome Program

Solicitations for proposals will be announced in early spring issues of the *Federal Register* and *Science* and in other publications. Formal proposals are due in August.

For further information, contact the program office via

- 301/903-5037, Fax: -5051, or Internet: *drell@mailgw.er.doe.gov*

SBIR Grants

DOE also invites small business firms to submit grant applications addressing the human genome topic of SBIR programs, which are designed to strengthen innovative firms in areas of research and development and to contribute to the growth and strength of the nation's economy. The human genome topic emphasizes instrumentation development for automated clone processing, improvements in DNA sequencing technologies, and enhanced sequence data storage and processing capabilities. Next submission date: March 8, 1993. For more information and a copy of the solicitation, contact

- Samuel Barish; SBIR Program Manager, ER-16; DOE; Washington, DC 20585; 301/903-5707.

Human Genome Distinguished Postdoctoral Fellowships

Next deadline: February 1, 1993. For further information, see p. 9 or contact

- Linda Holmes, Oak Ridge Institute for Science and Education: 615/576-4805.◊

Address Update

Please check the address label on this copy of *Human Genome News* to be sure the information is complete and current. Incomplete addresses may work for a time, but newsletters returned to this office result in the subscriber's name being removed from the mailing list.

Make corrections on the Subscription/Document Request Form (see the back of the newsletter) and return, **with the address label attached**, to the following address:

Laura Yust
Human Genome Management Information System
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN 37831-6050
Fax: 615/574-9888

If you are receiving duplicate mailings, please write *duplication* on the form and return with the incorrect address label.◊

March 1993

1–5. Recombinant DNA Methodology; Exon-Intron, Inc., Columbia, MD (also offered at later dates) [see contact: Jan. 4–7]

2–5. PCR Reaction/Cycle DNA Sequencing; ATCC, Rockville, MD [see contact: Dec. 8–11]

6–10. Kennedy Institute of Ethics Advanced Bioethics Course IV; Washington, DC [D. Michutka, 202/687-6771]

22–26. In Situ Hybridization & Recombinant DNA Technology; Exon-Intron, Inc., Columbia, MD (also offered Nov. 29–Dec. 3) [see contact: Jan. 4–7]

May 1993

17–21. Advanced Topics in Recombinant DNA; Exon-Intron, Inc., Columbia, MD (also offered July 19–23) [see contact: Jan. 4–7]

June 1993

6–12. Kennedy Institute of Ethics 19th Annual Intensive Bioethics Course; Washington, DC [see contact: Mar. 6–10]

(Late June) Workshop for Secondary Science Teachers; University of Kansas Medical Center, Kansas City [D. Collins, Fax: 913/588-3995] (See p. 5.)

August 1993

16–20. †Partnerships in Teaching Genome Technology Workshop; Ann Arbor, MI [P. Gregory, 313/747-3414, Fax: /763-4692]

October 1993

4–8. RNA Isolation and Characterization; Exon-Intron, Inc., Columbia, MD [see contact: Jan. 4–7]◊

Human Genome Management Information System

Subscription/Document Request (Vol. 4, No. 4)

Please type or print legibly. Enclose a business card, if possible. To change your name/address/affiliation or to be dropped from the mailing list, you must enclose your current newsletter address label to ensure that the proper entry is modified.

Name _____
(First) (MI) (Last)

Affiliation _____

Department/Division _____

Street and/or P.O. Box _____

City/State/Zip Code/Country _____

Phone _____ **Fax** _____

E-Mail Address _____ **Area of Interest** _____

1. ___ *Human Genome News*
   ___New Subscriber ___Change of Name/Affiliation/Address (circle all that apply) ___Drop Name from Mailing List
2. ___ *Understanding Our Genetic Inheritance, The U.S. Human Genome Project: The First Five Years, FY 1991–1995* (Joint DOE-NIH 5-Year Plan)
3. ___ DOE *Human Genome 1991–92 Program Report.*
4. ___ *Primer on Molecular Genetics.* (Included in program report above. Extracted as separate document for educational use.)

CALENDAR ACRONYMS

AAAS Am. Assoc. for the Advancement of Science

AACR Am. Assoc. for Cancer Research

AAP Association of American Physicians

ACEDB *Caenorhabditis elegans* Database

ACS Am. Chemical Soc.

AFCR American Federation of Clinical Research

ASCI American Society for Clinical Investigation

ATCC Am. *Type Culture* Collection

BTP Biotechnology Training Programs

CATCMB/CUA Center for Advanced Training in Cell and Molecular Biology/ Catholic Univ. of America

HGP Human Genome Project

MIT Mass. Institute of Tech.

NCHGR National Center for Human Genome Research

WSS Washington Seminar Series

Reader Comments:

HGMIS MAILING ADDRESS
Betty K. Mansfield
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN 37831-6050
U.S.A.