David Lipman, M.D.

National Library of Medicine
National Center for Biotechnology
Information
Building 38 Room 8N805
8600 Rockville Pike
Bethesda, MD 20894
USA

Dr. David Lipman is currently the Director of the National Center for Biotechnology Information (NCBI), which is a division of the National Library of Medicine within the National Institutes of Health. NCBI was created by Congress in 1988 to do basic research in computational biology, and to develop computational tools, databases and information systems for molecular biology. After medical training, Dr. Lipman joined the Mathematical Research Branch of the National Institute of Diabetes, Digestive and Kidney Diseases (NIDDK) as a Research Fellow. In his research on computational tools, he developed the most widely used methods for searching biological sequence databases. There are thousands of citations to Dr. Lipman's methods in papers that have used them to discover biological functions for unknown sequences and have thereby advanced the understanding of the molecular basis of human disease. Since 1989, Dr. Lipman has been the Director of the NCBI, a leading research center in computational biology and one of the most heavily used sites in the world for the search and retrieval of molecular biology information.

Supporting the gene expression effort

The National Center for Biotechnology Information (NCBI) of NIH has been taking a series of direct steps over the last year to insure that microarray technologies specifically, and gene expression quantifying technologies in general, have a full measure of bioinformatics support.

Essential steps in this support are taking place in two NCBI projects that are collecting, curating, annotating, organizing and publishing a host of gene-based information. The Reference Sequence project, RefSeq (http://www.ncbi.nlm.nih.gov/LocusLink/refseq.html), provides reference sequence standards for naturally occurring transcription- and translation-related polymers, including DNA, mRNA and proteins. These standards will provide stable reference points not only for mutation analyses and polymorphism discovery, but for gene expression studies as well. LocusLink (http://www.ncbi.nlm.nih.gov/LocusLink/), which is closely related to RefSeq, provides a single query interface to curated sequence and descriptive information about genetic loci. This information includes official nomenclature, aliases, sequence accession numbers, phenotypes, EC numbers, MIM numbers, UniGene clusters, map information and relevant web sites.

NCBI has been actively dealing with gene expression data through its storage and analysis of expressed sequence tag (EST) and serial analysis of gene expression (SAGE) data. For several years, NCBI has been acting as a public repository for EST data through the dbEST database (http://www.ncbi.nlm.nih.gov/dbEST/), and annotating this sequence information with gene, source and frequency information via the UniGene project (http://www.ncbi.nlm.nih.gov/UniGene) and participation in the Cancer Genome Anatomy Project (http://www.ncbi.nlm.nih.gov/CGAP). More recently, NCBI has developed a SAGE repository and several on-line tools for the exploration of that gene expression information (http://www.ncbi.nlm.nih.gov/SAGE).

As a culmination of our efforts in the gene expression arena, NCBI has been exploring the feasibility of a gene expression resource of wider scope. It is anticipated that this resource would accept gene expression data from various sources, including microarray, high-density array (DNA chip) and SAGE technologies, as well as providing a number of precomputed analyses and on-line tools for the exploration of this data.