Bioinformatics Toolbox 3

Read, analyze, and visualize genomic, proteomic, and microarray data

Bioinformatics Toolbox offers computational molecular biologists and other research scientists an open and extensible environment in which to explore ideas, prototype new algorithms, and build applications in drug research, genetic engineering, and other genomics and proteomics projects. The toolbox provides access to genomic and proteomic data formats, analysis techniques, and specialized visualizations for genomic and proteomic sequence and microarray analysis. Most functions are implemented in the open MATLAB® language, enabling you to customize the algorithms or develop your own.

Microarray Data Analysis and Visualization

Bioninformatics Toolbox enables you to analyze and comprehend raw microarray data.

Microarray Normalization

You can use several methods for normalizing microarray data, including lowess, global mean, median absolute deviation (MAD), and quantile normalization. You can apply these methods to the entire microarray chip or to specific regions or blocks. Filtering and imputation functions let you clean raw data before running analysis and visualization routines.

KEY FEATURES

- Capabilities for microarray data analysis and visualization
- Support for mass spectrometry preprocessing and analysis
- Graph theory and graph visualization tools
- Statistical learning functionality
- Gene Ontology functionality
- Sequence analysis tools including functions for pairwise and multiple sequence alignment
- Phylogenetic tree analysis tools
- Genomic, proteomic, and gene expression file formats
- Access to Web-based databases

Circular DNA map (far right), multiple sequence alignment (above, right), and secondary structure of RNA sequence (right). Bioinformatics Toolbox includes a variety of tools for visualizing sequence data.

The MathWorks

Accelerating the pace of engineering and science
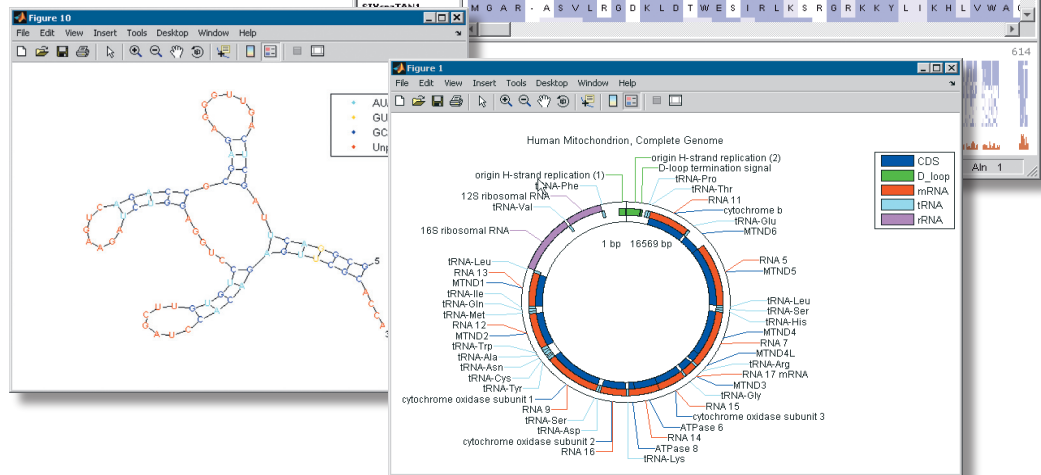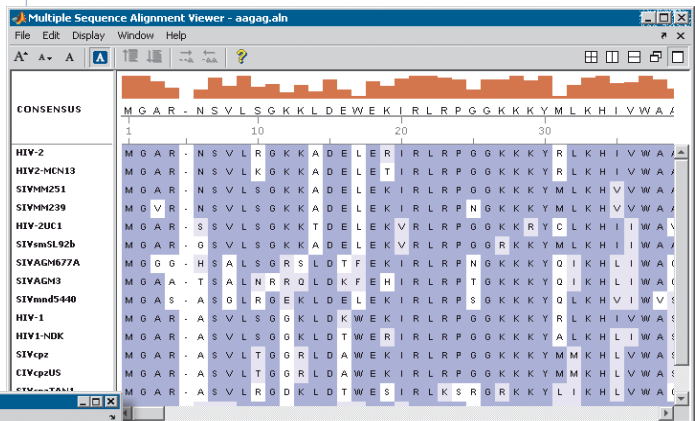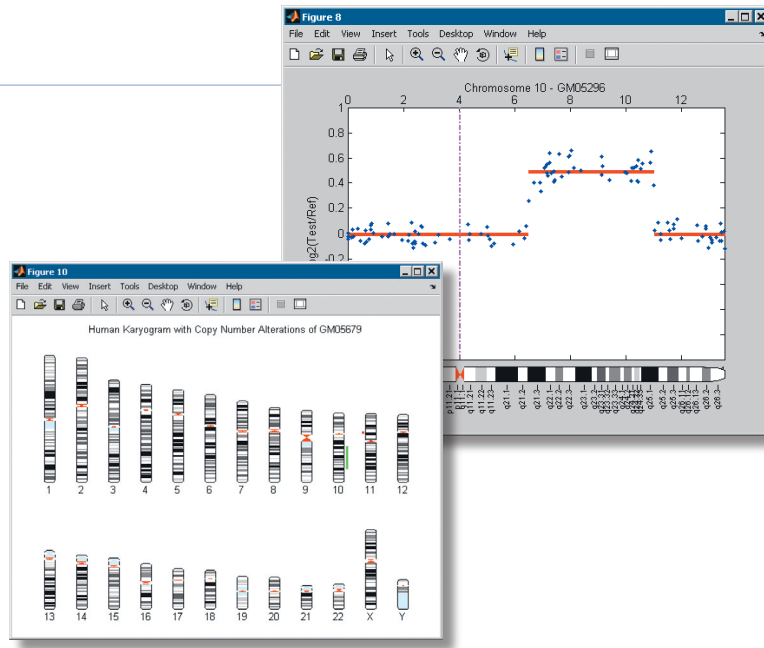
Copy number alterations calculated and viewed alongside ideograms using Bioinformatics Toolbox.

Data Analysis and Visualization

Bioinformatics Toolbox lets you perform background adjustment and calculate gene (probe set) expression values from Affymetrix microarray probe-level data using robust multiarray average (RMA) and GCRMA. You can apply circular binary segmentation (CBS) to array CGH data and estimate the false discovery rate (FDR) of multiple hypotheses testing of gene expression data from a microarray experiment. You can perform rank-invariant set normalization on either probe intensities for multiple Affymetrix CEL files or gene expression values from two different experimental conditions.

Specialized routines for visualizing microarray data include volcano plots, box plots, loglog plots, I-R plots, and spatial heat maps of the microarray. You can also visualize ideograms with G-banding patterns.

Using routines from Statistics Toolbox (required, and available separately), you can classify your results, perform hierarchical and K-means clustering, and represent your microarray data in statistical visualizations, such as two-dimensional clustergrams with optimal leaf ordering, heat maps, principle component plots, and classification trees.

Mass Spectrometry Data Analysis

A set of functions is provided for mass spectrometry data analysis. The tools are designed for preprocessing, classification, and marker identification from SELDI, MALDI, LC/MS, and GC/MS data. Preprocessing functions include baseline correction, smoothing, calibration, and resampling. You can align raw spectra data using the M/Z axis and perform retention time alignment on LC/MS and GC/MS data. A GUI lets you view multiple spectra simultaneously.

Tutorials provide step-by-step examples of how to smooth, align, and normalize spectra and then use classification and statistical learning tools to create classifiers and identify potential biomarkers.

Graph Theory and Visualization

Bioinformatics Toolbox includes functionality to apply basic graph theory to sparse matrices. Functions, objects, and methods are provided for creating, viewing, and manipulating graphs, such as interaction maps, hierarchy plots, and pathways. Examples include determining and viewing shortest paths in graphs, testing for cycles in directed graphs, and finding isomorphism between two graphs.

Statistical Learning and Visualization

Bioinformatics Toolbox provides functions that build on the classification and statistical learning tools in Statistics Toolbox. These include support vector machine (SVM) and K-nearest neighbor classifiers; functions for setting up cross-validation experiments and for measuring the performance of different classification methods; and tools for selecting discriminating features. Graph viewing and manipulation tools let you display interaction maps, hierarchy plots, and pathways.

Gene Ontology

Bioinformatics Toolbox provides functionality for accessing the Gene Ontology Database from within MATLAB, parsing Gene Ontology annotated files, and obtaining subsets of the ontology, such as ancestors, descendents, or relatives.

Sequence Analysis

Bioinformatics Toolbox provides functions for genomic and proteomic sequence analysis and visualization. You can perform a variety of analyses, including multiple sequence alignments and building and interactively viewing and manipulating phylogenetic trees.

Sequence Alignment

Bioinformatics Toolbox offers a comprehensive set of analysis methods for performing pairwise sequence, sequence profile, and multiple sequence alignment. These include:

- MATLAB implementations of standard algorithms for local and global sequence alignment, such as the Needleman-Wunsch, Smith-Waterman, and profile-hidden Markov model algorithms

- Progressive multiple sequence alignment

- Graphical representations of alignment results matrices

- Standard scoring matrices, such as the PAM and BLOSUM matrix families

- Consensus sequence calculation and sequence logo display

Sequence Utilities and Statistics

You can manipulate and analyze your sequences to gain a deeper understanding of your data. Bioinformatics Toolbox routines let you:

- Convert DNA or RNA sequences to amino acid sequences using the genetic code

- Perform statistical analysis on the sequences and search for specific patterns within a sequence

- Apply restriction enzymes and proteases to perform in-silico digestion of sequences or create random sequences for test cases

- Predict minimum free energy secondary structure of RNA sequences

Sequence Visualization

Bioinformatics Toolbox contains tools for visualizing sequences and alignments. You can view linear or circular maps of sequences annotated with GenBank features. You can visualize secondary structure diagrams of an RNA sequence. Interactive viewers let you view, modify, and explore pairwise and multiple sequence alignments.

Phylogenetic Tree Analysis

Bioinformatics Toolbox enables you to create and edit phylogenetic trees. You can calculate pairwise distances between aligned or unaligned nucleotide or amino acid sequences using a broad range of similarity metrics, such as Jukes-Cantor, p-distance, alignment-score, or a user-defined distance method. Phylogenetic trees are constructed using hierarchical linkage with a variety of techniques, including neighbor joining, single and complete linkage, and UPGMA.

Bioinformatics Toolbox includes tools for weighting and rerooting trees, calculating subtrees, and calculating canonical forms of trees. Through the graphical user interface (GUI), you can prune, reorder, and rename branches; explore distances; and read or write Newick-formatted files. You can also use the annotation tools in MATLAB to create presentation-quality trees.

Protein Feature Analysis

Bioinformatics Toolbox provides several protein analysis methods, as well as routines for calculating properties of a peptide sequence, such as atomic composition, isoelectric point, and molecular weight. You can determine the amino acid composition of protein sequences, cleave a protein with an enzyme, and create backbone plots and Ramachandran plots of PDB data. A GUI lets you view properties along the length of your sequence. Bioinformatics Toolbox also contains a Molecule Viewer tool for displaying and manipulating 3-D molecular structures.
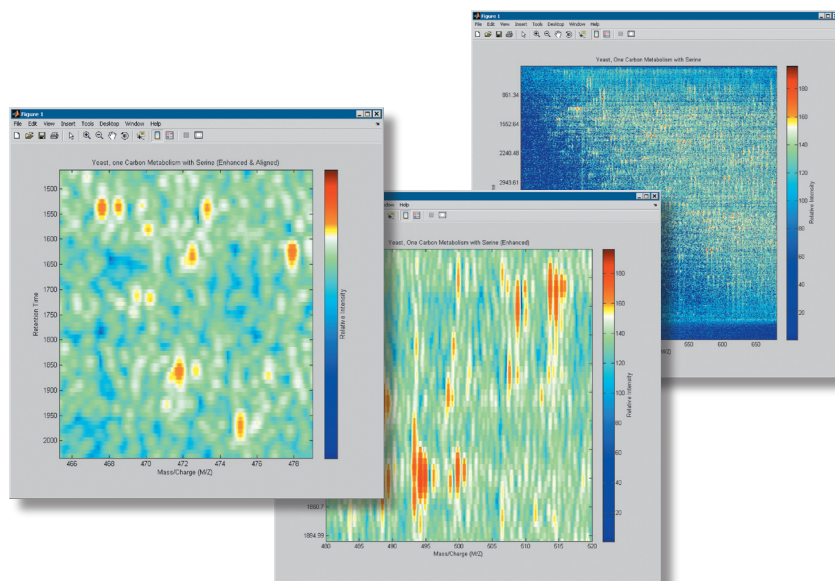
Label-free differential proteomics and metabolomics analysis using Bioinformatics Toolbox.
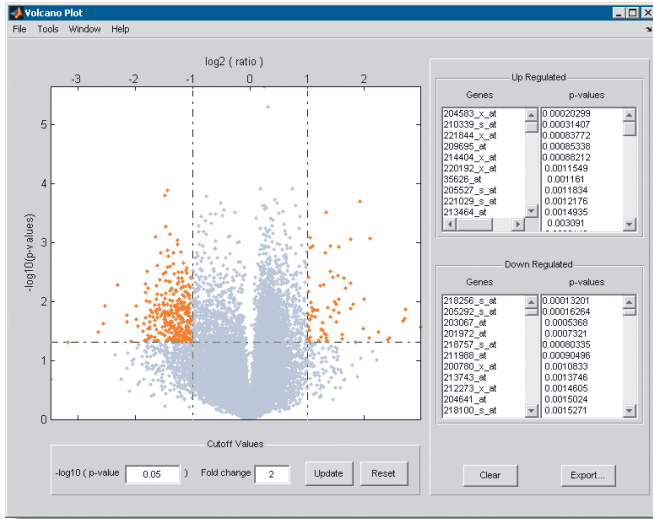
Scatter plot of microarray data showing significance versus gene expression ratio.

File Formats and Database Access

You can access many standard file formats for biological data, Web-based databases, and other online data sources from Bioinformatics Toolbox. For example, you can:

- Read sequence data from standard file formats, including FASTA, PDB, and SCF

- Read microarray data from file formats such as Affymetrix DAT, EXP, CEL, CHP, and CDF files; ImaGene results format data; Agilent Feature Extraction Software files; and GenePix GPR and GAL files

- Interface with major Web-based databases, such as GenBank, EMBL, NCBI BLAST, and PDB

- Import data directly from the NCBI Gene Expression Omnibus Web site using a single command

- Read cytogenetic banding information from NCBI ideograms or UCSC cytoband text files

- Read mass spectrometry data from MZXML and JCAMP-DX files

Sharing Algorithms and Deploying Applications with Bioinformatics Toolbox and MATLAB

MATLAB includes tools that let you turn your data analysis program into a customized software application. These include an inter-active GUI builder; programming tools, such as a visual debugger for algorithm development and refinement; and an algorithm performance profiler to accelerate development. You can share data analysis algorithms created in the MATLAB language across all platforms supported by MATLAB.

Tutorial examples show how to integrate MATLAB with commonly used bioinformatics tools, such as Bio Perl, SOAP-based Web services, and COM plug-ins. Using MATLAB application deployment products (available separately), you can integrate your MATLAB algorithms with existing C, C++, and Java applications, deploy the developed algorithms and GUIs as stand-alone applications, convert MATLAB algorithms into Microsoft .NET or COM components that can be accessed from any COM-based application, and create Microsoft Excel® add-ins.

Required Products

MATLAB
Statistics Toolbox

Related Products

Distributed Computing Toolbox. Perform distributed and parallel computations

Excel® Link. Use MATLAB from Microsoft® Excel

Image Processing Toolbox. Perform image processing, analysis, and algorithm development

MATLAB® Compiler. Build stand-alone executables and software components from MATLAB code

SimBiology. Model, simulate, and analyze biochemical pathways

For more information on related products, visit *www.mathworks.com/products/bioinfo*

Platform and System Requirements

For platform and system requirements, visit *www.mathworks.com/products/bioinfo* ∎

Resources

VISIT
www.mathworks.com

TECHNICAL SUPPORT
www.mathworks.com/support

ONLINE USER COMMUNITY
www.mathworks.com/matlabcentral

DEMOS
www.mathworks.com/demos

TRAINING SERVICES
www.mathworks.com/training

THIRD-PARTY PRODUCTS AND SERVICES
www.mathworks.com/connections

WORLDWIDE CONTACTS
www.mathworks.com/contact

E-MAIL
info@mathworks.com

The MathWorks

Accelerating the pace of engineering and science