Converting scientific knowledge into commercially useful products

Transferring technology to the private sector, a primary mission of DOE, is strongly encouraged in the Human Genome Program to enhance the nation's investment in research and technological competitiveness. Human genome centers at Lawrence Berkeley National Laboratory (LBNL), Lawrence Livermore National Laboratory (LLNL), and Los Alamos National Laboratory (LANL) provide opportunities for private companies to collaborate on joint projects or use laboratory resources. These opportunities include access to information (including databases), personnel, and special facilities; informal research collaborations; Cooperative Research and Development Agreements (CRADAs); and patent and software licensing. For information on recently developed resources, contact individual genome research centers or see Research Highlights, beginning on p. 9. Many universities have their own licensing and technology transfer offices.

Some collaborations and technology-transfer highlights from FY 1994 through FY 1996 are described below.

Collaborations

Involvement of the private sector in research and development can facilitate successful transfer of technology to the marketplace, and collaborations can speed production of essential tools for genome research. A number of interactive projects are now under way, and others are in preliminary stages.

CRADAs

One technology-transfer mechanism used by DOE national laboratories is the CRADA, a legal agreement with a nongovernmental organization to collaborate on a defined research project. Under a CRADA, the two entities share scientific and technological expertise, with the governmental organization providing personnel, services, facilities, equipment, or other resources. Funds must come from the nongovernmental partner. A benefit to participating companies is the opportunity to negotiate exclusive licenses for inventions arising from these collaborations. For periods through 1996, the CRADAs in place in the DOE Human Genome Program included the following:

- LLNL with Applied Biosystems Division of Perkin-Elmer Corporation to develop analytical instrumentation for faster DNA sequencing instrumentation;
- LANL with Amgen, Inc., to develop bioassays for cell growth factors;
- Oak Ridge National Laboratory (ORNL) with Darwin Molecular, Inc., for mouse models of human immunologic disease;
- ORNL with Proctor & Gamble, Inc., for analyses of liver regeneration in a mouse model; and
- Brookhaven National Laboratory with U.S. Biochemical Corporation to identify proteins useful for primer-walking methods and large-scale sequencing.

Work for Others

In other collaborations, the LBNL genome center is participating in a Work for Others agreement with Amgen to automate the isolation and characterization of large numbers of mouse cDNAs. The center group is focusing on adapting LBNL's automated colony-picking system to cDNA protocols and applying methods to generate large numbers of filter replicas for colony

Technology Transfer Legislation

Technology transfer involves converting scientific knowledge into commercially useful products. Through the 1980s, a series of laws was enacted to encourage the development of commercial applications of federally funded research at universities and federal laboratories. Such laws [chiefly the Bayh-Dole Act of 1980, Stevenson-Wydler Act of 1980, and Federal Technology Transfer Act of 1986 (Public Laws 96-517, 96-480, and 99-502, respectively)] were not aimed specifically at genome or even biomedical research. However, such research and the surrounding commercial biotechnology enterprises clearly have benefited from them. The biotechnology sector's success owes much to federal policies on technology transfer and intellectual property. [Source: U.S. Congress, Office of Technology Assessment, *Federal Technology Transfer and the Human Genome Project*, OTA-BP-EHR-162 (Washington, D.C.: U.S. Government Printing Office, September 1995)]

filter hybridization and subsequent analysis. ["Work for Others" projects supported by an agency or organization other than DOE (e.g., NIH, National Cancer Institute, or a private company) can be conducted at a DOE installation because this work is complementary to DOE research missions and usually requires multidisciplinary DOE facilities and technologies.]

The Resource for Molecular Cytogenetics was established at LBNL and the University of California (UC), San Francisco, with the support of the Office of Biological and Environmental Research and Vysis, Inc. (formerly Imagenetics). The Resource aims to apply fluorescent in situ hybridization (FISH) techniques to genetic analysis of human tissue samples; produce probe reagents; design and develop digital-imaging microscopy; distribute probes, analysis technology, and educational materials in the molecular cytogenetic community; and transfer useful reagents, processes, and instruments to the private sector for commercialization.

NIST Advanced Technology Program

Several commercial applications of research sponsored by the U.S. Human Genome Project have been furthered by the Advanced Technology Program (ATP) of the U.S. National Institute of Standards and Technology. ATP's mission is to stimulate economic growth and industrial competitiveness by encouraging high-risk but powerful new technologies. Its Tools for DNA Diagnostics program uses collaborations among researchers and industry to develop (1) cost-effective methods for determining, analyzing, and storing DNA sequences for a wide variety of diagnostic applications ranging from healthcare to agriculture to the environment and (2) a new and potentially very large market for DNA diagnostic systems.

Awardees have included companies developing DNA diagnostic chips, more powerful cytogenetic diagnostic techniques based on comparative genomic hybridization, DNA sequencing instrumentation, and DNA analysis technology. Eventually, commercialization of these underlying technologies is expected to generate hundreds of thousands of jobs. *[*800/287-3863, Fax: 301/926-9524, *atp@micf.nist.gov*, *http://www.atp.nist.gov]*

Patenting and Licensing Highlights, FY 1994–96

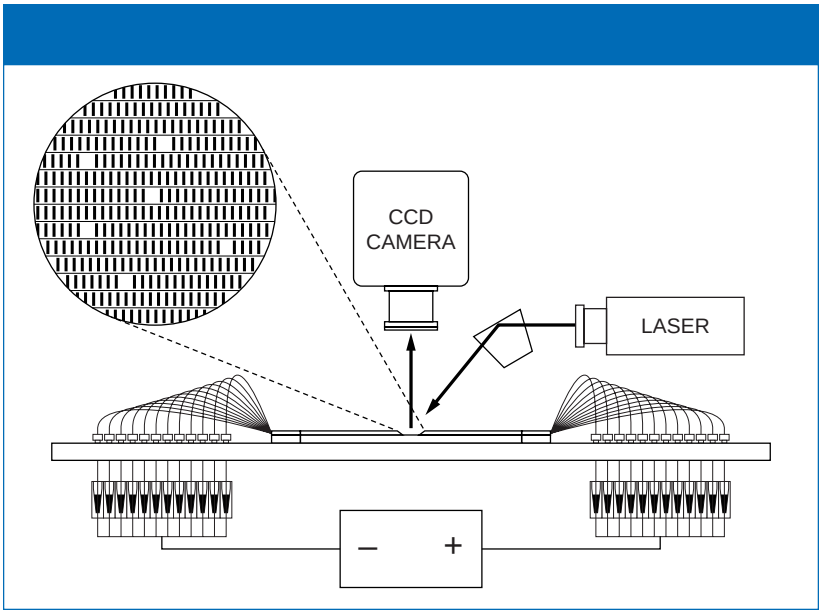
- A development license for single-molecule DNA sequencing replaced the 1991–94 CRADA (the first CRADA to be established in the U.S. Human Genome Project) between LANL and Life Technologies, Inc. (LTI).
- In 1995, a broad patent was awarded to UC for chromosome painting. This technology uses FISH to stain specific locations in cells and chromosomes for diagnosing, imaging, and studying chromosomal abnormalities and cancer. Resulting from a 1989 CRADA between LLNL and UC, FISH was licensed exclusively to Vysis.
- Hyseq, Inc., was founded in 1993 by former Argonne National Laboratory researchers Radoje Drmanac and Radomir Crkvenjakov to commercialize the sequencing by hybridization (SBH) technology. Hyseq has exclusive patent rights to a variation known as format 3 of SBH or the "super chip." Hyseq later won an Advanced Technology Program award from the U.S. National Institute of Standards and Technology to develop the technology further.
- Oligomers—short, single-stranded DNAs—are crucial reagents for genome research and biomedical diagnostics. ProtoGene Laboratories, Inc., was founded to commercialize new DNA synthesis technology (developed initially at LBNL with completed prototypes at Stanford University) and to offer the first lower-cost custom oligomer synthesis. The Parallel Array Synthesis system, which independently synthesizes 96 oligomers per run in a standard 96-well microtiter plate format, shows great promise for significant cost reductions. ProtoGene first

licensed sales and distribution to LTI and, later, production rights as well. LTI operates production centers in the United States, Europe, and Japan.

- The GRAIL-genQuest sequence-analysis software developed at ORNL was licensed by Martin Marietta Energy Systems (now Lockheed Martin Energy Research) to ApoCom, Inc., for pharmaceutical and biotechnology company researchers who cannot use the Internet because of data-security concerns. The public GRAIL-genQuest service remains freely available on the Internet (see box, p. 17).
- In 1995, an exclusive license was granted to U.S. Biochemical Corporation for a genetically engineered, heat-stable, DNA-replicating enzyme with much-improved sequencing properties. The enzyme was developed by Stanley Tabor at Harvard University Medical School.
- In 1995, an advanced capillary array electrophoresis system for sequencing DNA was patented by Iowa State University. The system was licensed to Premier American Technologies Corporation for commercialization (see graphic at right and R&D 100 Awards, next page).
- In 1996, a patent was granted to LANL researchers for DNA fragment sizing and sorting by laser-induced fluorescence. An exclusive license was awarded to Molecular Technology, Inc., for commercialization of the single-molecule detection capability related to DNA sizing (see R&D 100 Awards, next page).

SBIR and STTR

Small Business Innovation Research (SBIR) Program awards are designed to stimulate commercialization of new technology for the benefit of both the private and public sectors. The highly competitive program emphasizes cutting-edge, high-risk research with potential for high payoff in different areas, including human genome research. Small business firms with fewer than 500 employees are invited to submit applications. SBIR human genome topics concentrate on innovative and experimental approaches for carrying out the goals of the Human Genome Project (see SBIR, p. 63, in Part 2 of this report). The Small Business Technology Transfer (STTR) Program fosters transfers between research institutions and small businesses. *[*DOE SBIR and STTR contact: Kay Etzler (301/903-5867, Fax: -5488, *Kay.Etzler@oer.doe.gov*), *http://sbir.er.doe.gov/sbir, http://sttr.er.doe.gov/sttr]*



Capillary Array Electrophoresis (CAE). *CAE systems promise dramatically faster and higher-resolution fragment separation for DNA sequencing. A multiplexed CAE system designed by Edward Yeung (Iowa State University) has been developed for commercial production by Premier American Technologies Corporation (PATCO). In the PATCO ESY9600 model, DNA samples are introduced into the 96-capillary array; as the separated fragments pass through the capillaries, they are irradiated all at once with laser light. Fluorescence is measured by a charged coupled device that acts as a simultaneous multichannel detector. (Inset circle at upper left: Closeup view of individual capillary lanes with separated samples.) Because every fragment length exists in the sample, bases are identified in order according to the time required for them to reach the laser-detector region. [Source: Thomas Kane, PATCO]*

Technology Transfer Award

A Federal Laboratory Consortium Award for Excellence in Technology Transfer was presented to Edward Yeung and a research team at Iowa State University's Ames Laboratory in 1993. Their laser-based method for indirect fluorescence of biological samples may have applications for routine high-speed DNA sequencing (see graphic, p. 23). Yeung also won the 1994 American Chemical Society Award for Analytical Chemistry.

1997 R&D 100 Awards

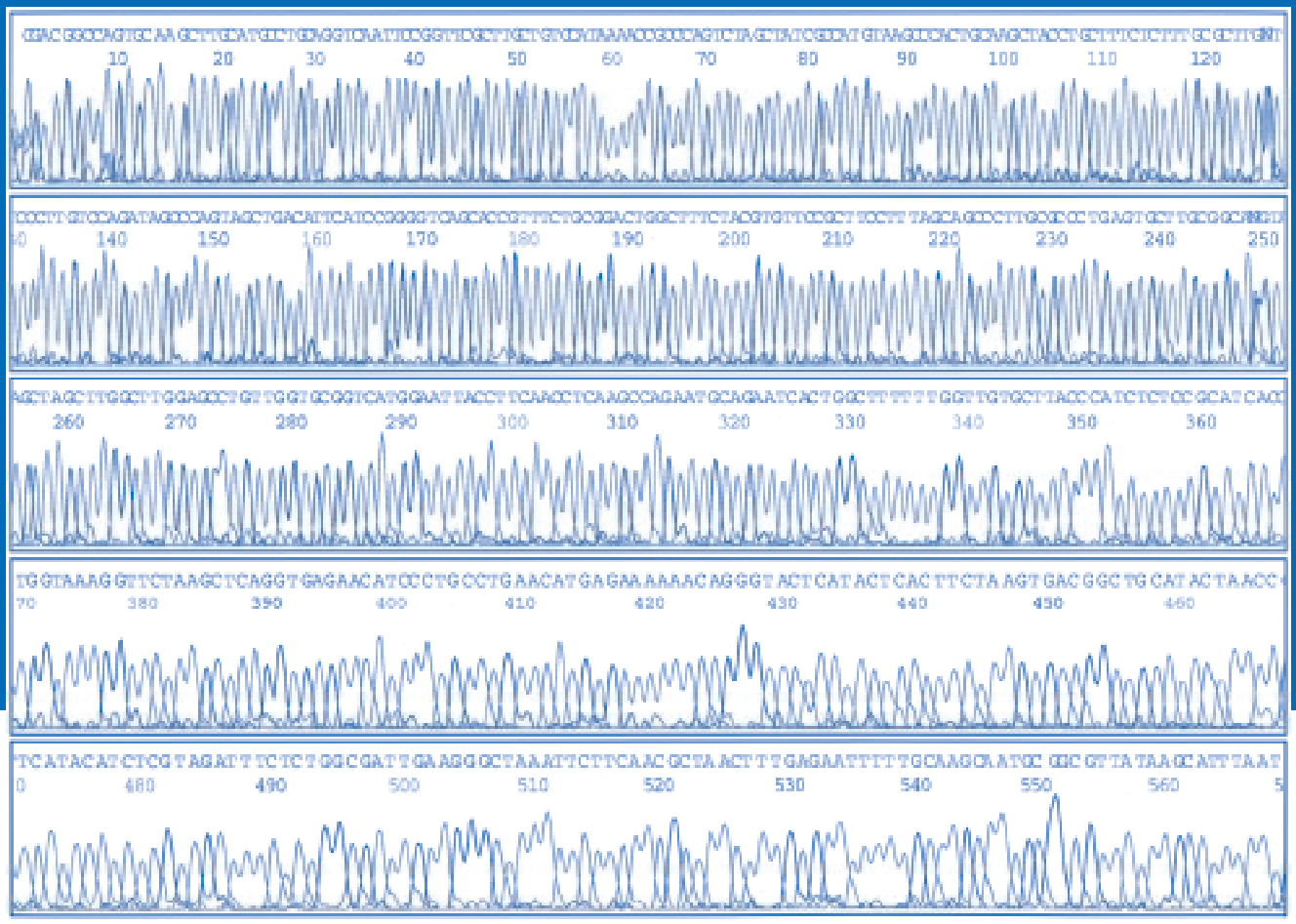
DOE researchers in 12 facilities across the country won 36 of the R&D 100 Awards given by *Research and Development Magazine* for 1996 work. DOE award-winning research ranged from advances in supercomputing to the biological recycling of tires. Announced in July 1997, these awards bring DOE's R&D 100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

Two DOE genome-related research projects received 1997 R&D 100 Awards. One was to Yeung (see text at left and graphic, p. 23) for "ESY9600 Multiplexed Capillary Electrophoresis DNA Sequencer."

The other award was to Richard Keller and James Jett (LANL) with Amy Gardner (Molecular Technologies, Inc.) for "Rapid-Size Analysis of Individual DNA Fragments." This technology speeds determination of DNA fragment sizes, making DNA fingerprinting applications in biotechnology and other fields more reliable and practical.

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

Readout from an automated DNA sequencing machine depicts the order of the four DNA bases (A, T, C, and G) in a DNA fragment of more than 500 bases. [Source: Linda Ashworth, LLNL]



Joint Genome Institute

DOE Merges Sequencing Efforts of Genome Centers

http://www.jgi.doe.gov

Elbert Branscomb
JGI Scientific Director
Lawrence Livermore
National Laboratory
7000 East Avenue, L-452
Livermore, CA 94551
510/422-5681
***elbert@alu.llnl.gov* or**
elbert@shotgun.llnl.gov

In a major restructuring of its Human Genome Program, on October 23, 1996, the DOE Office of Biological and Environmental Research established the Joint Genome Institute (JGI) to integrate work based at its three major human genome centers.

The JGI merger represents a shift toward large-scale sequencing via intensified collaborations for more effective use of the unique expertise and resources at Lawrence Berkeley National Laboratory (LBNL), Lawrence Livermore National Laboratory (LLNL), and Los Alamos National Laboratory (see Research Narratives, beginning on p. 27 in this report). Elbert Branscomb (LLNL) serves as JGI's Scientific Director. Capital equipment has been ordered, and operational support of about $30 million is projected for the 1998 fiscal year.

With easy access to both LBNL and LLNL, a building in Walnut Creek, California, is being modified. Here, starting in late FY 1998, production DNA sequencing will be carried out for JGI. Until that time, large-scale sequencing will continue at LANL, LBNL, and LLNL. Expectations are that within 3 to 4 years the Production Sequencing Facility will house some 200 researchers and technicians working on high-throughput DNA sequencing using state-of-the-art robotics.

Initial plans are to target gene-rich regions of around 1 to 10 megabases for sequencing. Considerations include gene density, gene families (especially clustered families), correlations to model organism results, technical capabilities, and relevance to the DOE mission (e.g., DNA repair, cancer susceptibility, and impact of genotoxins). The JGI program is subject to regular peer review.

Sequence data will be posted daily on the Web; as the information progresses to finished quality, it will be submitted to public databases.

As JGI and other investigators involved in the Human Genome Project are beginning to reveal the DNA sequence of the 3 billion base pairs in a reference human genome, the data already are becoming valuable reagents for explorations of DNA sequence function in the body, sometimes called "functional genomics." Although large-scale sequencing is JGI's major focus, another important goal will be to enrich the sequence data with information about its biological function. One measure of JGI's progress will be its success at working with other DOE laboratories, genome centers, and non-DOE academic and industrial collaborators. In this way, JGI's evolving capabilities can both serve and benefit from the widest array of partners.

Production DNA Sequencing Begun Worldwide

The year 1996 marked a transition to the final and most challenging phase of the U.S. Human Genome Project, as pilot programs aimed at refining large-scale sequencing strategies and resources were funded by DOE and NIH (see Research Highlights, DNA Sequencing, p. 14). Internationally, large-scale human genome sequencing was kicked off in late 1995 when The Wellcome Trust announced a 7-year, $75-million grant to the private Sanger Centre to scale up its sequencing capabilities. French investigators also have announced intentions to begin production sequencing.

Funding agencies worldwide agree that rapid and free release of data is critical. Other issues include sequence accuracy, types of annotation that will be most useful to biologists, and how to sustain the reference sequence.

The international Human Genome Organisation maintains a Web page to provide information on current and future sequencing projects and links to sites of participating groups (*http://hugo.gdb.org*). The site also links to reports and resources developed at the February 1996 and 1997 Bermuda meetings on large-scale human genome sequencing, which were sponsored by The Wellcome Trust.

Lawrence Livermore National Laboratory Human Genome Center

http://www-bio.llnl.gov/bbrp/genome/genome.html

The Human Genome Center at Lawrence Livermore National Laboratory (LLNL) was established by DOE in 1991. The center operates as a multidisciplinary team whose broad goal is understanding human genetic material. It brings together chemists, biologists, molecular biologists, physicists, mathematicians, computer scientists, and engineers in an interactive research environment focused on mapping, DNA sequencing, and characterizing the human genome.

Goals and Priorities

In the past 2 years, the center's goals have undergone an exciting evolution. This change is the result of several factors, both intrinsic and extrinsic to the Human Genome Project. They include: (1) successful completion of the center's first-phase goal, namely a high-resolution, sequence-ready map of human chromosome 19; (2) advances in DNA sequencing that allow accelerated scaleup of this operation; and (3) development of a strategic plan for LLNL's Biology and Biotechnology Research Program that will integrate the center's resources and strengths in genomics with programs in structural biology, individual susceptibility, medical biotechnology, and microbial biotechnology.

The primary goal of LLNL's Human Genome Center is to characterize the mammalian genome at optimal resolution and to provide information and material resources to other in-house or collaborative projects that allow exploitation of genomic biology in a synergistic manner. DNA sequence information provides the biological driver for the center's priorities:

- Generation of highly accurate sequence for chromosome 19.
- Generation of highly accurate sequence for genomic regions of high biological interest to the mission of the DOE Office of Biological and Environmental Research (e.g., genes involved in DNA repair, replication, recombination, xenobiotic metabolism, and cell-cycle control).
- Isolation and sequence of the full insert of cDNA clones associated with genomic regions being sequenced.
- Sequence of selected corresponding regions of the mouse genome in parallel with the human.
- Annotation and position of the sequenced clones with physical landmarks such as linkage markers and sequence tagged sites (STSs).
- Generation of mapped chromosome 19 and other genomic clones [cosmids, bacterial artificial chromosomes (BACs), and P1 artificial chromosomes (PACs)] for collaborating groups.
- Sharing of technology with other groups to minimize duplication of effort.
- Support of downstream biology projects, for example, structural biology, functional studies, human variation, transgenics, medical biotechnology, and microbial biotechnology with know-how, technology, and material resources.

Center Organization and Activities

Completion and publication of the metric physical map of human chromosome 19 (see p. 28) in 1995 has led to consolidation of many functions associated with physical mapping, with increased emphasis on DNA sequencing. The center is organized into five broad areas of research and support: sequencing, resources, functional genomics, informatics and analytical genomics, and instrumentation. Each area consists of multiple projects, and extensive interaction occurs both within and among projects.

Human Genome Center
Lawrence Livermore National Laboratory
Biology and Biotechnology Research Program
7000 East Avenue, L-452
Livermore, CA 94551

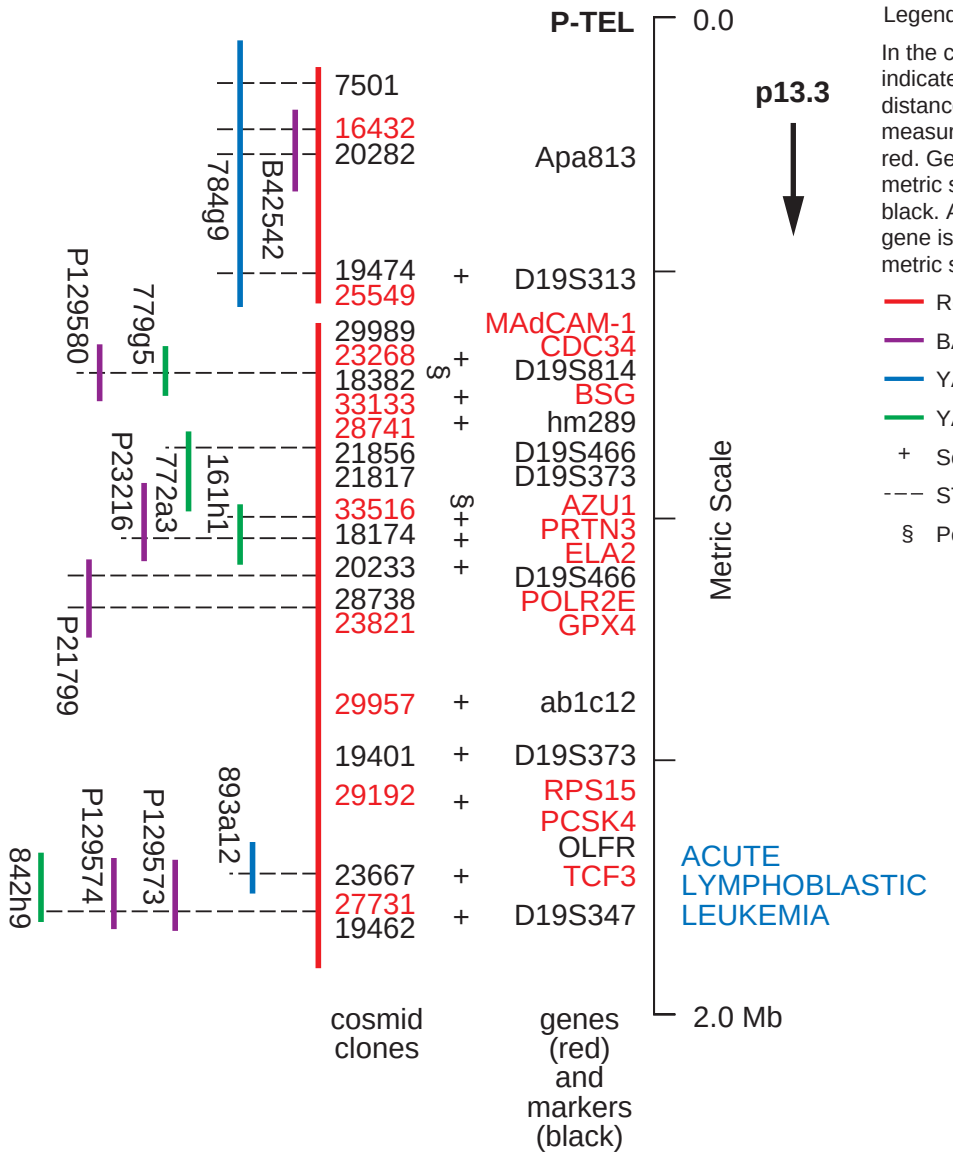
Anthony V. Carrano
Director
510/422-5698, Fax: /423-3110
carrano1@llnl.gov

Linda Ashworth
Assistant to Center Director
510/422-5665, Fax: -2282
ashworth1@llnl.gov

In lieu of individual abstracts, research projects and investigators at the LLNL Human Genome Center are represented in this narrative. More information can be found on the center's Web site (see URL above).

Update

In 1997 Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, and Los Alamos National Laboratory began collaborating in a Joint Genome Institute to implement high-throughput sequencing [see p. 26 and *Human Genome News* **8**(2), 1–2].



Legend

In the column labeled cosmid clones, black indicates a FISH-ordered clone where distance between clones has been measured. Other cosmids are shown in red. Genes are in red to the left of the metric scale. Other markers are labeled in black. A disease associated with a specific gene is shown in blue to the right of the metric scale.

- Restriction-mapped contig
- BAC, PAC, or P1 clone
- YAC with known and concordant size
- YAC with unknown or discordant size
- \+ Sequence tagged site (STS)
- STS and/or hybridization results
- § Polymorphic marker

Chromosome 19 Map. *In the current map (at left) of the first 2 million bases at the p-telomere end of chromosome 19, the* EcoR I *restriction-mapped contigs (represented by red lines) provide the starting material for genomic sequencing across a region.*

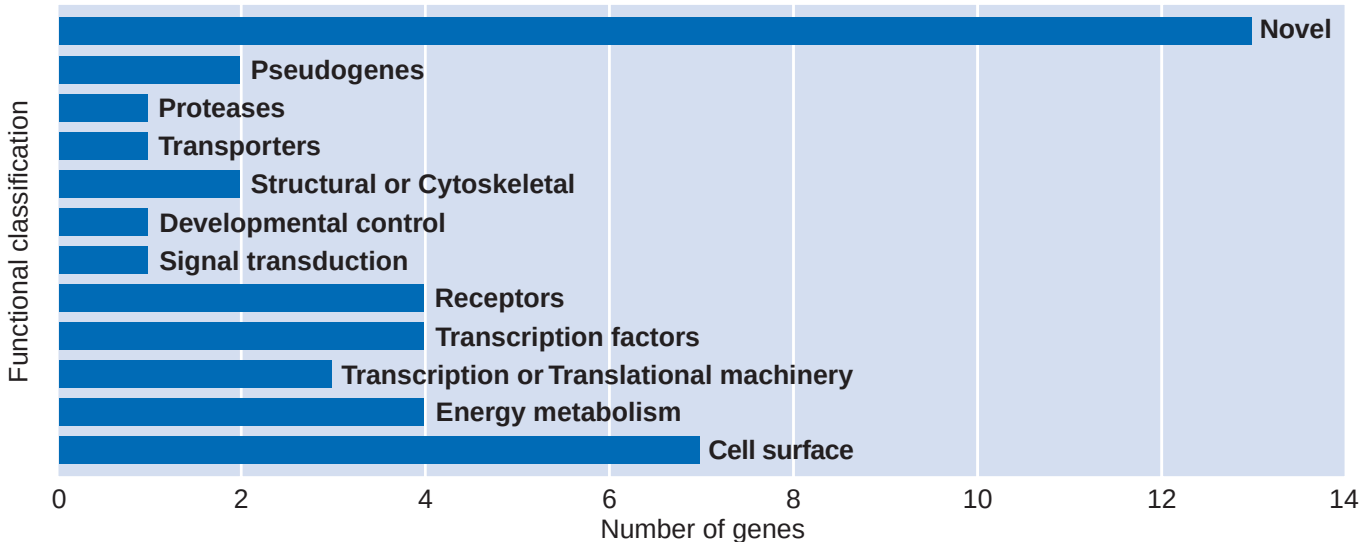
Construction of the human chromosome 19 physical map was based on a similar strategy for mapping the roundworm Caenorhabditis elegans. *View the complete map on the World Wide Web (*http://www-bio.llnl.gov/genome/html/chrom_map.html). *[Source: Adapted from figure provided by Linda Ashworth, LLNL]*

Sequencing

The sequencing group is divided into several subprojects. The core team is responsible for the construction of sequence libraries, sequencing reactions, and data collection for all templates in the random phase of sequencing. The finishing team works with data produced by the core team to produce highly redundant, highly accurate "finish" sequence on targets of interest. Finally, a team of researchers focuses specifically on development, testing, and implementation of new protocols for the entire group, with an emphasis on improving the efficiency and cost basis of the sequencing operation.

Resources

The resources group provides mapped clonal resources to the sequencing teams. This group performs physical mapping as needed for the DNA sequencing group by using fingerprinting, restriction mapping, fluorescence in situ hybridization, and other techniques. A small mapping effort is under way to identify, isolate, and characterize BAC



Putative-Gene Classification. *The figure depicts the functional classification of putative genes identified in a 1.02-Mb region on the long arm of human chromosome 19. Analysis of the completed sequence between markers D19S208 and COX7A1 revealed 43 open reading frames (ORFs) or putative genes. (An ORF is a DNA region containing specific sequences that signal the beginning and ending of a gene.)*

Thirty of these putative genes were found to have sequence similarities to a wide variety of known genes or proteins, including some involved in transcription, cell adhesion and signaling, and metabolism. Many appear to be related functionally to such known proteins as the GTP-ase activating proteins or the ETS family of transcription factors. Others seem to be new members of existing gene families, for example, the mRNA splicing factor, or of such pseudogenes as the elongation factor Tu.

In addition to those that could be classified, 13 novel genes were identified, including one with high similarity to a predicted ORF of unknown function in the roundworm Caenorhabditis elegans**.** *[Source: Adapted from graph provided by Linda Ashworth, LLNL]*

clones (from anywhere in the human genome) that relate to susceptibility genes, for example, DNA repair. These clones will be characterized and provided for sequencing and at the same time contribute to understanding the biology of the chromosome, the genome, and susceptibility factors. The mapping team also collaborates with others using the chromosome 19 map as a resource for gene hunting.

Functional Genomics

The functional genomics team is responsible for assembling and characterizing clones for the Integrated Molecular Analysis of Gene Expression (called IMAGE) Consortium and cDNA sequencing, as well as for work on gene expression and comparative mouse genomics. The effort emphasizes genes involved in DNA repair and links strongly to LLNL's gene-expression and structural biology efforts. In addition, this team is working closely with Oak Ridge National Laboratory (ORNL) to develop a comparative map and the sequence data for mouse regions syntenic to human chromosome 19 (see p. 32).

Informatics and Analytical Genomics

The informatics and analytical genomics group provides computer science support to biologists. The sequencing informatics team works directly with the DNA sequencing group to facilitate and automate sample handing, data acquisition and storage, and DNA sequence analysis and annotation. The

analytical genomics team provides statistical and advanced algorithmic expertise. Tasks include development of model-based methods for data capture, signal processing, and feature extraction for DNA sequence and fingerprinting data and analysis of the effectiveness of newly proposed methods for sequencing and mapping.

Instrumentation

The instrumentation group also has multiple components. Group members provide expertise in instrumentation and automation in high-throughput electrophoresis, preparation of high-density replicate DNA and colony filters, fluorescence labeling technologies, and automated sample handling for DNA sequencing. To facilitate seamless integration of new technologies into production use, this group is coupled tightly to the biologist user groups and the informatics group.

Collaborations

The center interacts extensively with other efforts within the LLNL Biology and Biotechnology Research Program and with other programs at LLNL, the academic community, other research institutes, and industry. More than 250 collaborations range from simple probe and clone sharing to detailed gene family studies. The following list reflects some major collaborations.

- Integration of the genetic map of human chromosome 19 with corresponding mouse chromosomes (ORNL).
- Miniaturized polymerase chain reaction instrumentation (LLNL).
- Sequencing of IMAGE Consortium cDNA clones (Washington University, St. Louis).
- Mapping and sequencing of a gene associated with Finnish congenital nephrotic syndrome (University of Oulu, Finland).

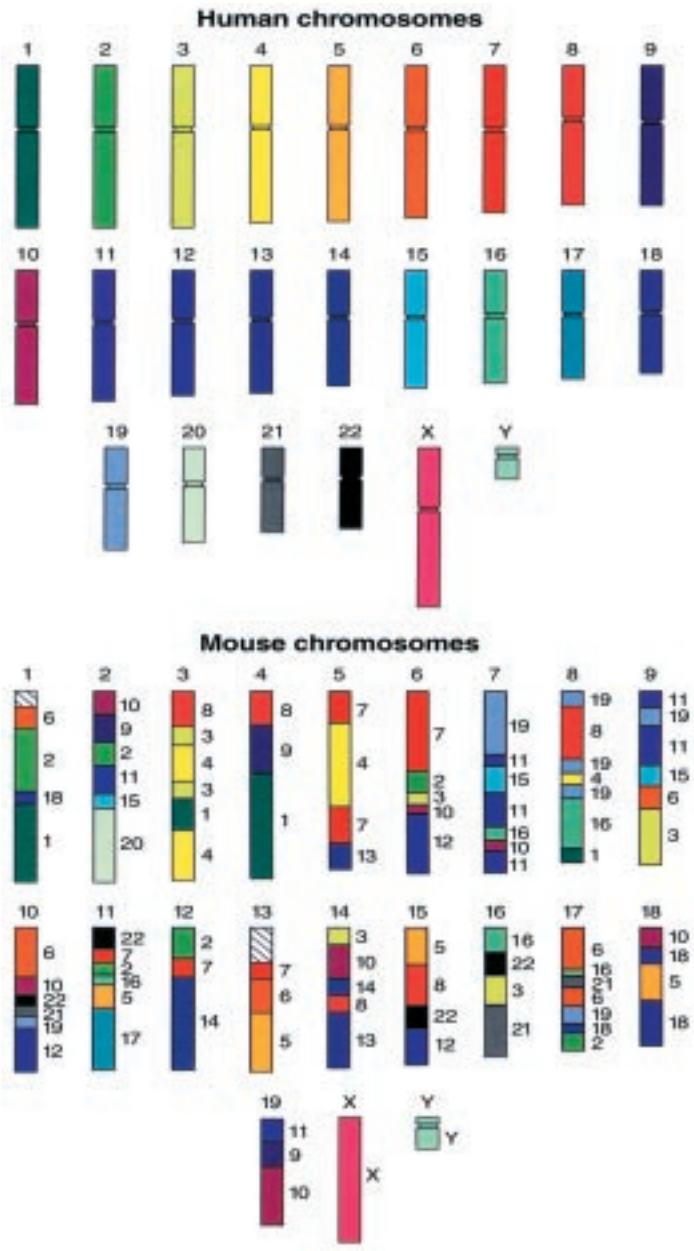
Accomplishments

The LLNL Human Genome Center has excelled in several areas, including comparative genomic sequencing of DNA repair genes in human and rodent species, construction of a metric physical map of human chromosome 19, and development and application of new biochemical and mathematical approaches for constructing ordered clone maps. These and other major accomplishments are highlighted below.

- Completion of highly accurate sequencing totaling 1.6 million bases of DNA, including regions spanning human DNA repair genes, the candidate region for a congenital kidney disease gene, and other regions of biological interest on chromosome 19.
- Completion of comparative sequence analysis of 107,500 bases of genomic DNA encompassing the human DNA repair gene *ERCC*2 and the corresponding regions in mouse and hamster (p. 32). In addition to *ERCC*2, analysis revealed the presence of two previously undescribed genes in all three species. One of these genes is a new member of the kinesin motor protein family. These proteins play a wide variety of roles in the cell, including movement of chromosomes before cell division.
- Complete sequencing of human genomic regions containing two additional DNA repair genes. One of these, *XRCC*3, maps to human chromosome 14 and encodes a protein that may be required for chromosome stability. Analysis of the genomic sequence identified another kinesin motor protein gene physically linked to *XRCC*3. The second human repair gene, *HHR*23A, maps to 19p13.2. Sequence analysis of 110,000 bases containing *HHR*23A identified six other genes, five of which are new genes with similarity

to proteins from mouse, human, yeast, and *Caenorhabditis elegans*.

- Complete sequencing of full-length cDNAs for three new DNA repair genes (*XRCC*2, *XRCC*3, and *XRCC*9) in collaboration with the LLNL DNA repair group.
- Generation of a metric physical map of chromosome 19 spanning at least 95% of the chromosome. This unique map incorporates a metric scale to estimate the distance between genes or other markers of interest to the genetics community.
- Assembly of nearly 45 million bases of *Eco*R I restriction-mapped cosmid contigs for human chromosome 19 using a combination of fingerprinting and cosmid walking. Small gaps in cosmid continuity have been spanned by BAC, PAC, and P1 clones, which are then integrated into the restriction maps. The high depth of coverage of these maps (average redundancy, 4.3-fold) permits selection of a minimum overlapping set of clones for DNA sequencing.
- Placement of more than 400 genes, genetic markers, and other loci on the chromosome 19 cosmid map. Also, 165 new STSs associated with pre-mapped cosmid contigs were generated and added to the physical map.
- Collaborations to identify the gene (*COMP*) responsible for two allelic genetic diseases, pseudoachondroplasia and multiple epiphyseal dysplasia, and the identification of specific mutations causing each condition.
- Through sequence analysis of the 2A subfamily of the human cytochrome P450 enzymes, identification of a new variant that exists in 10% to 20% of individuals and results in reduced ability to metabolize nicotine and the antiblood-clotting drug Coumadin.
- Location of a zinc finger gene that encodes a transcription factor regulating blood-cell development adjacent to telomere repeat sequences, possibly the gene nearest one end of chromosome 19.
- Completion of the genomic and cDNA sequence of the gene for the human Rieske Fe-S protein involved in mitochondrial respiration.
- Expansion of the mouse-human comparative genomics collaboration with ORNL to include study of new groups of clustered transcription factors found on human chromosome 19q and as syntenic homologs on mouse chromosome 7 (p. 32).
- Numerous collaborations (in particular, with Washington University and Merck) continuing to expand the LLNL-based IMAGE Consortium, an effort to characterize the transcribed human genome. The IMAGE clone collection is now the largest public collection of sequenced cDNA clones, with more than one million arrayed clones, 800,000 sequences in public databases, and 10,000 mapped cDNAs.
- Development and deployment of a comprehensive system to handle sample tracking needs of production DNA sequencing. The system combines databases and graphical interfaces running on both Mac and Sun platforms and scales easily to handle large-scale production sequencing.
- Expansion of the LLNL genome center's World Wide Web site to include tables that link to each gene being sequenced, to the quality scores and assembled bases collected each night during the sequencing process, and to the submitted GenBank sequence when a clone is completed. *[http://bbrp.llnl.gov/test-bin/projqcsummary]*

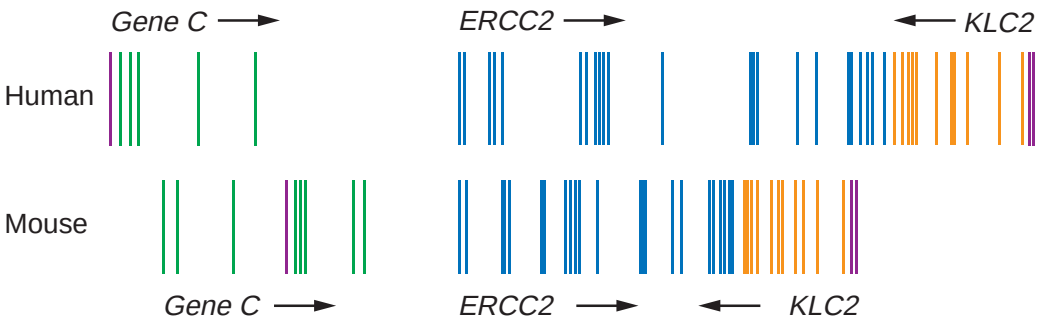


Human-Mouse Homologies. *LLNL researcher Lisa Stubbs (above) is shown in the Mouse Genetics Research Facility at ORNL.* *[ORNL photo]*

The figure at left demonstrates the genetic similarity (homology) of the superficially dissimilar mouse and human species. The similarity is such that human chromosomes can be cut (schematically at least) into about 150 pieces (only about 100 are large enough to appear here), then reassembled into a reasonable approximation of the mouse genome. The colors and corresponding numbers on the mouse chromosomes indicate the human chromosomes containing homologous segments. *[Source: Lisa Stubbs, LLNL]*

Comparative sequencing of homologous regions in human and mouse at LLNL has enhanced the ability to identify protein-coding (exon) and noncoding DNA regions that have remained unchanged over the course of evolution. Colors in the figure below depict similarities in mouse and human genes involved in DNA repair, a research interest rooted in DOE's mission to develop better technologies for measuring health effects, particularly mutations. *[Source: Linda Ashworth, LLNL]*

ERCC2 Region



- Implementation of a new database to support sequencing and mapping work on multiple chromosomes and species. Web-based automated tools were developed to facilitate construction of this database, the loading of over 100 million bytes of chromosome 19 data from the existing LLNL database, and automated generation of Web-based input interfaces.
- Significant enhancement of the LLNL Genome Graphical Database Browser software to display and link information obtained at a subcosmid resolution from both restriction map hybridization and sequence feature data. Features, such as genes linked to diseases, allow tracking to fragments as small as 500 base pairs of DNA.
- Development of advanced microfabrication technologies to produce electrophoresis microchannels in large glass substrates for use in DNA sequencing.
- Installation of a new filter-spotting robot that routinely produces $6 \times 6 \times 384$ filters. A $16 \times 16 \times 384$ pattern has been achieved.
- Upgrade of the Lawrence Berkeley National Laboratory colony picker using a second computer so that imaging and picking can occur simultaneously.

Future Plans

Genomic sequencing currently is the dominant function of Livermore's Human Genome Center. The physical mapping effort will ensure an ample supply of sequence-ready clones. For sequencing targets on chromosome 19, this includes ensuring that the most stable clones (cosmids, BACs, and PACs) are available for sequencing and that regions with such known physical landmarks as STSs and expressed sequenced tags (ESTs) are annotated to facilitate sequence assembly and analysis. The following targets are emphasized for DNA sequencing:

- Regions of high gene density, including regions containing gene families.
- Chromosome 19, of which at least 42 million bases are sequence ready.
- Selected BAC and PAC clones representing regions of about 0.2 million to 1 million bases throughout the human genome; clones would be selected based on such high-priority biological targets as genes involved in DNA repair, replication, recombination, xenobiotic metabolism, cell-cycle checkpoints, or other specific targets of interest.
- Selected BAC and PAC clones from mouse regions syntenic with the genes indicated above.
- Full-insert cDNAs corresponding to the genomic DNA being sequenced.

The informatics team is continuing to deploy broader-based supporting databases for both mapping and sequencing. Where appropriate, Web- and Java-based tools are being developed to enable biologists to interact with data. Recent reorganization within this group enables better direct support to the sequencing group, including evaluating and interfacing sequence-assembly algorithms and analysis tools, data and process tracking, and other informatics functions that will streamline the sequencing process.

The instrumentation effort has three major thrusts: (1) continued development or implementation of laboratory automation to support high-throughput sequencing; (2) development of the next-generation DNA sequencer; and (3) development of robotics to support high-density BAC clone screening. The last two goals warrant further explanation.

The new DNA sequencer being developed under a grant from the National Institutes of Health, with minor support

through the DOE genome center, is designed to run 384 lanes simultaneously with a low-viscosity sieving medium. The entire system would be loaded automatically, run, and set up for the next run at 3-hour intervals. If successful, it should provide a 20- to 40-fold increase in throughput over existing machines.

An LLNL-designed high-precision spotting robot, which should allow a density of 98,304 spots in 96 cm^2, is now operating. The goal of this effort is to create high-density filters representing a 10× BAC coverage of both human and mouse genomes (30,000 clones = 1× coverage). Thus each filter would provide ~3× coverage, and eight such filters would provide the desired coverage for both genomes. The filters would be hybridized with amplicons from individual or region-specific cDNAs and ESTs; given the density of the BAC libraries, clones that hybridize should represent a binned set of BACs for a region of interest. These BACs could be the initial substrate for a BAC sequencing strategy. Performing hybridizations in parallel in mouse and human DNA facilitates the development of the mouse map (with ORNL involvement), and sequencing BACs from both species identifies evolutionarily conserved and, perhaps, regulatory regions.

Information generated by sequencing human and mouse DNA in parallel is expected to expand LLNL efforts in functional genomics. Comparative sequence data will be used to develop a high-resolution synteny map of conserved mouse-human domains and incorporate automated northern expression analysis of newly identified genes. Long range, the center hopes to take advantage of a variety of forms of expression analysis, including site-directed mutation analysis in the mouse.

Summary

The Livermore Human Genome Center has undergone a dramatic shift in emphasis toward commitment to large-scale, high-accuracy sequencing of chromosome 19, other chromosomes, and targeted genomic regions in the human and mouse. The center also is committed to exploiting sequence information for functional genomics studies and for other programs, both in house and collaboratively.

Los Alamos National Laboratory Center for Human Genome Studies

http://www-ls.lanl.gov/masterhgp.html

Center for Human Genome Studies
Los Alamos National Laboratory
P.O. Box 1663
Los Alamos, NM 87545

Larry L. Deaven
Acting Director
505/667-3912, Fax: -2891
ldeaven@telomere.lanl.gov

Lynn Clark
Technical Coordinator
505/667-9376, Fax: -2891
clark@telomere.lanl.gov

Robert K. Moyzis
Director, 1989–97*

Biological research was initiated at Los Alamos National Laboratory (LANL) in the 1940s, when the laboratory began to investigate the physiological and genetic consequences of radiation exposure. Eventual establishment of the national genetic sequence databank called GenBank, the National Flow Cytometry Resource, numerous related individual research projects, and fulfillment of a key role in the National Laboratory Gene Library Project all contributed to LANL's selection as the site for the Center for Human Genome Studies in 1988.

Center Organization and Activities

The LANL genome center is organized into four broad areas of research and support: Physical Mapping, DNA Sequencing, Technology Development, and Biological Interfaces. Each area consists of a variety of projects, and work is distributed among five LANL Divisions (Life Sciences; Theoretical; Computing, Information, and Communications; Chemical Science and Technology; and Engineering Sciences and Applications). Extensive interdisciplinary interactions are encouraged.

Physical Mapping

The construction of chromosome- and region-specific cosmid, bacterial artificial chromosome (BAC), and yeast artificial chromosome (YAC) recombinant DNA libraries is a primary focus of physical mapping activities at LANL. Specific work includes the construction of high-resolution maps of human chromosomes 5 and 16 and associated informatics and gene discovery tasks.

Accomplishments

- Completion of an integrated physical map of human chromosome 16 consisting of both a low-resolution YAC contig map and a high-resolution cosmid contig map (pp. 37–39). With sequence tagged site (STS) markers provided on average every 125,000 bases, the YAC-STS map provides almost-complete coverage of the chromosome's euchromatic arms. All available loci continue to be incorporated into the map.
- Construction of a low-resolution STS map of human chromosome 5 consisting of 517 STS markers regionally assigned by somatic-cell hybrid approaches. Around 95% mega-YAC–STS coverage (50 million bases) of 5p has been achieved. Additionally, about 40 million bases of 5q mega-YAC–STS coverage have been obtained collaboratively.
- Refinement of BAC cloning procedures for future production of chromosome-specific libraries. Successful partial digestion and cloning of microgram quantities of chromosomal DNA embedded in agarose plugs. Efforts continue to increase the average insert size to about 100,000 bases.

DNA Sequencing

DNA sequencing at the LANL center focuses on low-pass sample sequencing (SASE) of large genomic regions. SASE data is deposited in publicly available databases to allow for wide distribution. Finished sequencing is prioritized from initial SASE analysis and pursued by parallel primer walking. Informatics development includes data tracking, gene-discovery integration with the Sequence Comparison ANalysis (SCAN) program, and functional genomics interaction.

Accomplishments

- SASE sequencing of 1.5 million bases from the p13 region of human chromosome 16.
- Discovery of more than 100 genes in SASE sequences.

In lieu of individual abstracts, research projects and investigators at the LANL Center for Human Genome Studies are represented in this narrative. More information can be found on the center's Web site (see URL above).

Update

In 1997 Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, and Los Alamos National Laboratory began collaborating in a Joint Genome Institute to implement high-throughput sequencing [see p. 26 and *Human Genome News* **8**(2), 1–2].

*Now at University of California, Irvine

- Generation of finished sequence for a 240,000-base telomeric region of human chromosome 7q. From initial sequences generated by SASE, oligonucleotides were synthesized and used for primer walking directly from cosmids comprising the contig map. Complete sequencing was performed to determine what genes, if any, are near the 7q terminus. This intriguing region lacks significant blocks of subtelomeric repeat DNA typically present near eukaryotic telomeres.
- Complete single-pass sequencing of 2018 exon clones generated from LANL's flow-sorted human chromosome 16 cosmid library. About 950 discrete sequences were identified by sequence analysis. Nearly 800 appear to represent expressed sequences from chromosome 16.
- Development of Sequence Viewer to display ABI sequences with trace data on any computer having an Internet connection and a Netscape World Wide Web browser.
- Sequencing and analysis of a novel pericentromeric duplication of a gene-rich cluster between 16p11.1 and Xq28 (in collaboration with Baylor College of Medicine).

Technology Development

Technology development encompasses a variety of activities, both short and long term, including novel vectors for library construction and physical mapping; automation and robotics tools for physical mapping and sequencing; novel approaches to DNA sequencing involving single-molecule detection; and novel approaches to informatics tools for gene identification.

Accomplishments

- Development of SCAN program for large-scale sequence analysis and annotation, including a translator converting SCAN data to GIO format for submission to Genome Sequence DataBase.
- Application of flow-cytometric approach to DNA sizing of P1 artificial chromosome (PAC) clones. Less than one picogram of linear or supercoiled DNA is analyzed in under 3 minutes. Sizing range has been extended down to 287 base pairs. Efforts continue to extend the upper limit beyond 167,000 bases.
- Characterization of the detection of single, fluorescently tagged nucleotides cleaved from multiple DNA fragments suspended in the flow stream of a flow cytometer (see picture, p. 70). The cleavage rate for Exo III at 37°C was measured to be about 5 base pairs per second per M13 DNA fragment. To achieve a single-color sequencing demonstration, either the background burst rate (currently about 5 bursts per second) must be reduced or the exonuclease cleavage rate must be increased significantly. Techniques to achieve both are being explored.
- Construction of a simple and compact apparatus, based on a diode-pumped Nd:YAG laser, for routine DNA fragment sizing.
- Development of a new approach to detect coding sequences in DNA. This complete spectral analysis of coding and noncoding sequences is as sensitive in its first implementations as the best existing techniques.
- Use of phylogenetic relationships to generate new profiles of amino acid usage in conserved domains. The profiles are particularly useful for classification of distantly related sequences.

Biological Interfaces

The Biological Interfaces effort targets genes and chromosome regions associated with DNA damage and repair, mitotic stability, and chromosome structure and function as primary subjects for physical mapping and sequencing. Specific disease-associated genes on human chromosome 5 (e.g., Cri-du-Chat syndrome) and on 16 (e.g., Batten's disease and Fanconi anemia) are the subjects of collaborative biological projects.

Accomplishments

- Identification of two human 7q exons having 99% homology to the cDNA of a known human gene, vasoactive intestinal peptide receptor 2A. Preliminary data suggests that the *VIPR*2A gene is expressed.
- Identification of numerous expressed sequence tags (ESTs) localized to the 7q region. Since three of the ESTs contain at least two regions with high confidence of homology (~90%), genes in addition to *VIPR*2A may exist in the terminal region of 7q.
- Generation of high-resolution cosmid coverage on human chromosome 5p for the larynx and critical regions identified with Cri-du-Chat syndrome, the most common human terminal-deletion syndrome (in collaboration with Thomas Jefferson University).
- Refinement of the Wolf-Hirschhorn syndrome (WHS) critical region on human chromosome 4p. Using the SCAN program to identify genes likely to contribute to WHS, the project serves as a model for defining the interaction between genomic sequencing and clinical research.
- Collaborative construction of contigs for human chromosome 16, including 1.05 million bases in cosmids through the familial Mediterranean fever (FMF) gene region (with members of the FMF Consortium) and 700,000 bases in P1 clones encompassing the polycystic kidney disease gene (with Integrated Genetics, Inc.).
- Collaborative identification and determination of the complete genomic structure of the Batten's disease gene (with members of the BDG Consortium), the gamma subunit of the human amiloride-sensitive epithelial channel (Liddle's syndrome, with University of Iowa), and the polycystic kidney disease gene (with Integrated Genetics).
- Participation in an international collaborative research consortium that successfully identified the gene responsible for Fanconi anemia type A.

Chromosome 16 Physical Map (pp. 38–39). *A condensed chromosome 16 physical map constructed at Los Alamos National Laboratory (LANL) is shown in two parts on the following pages. Besides facilitating the isolation and characterization of disease genes, the map provides the framework for a large-scale sequencing effort by LANL, The Institute for Genomic Research, and the Sanger Centre.*

Distinct types of maps and data are shown as levels or tiers on the integrated map. At the top of each page is a view of the banded human chromosome to which the map is aligned. A somatic-cell hybrid breakpoint map, which divides the chromosome into 90 intervals, was used as a backbone for much of the map integration.

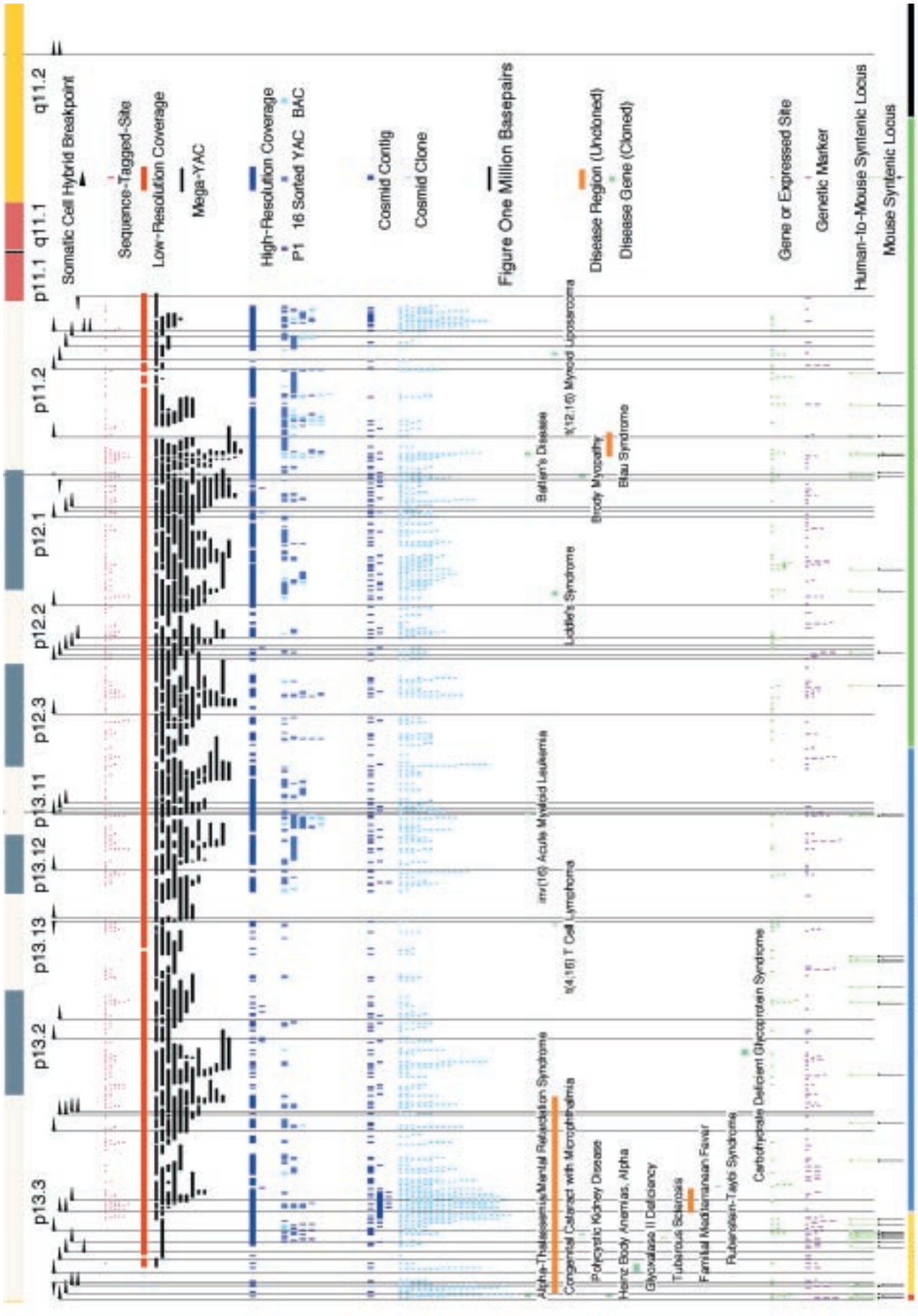
The physical map consists of both a low-resolution yeast artificial chromosome (YAC) contig map localized to and ordered within the breakpoint intervals with sequence tagged sites (STSs) and a high-resolution bacteria-based clone map. The YAC-STS map provides almost complete coverage of the chromosome's euchromatic arm, with STS markers on average every 100,000 bases.

A high-resolution, sequence-ready cosmid contig map is anchored to the YAC and breakpoint maps via STSs developed from cosmid contigs and by hybridizations between YACs and cosmids.

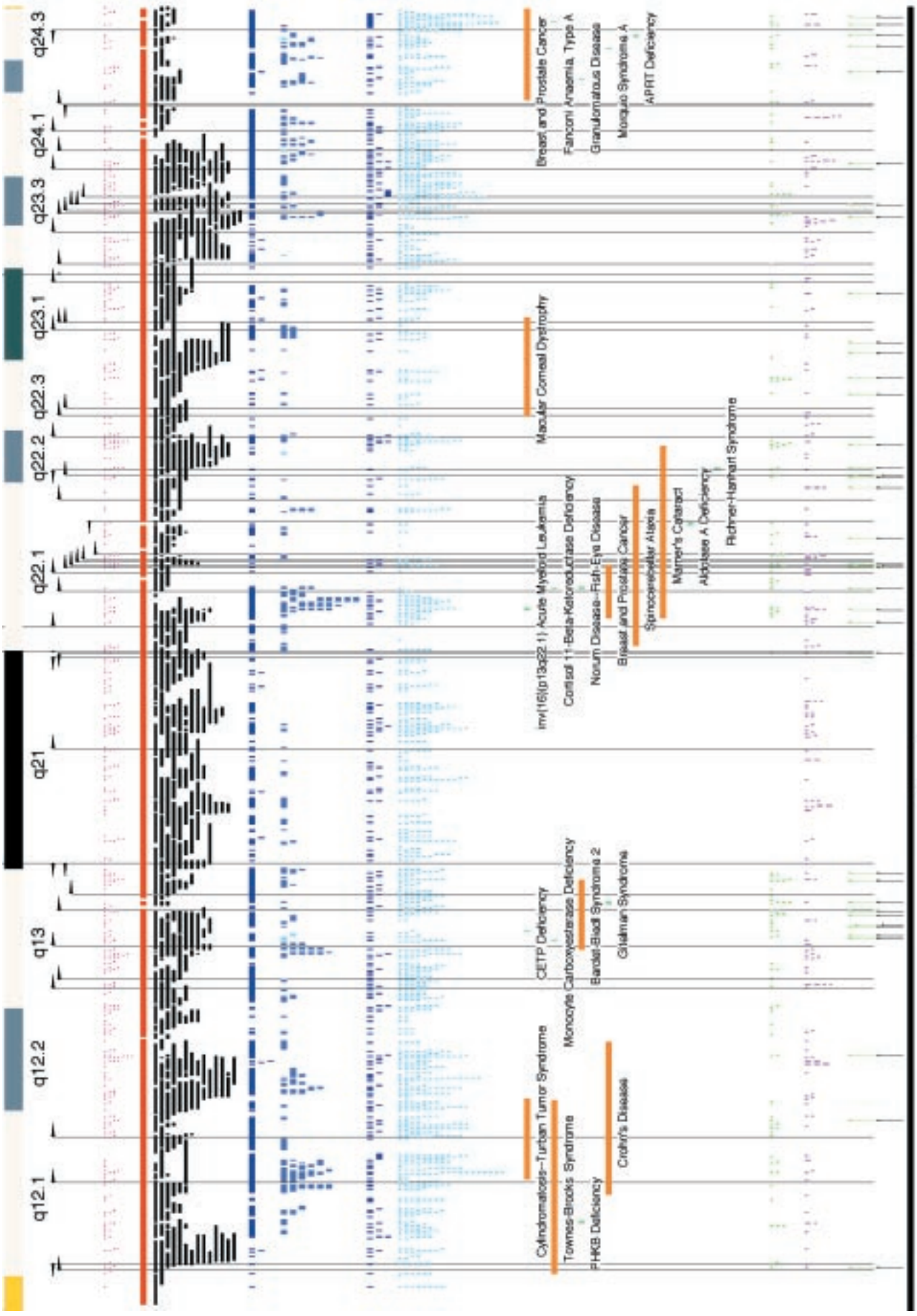
As part of the ongoing effort to incorporate all available loci onto a single map of this chromosome, the integrated map also features genes, expressed sequence tags, exons (gene-coding regions), and genetic markers.

The mouse chromosome segments at the bottom of the map contain groups that correspond to human genes mapped to the regions shown above them. *[Source: Norman Doggett, LANL]*

Human Chromosome 16



Human Chromosome 16



The exhibit "Understanding Our Genetic Inheritance" at the Bradbury Science Museum in Los Alamos, New Mexico, describes the LANL Center for Human Genome Studies' contributions to the Human Genome Project. The exhibit's centerpiece is a 16-foot-long version of LANL's map of human chromosome 16. [Source: LANL Center for Human Genome Studies]

Patents, Licenses, and CRADAs

- Rhett L. Affleck, James N. Demas, Peter M. Goodwin, Jay A. Schecker, Ming Wu, and Richard A. Keller, "Reduction of Diffusional Defocusing in Hydrodynamically Focused Flows by Complexing with a High Molecular Weight Adduct," United States Patent, filed December 1996.
- R.L. Affleck, W.P. Ambrose, J.D. Demas, P.M. Goodwin, M.E. Johnson, R.A. Keller, J.T. Petty, J.A. Schecker, and M. Wu, "Photobleaching to Reduce or Eliminate Luminescent Impurities for Ultrasensitive Luminescence Analysis," United States Patent, S-87, 208, accepted September 1997.
- J.H. Jett, M.L. Hammond, R.A. Keller, B.L. Marrone, and J.C. Martin, "DNA Fragment Sizing and Sorting by Laser-Induced Fluorescence," United States Patent, S.N. 75,001, allowed May 1996.
- James H. Jett, "Method for Rapid Base Sequencing in DNA and RNA with Three Base Labeling," in preparation.
- Development license and exclusive license to LANL's DNA sizing patent obtained by Molecular Technology, Inc., for commercialization of single-molecule detection capability to DNA sizing.

Future Plans

LANL has joined a collaboration with California Institute of Technology and The Institute for Genomic Research to construct a BAC map of the *p* arm of human chromosome 16 and to complete the sequence of a 20-million–base region of this map.

In its evolving role as part of the new DOE Joint Genome Institute, LANL will continue scaleup activities focused on high-throughput DNA sequencing. Initial targets include genes and DNA regions associated with chromosome structure and function, syntenic breakpoints, and relevant disease-gene loci.

A joint DNA sequencing center was established recently by LANL at the University of New Mexico. This facility is responsible for determining the DNA sequence of clones constructed at LANL, then returning the data to LANL for analysis and archiving.