format. Overnight bacteria growths are lysed, and samples are separated from cell debris by centrifugation. The DNA is recovered by ethanol precipitation.

Informatics

The informatics group is focused on hardware and software support and system administration, software development for end sequencing, transposon mapping and sequence template selection, data-flow automation, gene finding, and sequence analysis. Data-flow automation is the main emphasis. Six key steps have been identified in this process, and software is being written and tested to automate all six. The first step involves controlling gel quality, trimming vector sequence, and storing the sequences in a database. A program module called Move-Track-Trim, which is now used in production, was written to handle these steps. The second through fourth steps in this process involve assembling, editing, and reconstructing P1 clones of 80,000 base pairs from 400-base traces. The fifth step is sequence annotation, and the sixth is data submission.

Annotation can greatly enhance the biological value of these sequences. Useful annotations include homologies to known genes, possible gene locations, and gene signals such as promoters. LBNL is developing a workbench for automatic sequence annotation and annotation viewing and editing. The goal is to run a series of sequence-analysis tools and display the results to compare the various predictions. Researchers then will be able to examine all the annotations (for example, genes predicted by various gene-finding methods) and select the ones that look best.

Nomi Harris developed Genotator, an annotation workbench consisting of a stand-alone annotation browser and several sequence-analysis functions. The back end runs several gene finders, homology searches (using BLAST), and signal searches and saves the results in ".ace" format. Genotator thus automates the tedious process of operating a dozen different sequence-analysis programs with many different input and output formats. Genotator can function via command-line arguments or with the graphical user interface (*http://www-hgc.lbl.gov/inf/annotation.html*).

Progress to Date

Chromosome 5

Over the last year, the center has focused its production genomic sequencing on the distal 40 megabases of the human chromosome 5 long arm. This region was chosen because it contains a cluster of growth factor and receptor genes and is likely to yield new and functionally related genes through long-range sequence analysis. Results to date include:

- 40-megabase nonchimeric map containing 82 yeast artificial chromosomes (YACs) in the chromosome 5 distal long arm.
- 20-megabase contig map in the region of 5q23-q33 that contains 198 P1s, 60 P1 artificial chromosomes, and 495 bacterial artificial chromosomes (BACs) linked by 563 sequenced tagged sites (STSs) to form contigs.
- 20-megabase bins containing 370 BACs in 74 bins in the region of 5q33-q35.

Chromosome 21

An early project in the study of Down syndrome (DS), which is characterized by chromosome 21 trisomy, constructed a high-resolution clone map in the chromosome 21 DS region to be used as a pilot study in generating a contiguous gene map for all of chromosome 21. This project has integrated P1 mapping efforts with transgenic studies in the Life Sciences Division. P1 maps provide a suitable form of genomic DNA for isolating and mapping cDNA.

- 186 clones isolated in the major DS region of chromosome 21 comprising about 3 megabases of genomic DNA extending from D21S17 to ETS2. Through cross-hybridization, overlapping P1s were identified, as well as gaps between two P1 contigs, and transgenic mice were created from P1 clones in the DS region for use in phenotypic studies.

Transgenic Mice

One of the approaches for determining the biological function of newly identified genes uses YAC transgenic mice. Human sequence harbored by YACs in transgenic mice has been shown to be correctly regulated both temporally and spatially. A set of nonchimeric overlapping YACs identified from the 5q31 region has been used to create transgenic mice. This set of transgenic mice, which together harbor 1.5 megabases of human sequence, will be used to assess the expression pattern and potential function of putative genes discovered in the 5q31 region. Additional mapping and sequencing are under way in a region of human chromosome 20 amplified in certain breast tumor cell lines.

Resource for Molecular Cytogenetics

Divining landmarks for human disease amid the enormous plain of the human genetic map is the mission of an ambitious partnership among the Berkeley Lab; University of California, San Francisco; and a diagnostics company. The collaborative Resource for Molecular Cytogenetics is charting a course toward important sites of biological interest on the 23 pairs of human chromosomes (*http://rmc-www.lbl.gov*).