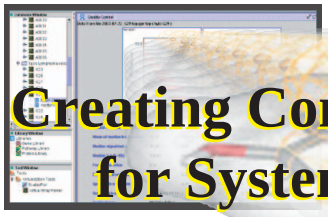
BIOINFORMATICS, CHEMOINFORMATICS & COMPUTATIONAL TOXICOLOGY

Center for Toxicoinformatics



Creating Computational Solutions for Systems Toxicology and Predictive Toxicology

Modern toxicological research faces a new challenge: data are prodigious, hyper-dimensional, complex, and very noisy - innovative bioinformatics is INDISPENSABLE for biological interpretation. The Center for Toxicoinformatics at the National Center for Toxicological Research integrates the disciplines of bioinformatics, chemoinformatics, and computational toxicology providing an encompassing environment to integrate knowledge with experimental data for biological interpretation and elucidation of mechanisms of toxicity.

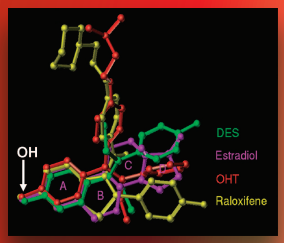
Bioinformatics:

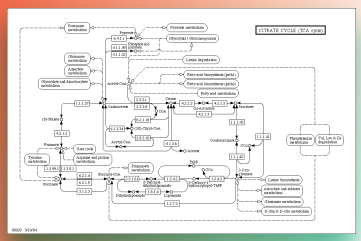
Novel methods in software design and data analysis enable the integration and interrogation of "omics" data together with traditional toxicology data and public data for interpretation in biological context.

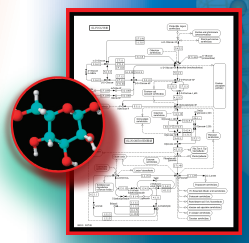
Chemoinformatics and Computational Toxicology:

Computational methods are developed to predict toxicity solely on chemical structure to complement or serve as a possible alternative to extensive and expensive animal-based studies. Knowledge bases are constructed by combining data and chemical structure with predictive models for leveraging data and resources.

Determining the differential expression patterns of disease (biomarkers) for diagnostics, prognostics treatment discovery, treatment intervention selection, and elucidation of molecular mechanisms



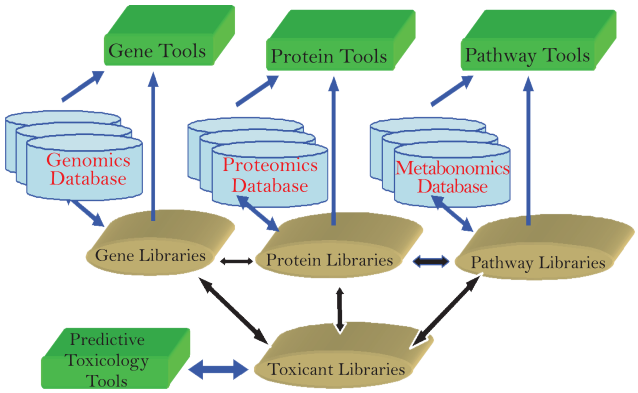




Toxicoinformatics Integrated System (TIS)

The Unified Computing Environment for Systems Toxicology Research

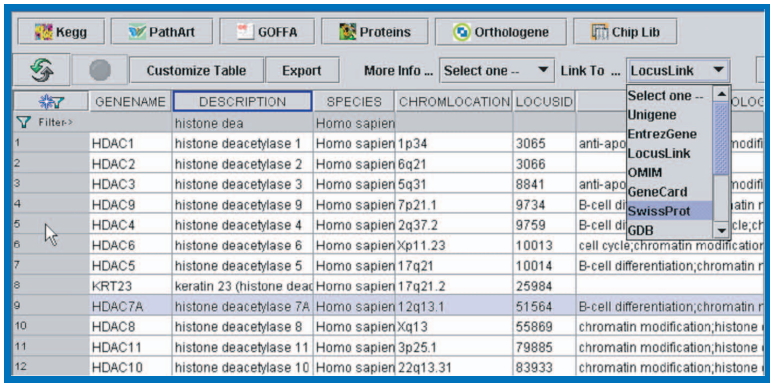
By determining the interrelationships of data from the genome, proteome, and metabonome, science will begin to unravel the networks and the feedbacks of the living cell, organ or organism. TIS integrates genomic, proteomic and metabonomic experimental data with data from the public domain as well as data from conventional toxicology to enable systems toxicology research.



BioInfoTrack

A stand-alone version of the ArrayTrack Libraries

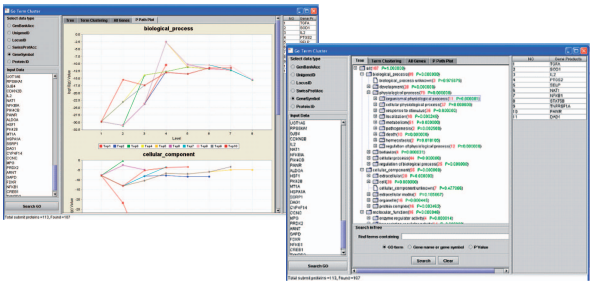
- Most relevant functional data from the public repositories are locally available
- Data from each library are interconnected on the basis of gene identity
- Library data are searchable across multiple libraries for one or many genes, proteins or pathways
- Library data are hyperlinked to data repository websites
- Pathway- and GO-based analysis are provided

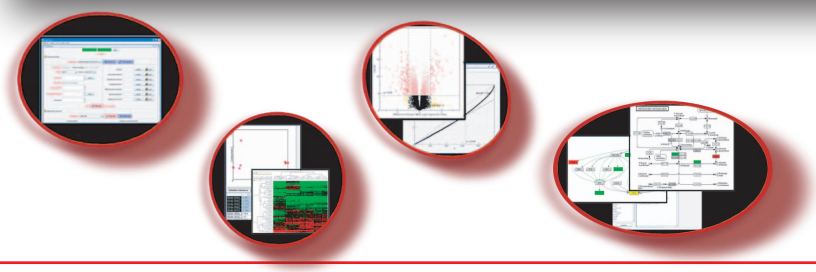


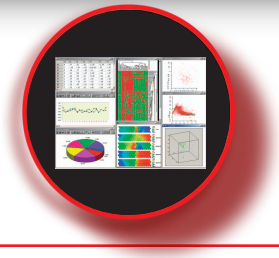
GOFFA

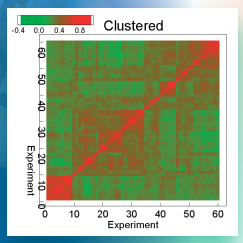
Gene Ontology For Functional Analysis

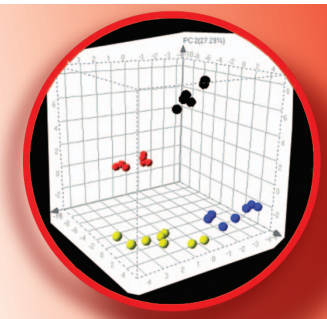
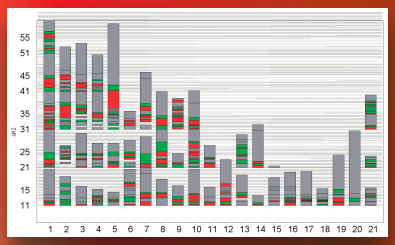
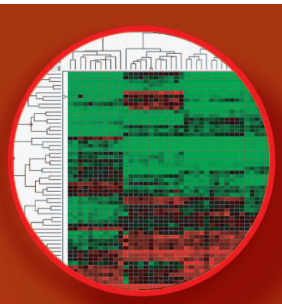
Perform gene ontology-based data interpretation with libraries containing hierarchical-organized data on biological process, cellular component, and molecular function





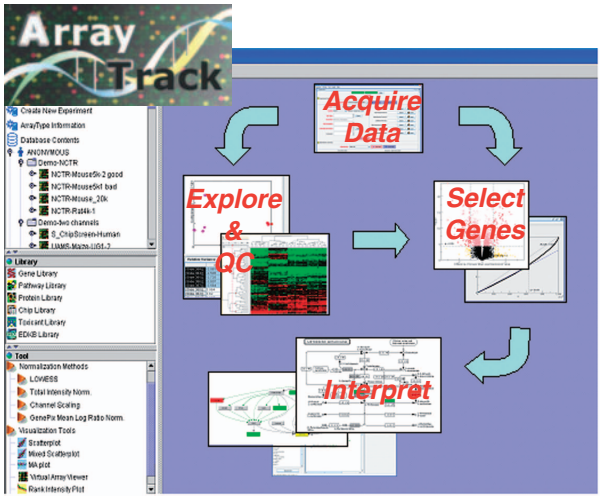




ArrayTrack

Microarray Data management, Analysis and Interpretation Software



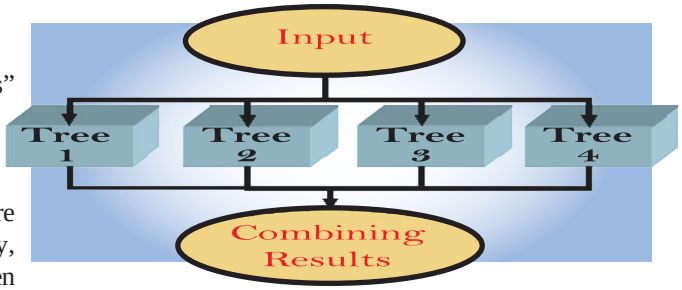
DNA microarrays offer the prospect of biomarker identification by measuring expression of tens of thousands of genes in a single experiment. The amount and complexity of microarray data impose challenges making bioinformatics indispensable. ArrayTrack is a bioinformatics solution for DNA microarray-based research that provides:

* Database with automated data import across multiple array platforms, including affymetrix GeneChip, supporting the MIAME guideline for protocols and data, and interactive graphics for QA/QC
* Libraries combining the most germane public archive data for gene and protein function/ontology, signaling and cellular pathways, and cross-species orthologs – with built in gene protein hyperlinks to data repository websites;
* A rich collection of mathematical, statistical and interactive graphical tools for significant gene selection, data exploration, and ultimately biological interpretation

http://www.fda.gov/nctr/science/centers/toxicoinformatics/ArrayTrack

Decision Forest (DF)

A pattern recognition and prediction system with versatile applicability



Data from high throughput technology-based "omics" research contain patterns characterizing disease and exposure that are challenging to extract with statistical certainty. DF is a novel consensus pattern recognition method based on the combination of multiple heterogeneous models to reach more reliable forecast. DF is versatile for analysis of microarray, proteomic, pharmacogenomic, and toxicology data. DF has been successfully applied to:

DF-SNPs: Determine SNP patterns associated with disease for use as biomarkers

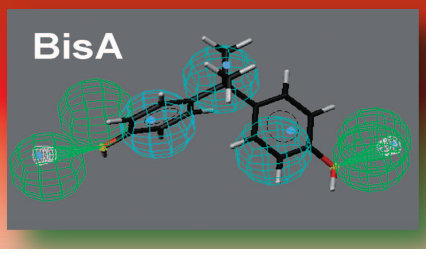
DF-Microarray: Determine differential gene expression patterns from microarrays associated with disease and exposure

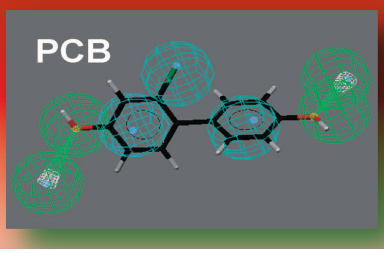
DF-SELDI: Determine differential protein expression patterns from mass spectroscopy associated with disease and exposure

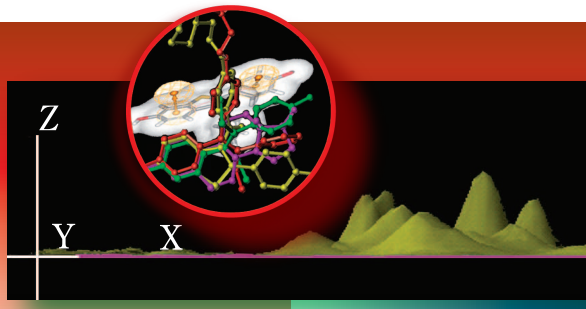
DF-Seq: Determine protein function based on amino acid sequence

DF-SAR: Predict toxicological endpoints solely based on chemical structure

http://www.fda.gov/nctr/science/centers/toxicoinformatics/DecisionForest





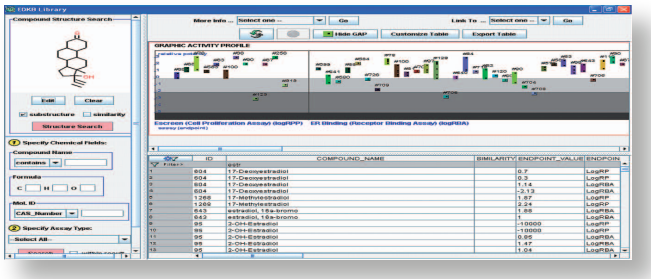


Computational Toxicology

Correlating chemical structure with biological endpoint

The decades old toxicology paradigm for animal testing is expensive in terms of resource consumption. As mechanistic knowledge is gained, it becomes increasingly feasible to use analytical models to predict toxicity of an untested chemical solely on chemical structure. NCTR scientists are experienced with the full spectrum of methods and software for developing predictive toxicology models. The methods are essentially the same as used in rational drug design, such as:

* SAR and pharmacophore motifs
* Quantitative Structure Activity Relationships (QSAR)
* Supervised and unsupervised approaches.



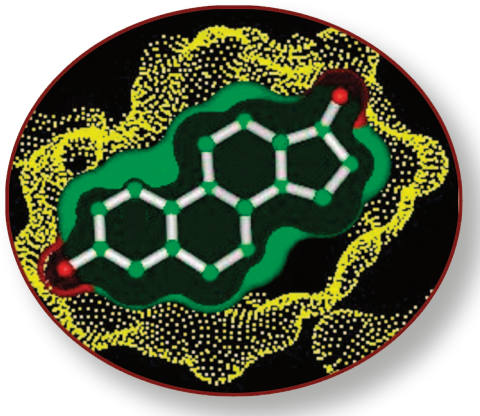
Endocrine Disruptor Knowledge Base (EDKB)

A knowledge base system supporting assessment of a chemical's ability to mimic hormones and disrupt endocrine function

The NCTR EDKB comprises a large collection of online endocrine disruptor data, citations, and computational models to predict estrogen and androgen activity. As a knowledge base, it is intended to serve as a resource for research and regulatory scientists, and other interested parties, to make use of an existing body of knowledge, foster the development of computational predictive toxicology models, and reduce dependency upon expensive animal experiments.

Chemometric and QSAR models of the EDKB are capable of predicting binding of an untested chemical to the estrogen and androgen protein receptors solely on chemical structure.

http://www.fda.gov/nctr/science/centers/toxicoinformatics/edkb/



NCTRDescriptor

Generation of more than 700 descriptors for SAR/QSAR study

Models to predict biological activity based on structure require descriptors to represent chemical structure. There are many commercial products to produce one, two and three-dimensional descriptors, but their proprietary nature precludes both the integration with "omics" infrastructure and Internet-based applications. To remove this impediment, the NCTRDescriptor has been developed. The descriptors are used both for the Decision Forest as well as other chemometric and QSAR methods.

March 10, 2006

Weida Tong, Ph.D.
Director, Center for Toxicoinformatics
3900 NCTR Road
Jefferson, AR 72079
Tel: 870-543-7142 Fax: 870-543-7382
E-mail: weida.tong@fda.hhs.gov

William Slikker, Jr.., Ph.D.
Acting Director, NCTR
3900 NCTR Road
Jefferson, AR 72079
Tel: 870-543-7950 Fax: 870-543-7576
E-mail: william.slikker@fda.hhs.gov