- 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. 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).