POSTER SESSIONS

ISMB
ECCB
2004

A-1

A Bayesian Approach to Discover Sparse and Flexible Patterns in Biological Sequences

Kazuhito Shida[1], Makoto Ikeda[2], Atsuo Kasuya
[1]shida@cir.tohoku.ac.jp, CIR Tohoku univ.; [2]CIR Tohoku univ.

Patterns in biological sequences may have sparse design. However, the most flexible kind of such pattern (letter-composition at the sites and the length of the gaps are both variable) is scarsely investigated. A Bayesian automated detector of such patterns was deviced and tested successfully onsome of known examples.

A-2

ProtExt: a Web-based Protein-Protein Interaction Extraction System for PubMed Abstracts

Chin-Lin Peng[1], Yung-Chung Lin[2], Hsuan-Cheng Huang, Cheng-Yan Kao, Shui-Tein Chen, Hsueh-Fen Juan
[1]r91060@csie.ntu.edu.tw, Department of Computer Science and Information Engineering, National Taiwan University; [2]xern@cpan.org, Department of Computer Science and Information Engineering, National Taiwan University

ProtExt is a web-based software package, which automatically extracts information of protein-protein interactions from the literature abstracts available at the NCBI Entrez-PubMed system and present the extracted information graphically and intuitively. ProtExt is available at http://protext.csie.org.

A-3

Automatic Hidden Markov Model Selection for Clustering Procedures

Schoenhuth, Alexander[1], Schliep, Alexander[2], Olof Persson, Christine Steinhoff
[1]schoenhuth@zpr.uni-koeln.de, Center for Applied Computer Science, University of Cologne; [2]schliep@molgen.mpg.de, Max Planck Institute for Molecular Genetics, Berlin

Model based clustering methods have proved to be superior for the analysis of gene expression profiles as they care for dependencies along the time axis. Choosing models for initialization is a crucial issue. We present an efficient method for inferring suitable HMMs from the data in a Bayesian framework.

A-4

What do Data Sets have in Common? The Yeast Stress Case

Samuel Kaski[1], Janne Nikkilä[2], Christophe Roos
[1]samuel.kaski@cs.helsinki.fi, University of Helsinki; [2]janne.nikkila@hut.fi, Helsinki University of Technology

We suggest using a combination of two new data analysis methods to search for properties that are a priori unknown but common to several data sets: dependency-preserving dimensionality reduction by generalization of canonical correlations, followed by dependency-searching clustering. These methods are especially well suited for analyzing yeast stress genes

A-5

BioMart: a Distributed and Query-Optimised Data Integration System

Damian Smedley[1], Darin London[2], Craig Melsopp, Damian Keefe, Andreas Kahari, Katerina Tzouvara, Ewan Birney, Arek Kasprzyk
[1]damian@ebi.ac.uk, EBI; [2]dlondon@ebi.ac.uk, EBI

BioMart is a simple, distributed data integration system allowing optimised querying across single or multiple data sources. The system consists of a schema specification, an XML-configuration system and tools for deployment, and web and standalone interfaces. This domain and RDBMS platform independent system is freely available at http://www.ebi.ac.uk/biomart.

A-6

Analyzing Clinical Records and Biomedical Texts with AMBIT

Rob Gaizauskas[1], Mark Hepple[2], H. Harkema, N. Davis, Y. Guo, A. Roberts, I. Roberts
[1]R.Gaizauskas@dcs.shef.ac.uk, University of Sheffield; [2]M.Hepple@dcs.shef.ac.uk, University of Sheffield

It is practically infeasible for researchers to find relevant information in clinical records and biomedical texts due to their unstructured, textual format and their volume. We introduce AMBIT, a system that extracts entities and relationships from texts and presents this information in a structured format for rapid and effective access.

A-7

Scalable Text Mining

Mark Cumiskey[1]
[1]mcumiske@inf.ed.ac.uk, University Of Edinburgh

To extract context information from large datasets a scalable approach is required. The method proposed uses a three-tier approach. The first pass is a search using a distributed index across multiple clusters. The second evaluates the text per-sentence. Finally, a classier trained upon correct data is used to disqualify hits

A-8

Automatic Extraction of Protein Interactions and Biological Functions from Text

A. Koike[1], Y. Niwa[2], T. Takagi
[1]akoike@hgc.jp, Univ. of Tokyo; [2], Hitachi, Ltd.

We have developed a method for automatically extracting the biological functions and interaction information of genes/proteins/families from biomedical text using a shallow parser and sentence structure analysis techniques. The extraction performance has been discussed. The results applied to PUBMED abstracts are open to the public in the PRIME database http://prime.ontology.ims.u-tokyo.ac.jp/

A-9

Prediction of Single Nucleotide Polymorphisms in Domestic Tomato: How Useful is EST Sequence Diversity?

Angela M. Baldo[1], Joanne Labate[2], Larry D. Robertson
[1]abaldo@pgru.ars.usda.gov, USDA-ARS Plant Genetic Resources Unit; [2]jl265@cornell.edu, USDA-ARS Plant Genetic Resources Unit

Cultivated tomato is low in genetic diversity. We screened an NCBI Unigene set for SNPs. 2,527 potential SNPs were found among 764 clusters. We are in the process of verifying these polymorphisms and comparing the results with sequence derived from introns, and regions of published, mapped markers.

A-10

Dinucleotide Relative Abundance (DrA)-Based Classification of Viruses; Full-Length Genome Sequence Analysis of Whole Papillomaviridae(PV) Family

HoJoon Lee[1], Sehee Hwang [2], Insuk Sohn, Sujong Kim, Young Sik Lee and Yong Sung Lee
[1]clue96@freechal.com, Department of Biology, Yonsei University; [2]hsehee@cj.net, Researcher Nutraceuticals & Functional Food Div. CJ Foods R/D

DrAs of whole genome sequences of PV family were calculated and their profiles were hierarchically clustered for a sequence - based classification of this virus family. The clusters are similar to PV categories from cladistic analysis based on similarity of specific coding regions. Our phonetic approach might describe the sequence signals embedded in genome sequences.

A-11

ALOB - Automating Local Database Building and DB Searching Tool for the Promoter Analysis of Microarray Data

Junsu Ko[1], Insuk Sohn[2], Sujong Kim, Sehee Hwang and Junghyun Namkung
[1]neosphere@korea.ac.kr, Life Science, School of Life Sciences and Biotechnology, Korea University; [2]sis46@korea.ac.kr, Department of Statistics, Korea University

ALOB is a Java application for extracting various types of data that are used for the promoter analysis of microarray data from public sources and building local Databases automatically. UniGene, LocusLink, Gene Ontology and Ensemble Databases are utilized to extract data and to make local Databases.

A-12

Report on the BioCreAtIvE Workshop, Granada 2004

Christian Blaschke[1], Lynette Hirschman[2], Alexander Yeh, Marc Colosimo, Alexander Morgan, Alfonso Valencia
[1]blaschke@GREDOS.CNB.UAM.ES, CNB, Madrid; [2]lynette@mitre.org, MITRE

The first BioCreAtIvE Workshop (Critical Assessment of Information Extraction in Biology) was held in Granada, Spain March 28-31, 2004. The workshop had 27 international participating groups, assessed on two tasks: gene/protein name extraction from MEDLINE abstract and functional (Gene Ontology) annotation from full text.

A-13

Evaluating the Usefulness of Stem-Loop Characteristics in Pinpointing Structural RNA Genes

Kirt Noel[1], Kay Wiese[2]
[1]kirtnoel@sfu.ca, Simon Fraser University;
[2]wiese@sfu.ca, Simon Fraser University

An efficient structural RNA gene-finder which works effectively across the GC content spectrum has remained elusive. Rather than rely on Free Energy calculations, our research explores the usefulness of stem-loop characteristics in pinpointing structural RNA genes. These include average stem-loop spacing and length observed in structural RNA's compared to their counterparts.

A-14

Analysis of Non-coding Regions in Arabidopsis Thaliana

Judith Lucia Gomez[1], Diego Mauricio Riaño[2], Ingo dreyer, Bernd Mueller-Roeber
[1]jgomez@rz.uni-potsdam.de, University of Potsdam;
[2]diriano@rz.uni-potsdam.de, University of Potsdam

Intergenic regions harbor short regulatory sequences that control gene expression. We determine some compositional properties for intergenic, upstream and conding regions of Arabidopsis thaliana, and investigate the possibility to extract a typical intergenic vocabulary (octamers and pentamers), that could be related with known or putative transcriptional-regulatory motifs.

A-15

METIS: Multiple Extraction Techniques for Informative Sentences

Anna Divoli[1], Alex Mitchell[2], Paul Bradley, Terri Attwood
[1]divoli@bioinf.man.ac.uk, The University of Manchester; [2]mitchell@ebi.ac.uk, EBI

METIS combines an existing tool, PRECIS, with two information extraction systems: probability-based Bayesian filters and the rule-based BioIE. Taking a single protein query sequence, the combined tool can first generate a report from related Swiss-Prot entries, and then extract pertinent biomedical information, in the form of sentences, from the literature.

A-16

Web Services and Client Packages as a Framework for Data Mining on the Neighborhood of Organized Sets of Biological Sequences

Barriot Roland[1], Lamiable Alexis[2]
[1]barriot@labri.fr, LaBRI; [2]alexis@ragondux.com, LaBRI

Web services dedicated to Data Mining on the neighborhood of sets of biological sequence have been deployed together with the release of Perl, Ruby and Java packages as a framework for the rapid development of client applications.

A-17

TextLens/SeC: A Sentence Classification System for Abstracts of Biomedical Literature

Yasunori Yamamoto[1], Toshihisa Takagi[2]
[1]yayamamo@hgc.jp, University of Tokyo; [2]tt@k.u-tokyo.ac.jp, University of Tokyo

TextLens/SeC is a sentence classification system which classifies sentences of abstracts of biomedical literature into four categories: background, method, result, and conclusion. It uses a ML approach (SVM) and is a component of our currently developing multidocument summarization system which allows researchers to briefly but sufficiently learn a research topic.

A-18

EVPPI - Extraction and Visualization of Protein-Protein Interactions

Mikael Andersson[1], Per Lilja[2], Dr. Per-Åke Jovall and Dr. Thorsteinn Rögnvaldsson
[1]contact@mickeandersson.net, Halmstad University;
[2]private@perlilja.net, Halmstad University

The EVPPI tool extracts information about protein-protein interactions from plain text and displays these in an interactive graphical environment. Protein names are extracted from text using a Support Vector Machine, which performs better than previously suggested algorithms for this. Interactions from more than one source can be merged and compared.

A-19

UVWORD: Fast Genome-Scale Analysis of DNA Words on a PC

Vicente Arnau[1], Miguel Gallach[2], Francisco Ferri, Ignacio Marín
[1]Vicente.Arnau@uv.es, Departamento de Informática, Universidad de Valencia. Spain; [2]Ignacio.Marin@uv.es, Departamento de Genética, Universidad de Valencia. Spain

We have developed a new algorithm that allows for the exhaustive determination of 1-14 nucleotides-long words in DNA sequences of any length. It is fast enough as to be used at a genomic scale running on a standard personal computer.

A-20

A Cataloguing System for Diverse Biological Datasets

T. G. Booth[1], A. J. Wood[2], P. Swift, B. Tiwari and D. Field.
[1]tbooth@ceh.ac.uk, NERC; [2]anjw@ceh.ac.uk, NERC

Biologists produce many diverse datasets in the course of their research. We present a cataloguing system to capture and search descriptions of all data and sample sets associated with a research project. Our approach uses only freely available software, and can be easily adapted to suit any metadata capture need.

A-21

Analysing Patterns of DNA Methylation Associated With Clinical Outcome

Hall, J[1], Paul, J[2], Brown, R
[1]j.hall@beatson.gla.ac.uk, University of Glasgow; [2]j.paul@clinmed.gla.ac.uk, University of Glasgow

Patterns of tumour CpG island DNA methylation associated with clinical outcome could aid disease management of patients, give insight into biological mechanisms and prioritiz future research areas. We investigated a supervised clustering methodology; initial results find associations but adequate power and validation are essential for correct interpretation of results.

A-22

Identification of Candidate Disease Genes using Text Mining with Controlled Vocabularies

N. Tiffin[1], J. F. Kelso[2], A. R. Powell; P. Hong; V. B. Bajic; W.A. Hide
[1]nicki@sanbi.ac.za, SANBI; [2]janet@sanbi.ac.za, SANBI

Complex-trait linkage analysis of a disease may identify hundreds of candidate disease-causing genes. Text-mining of Entrez-PubMed abstracts using eVOC ontology terms identifies tissues and cell types associated with the disease. Candidate genes expressed in these tissues may then be identified according to their expression profiles as defined by eVOC annotation.

A-23

GoPubMed: Exploring PubMed with the GeneOntology

Michael Schroeder[1]
[1]ms@mpi-cbg.de, TU Dresden

GoPubMed implements a novel approach to literature search. GoPubMed extracts GeneOntology (GO) terms from relevant abstracts obtained by keyword search and thus classifies the abstract into multiple categories. From the extracted GO-terms GoPubMed derives the induced ontology, which allows the users to systematically explore the search results

A-24

An Information System to Integrate Functional Genomic Data Related to Nutrition

F. Desiere[1], R Munro, Anne Donnet[1], Rainer Warth[1]
[1]rainer.warth@rdls.nestle.com, Nestle Research Center

Our system provides the infrastructure to collect and analyze functional genomic data. Based on the requirement analysis of 20 scientists, a system was designed to collectively data-mine transcriptomics, proteomics and ENSEMBL genome data. After two years of operation, app. 2000 proteins with an average of 10 different annotations are integrated.

A-25

Biological Named Entity Recognition System Using UMLS

Hyun-Sook Lee[1], Soo-Jun Park[2], Hyun-Chul Jang, Seon-Hee Park
[1]lhs63473@etri.re.kr, Bioinformatics Research Team, Electronics and Telecommunications Research Institute;
[2]psj@etri.re.kr, Bioinformatics Research Team, Electronics and Telecommunications Research Institute

In order to solve domain portability issue, we propose an automated method for recognizing biological named entity that minimizes the cost of resource construction and rule generation by using UMLS without the help of experts and curated large corpus. Our experiment with 100 abstracts on apoptosis shows a reliable result.

A-26

Integration of Annotation and Functional Data to Assesses the Quality of Protein Interaction Data

Mike Cornell[1], Norman W. Paton[2], Stephen G. Oliver
[1]mcornell@cs.man.ac.uk, Department of Computer Science, University of Manchester;
[2]norm@cs.man.ac.uk, Department of Computer Science, University of Manchester

Using microarray and annotation data we compared protein interactions common to multiple data sets with those found in a single set. Our results suggest a high frequency of false positives in protein interaction Databases, and we propose an approach to identify confirmed interactions for which there is corroborating evidence.

A-27

A Fast Subgraph Search Method Using A Local Index for Functional Annotation of Protein Structures

Deepak Bandyopadhyay[1], Jun (Luke) Huan[2], Wei Wang, Jack Snoeyink
[1]debug@cs.unc.edu, Dept. of Computer Science, University of North Carolina at Chapel Hill;
[2]huan@cs.unc.edu, Dept. of Computer Science, University of North Carolina at Chapel Hill

Structural genomics projects are generating new protein structures in high-throughput mode, some of which have unknown function. Given such a structure, we would like to annotate its functional regions by searching for family-specific structural motifs (signatures). We present a fast algorithm using local indices to restrict the search space.

A-28

A Simple Approach for Protein Name Identification

Katrin Fundel[1], Joannis Apostolakis[2], Katrin Fundel, Daniel Güttler, Ralf Zimmer, Joannis Apostolakis
[1]Katrin.Fundel@bio.ifi.lmu.de, Ludwig-Maximilians-Universität München;
[2]Joannis.Apostolakis@bio.ifi.lmu.de, Ludwig-Maximilians-Universität München

We present a simple and efficient approach for protein name identification in medline abstracts. The approach is based on extensive curation of synonym lists obtained from public Databases and consequent exact matching of synonyms against medline abstracts. The approach showed good performance in the BioCreative assessment

A-29

PROTEUS? Assisting Protein Sequence Analysis by Fusion, Comparison and Annotation of Selected Data from Different Databases

Michal J. Gajda[1], Janusz M. Bujnicki[2], ?ukasz Jancewicz, Mariusz Zawadzki, Micha? Kurowski, Grzegorz Papaj
[1]mgajda@mini.pw.edu.pl, Warsaw Technical University;
[2]iamb@genesilico.pl, International Institute of Molecular and Cell Biology

We developed PROTEUS, an object-oriented software framework for comparative analysis of protein sequences and structures. PROTEUS facilitates construction and annotation of superfamily meta-profiles by database searches, fold-recognition and structure superposition. We used PROTEUS to build a database of DNA repair enzymes and to identify their homologs in the human genome.

A-30

Towards Automating the Curation Decision for the Nuclear Protein Database (NPD)

Catherine Canevet[1], Wendy Bickmore[2], Bonnie Webber
[1]ccanevet@inf.ed.ac.uk, ICCS, School of Informatics, University of Edinburgh;
[2]Wendy.Bickmore@hgu.mrc.ac.uk, MRC Human Genetics Unit

The NPD is a hand-curated database, containing information drawn from the biological literature on vertebrate proteins in the cell nucleus. Automated classifiers deciding whether a paper should or should not be curated currently demonstrates 0.56 precision, 0.817 recall and 0.665 f-score on a representative corpus of 13945 MedLine abstracts

A-31

Benchmarking Classification Analysis Applied to Microarray Data

Benedikt Brors[1], Patrick Warnat[2], Roland Eils
[1]b.brors@dkfz.de, DKFZ; [2]p.warnat@dkfz.de, DKFZ

We suggest methods to compare different classification methods of microarray data and show results on two data sets from leukemia and breast cancer patients. In this study, we used support vector machines, artificial neural networks and multiple decision trees. Most classifiers reached nearly equal performance on both data sets.

A-32

A System to Select Articles from PubMed Search: PETER

Hiroko Ao[1], Yasunori Yamamoto[2], Toshihisa Takagi
[1]aohiroko@hgc.jp, Department of Computational Biology, University of Tokyo, and Basic Research Laboratory, Kanebo, LTD.; [2]yayamamo@ims.u-tokyo.ac.jp, Department of Computational Biology, University of Tokyo

With the rapid growth of machine-readable literature such as MEDLINE database, retrieval of articles is an important task. In this situation, we propose an efficient system to filter results of a PubMed search. This system allows a user to reflect his/her precious knowledge in its search, customize its search conditions.

A-33

OntoSieve: Detecting Themes in Protein or Gene Sets using Structured Annotations

Jonathan Swinton[1], Elgar Pichler[2], David deGraaf, Ian S. Peers, Larry Furlong
[1]joanthan.swinton@astrazeneca.com, AstraZeneca ; [2]elgar.pichler@astrzeneca.com, AstraZeneca

GO-like annotations can be used to find annotation terms which are enriched in a given list of proteins. Existing tools partly flatten the rich ontology graph into a simple association list. We use OntoSieve, designed to preserve hierarchical information, to detect bias in interactor detection in large-scale protein interaction datasets.

A-34

Sub-Quadratic Method for Finding Most Similar Pairs of Genes and Fast Approximate Hierarchical Clustering

Meelis Kull[1], Jaak Vilo[2]
[1]meelisk@ut.ee, Department of Computer Science, University of Tartu; [2]vilo@egeen.ee, EGeen

Gene expression data clustering methods group together similar genes. We have developed a method that avoids calculations of all-against-all distances, yet identifies rapidly most of the similar gene pairs. Using this method we have developed fast approximate versions of single- ,average-, and complete-linkage hierarchical clustering.

A-35

Cataloguing of Disease-Associated Genetic and Environmental Factors by a Text-Mining Approach

Naoki Nagata[1], Teruyoshi Hishiki[2], Yumi Yamaguchi-Kabata, Teruhiko Yoshida, Tadashi Imanishi, Takashi Gojobori
[1]nnagata@jbirc.aist.go.jp, Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan; [2]t-hishiki@jbirc.aist.go.jp, Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan

"Common diseases" are caused by combinations of multiple genetic and environmental factors. We developed a text-mining system for automatic detection of smoking-related disease genes from literatures. The resulting database will be important for both the clinical studies and members of the public interested in gene-linked smoking related diseases.

A-36

Mining Biological Networks

Catherine Hack[1]
[1]cj.hack@ulster.ac.uk, University of Ulster

A directed graph describing the gene regulatory network of Saccharomyces cerevisiae was constructed from genome wide transcriptome data for 274 single gene deletion mutants. Graph metrics: degree distribution; average degree; and clustering coefficient were determined and the use of the graph structure as a basis for Data Mining was explored.

A-37

GenomeCube - Systematic Categorization, Verification and Annotation of Genome Research Material and Data to Support Efficient Functional Analysis of Genes

Juerchott, K[1], Neubert, P[2], Korn, B, Heil, O, Maurer, J, Drescher, B
[1]juerchott@rzpd.de, RZPD; [2]neubert@rzpd.de, RZPD

RZPD is a charitable provider of clones, clone related products and high-throughput services for genome research world-wide. In order to allow researchers to take conveniently advantage of the biological resources as well as of data derived thereof, RZPD now systematically categorizes, verifies and annotates its resources and offers convenient access to them by the concept "GenomeCube".

A-38

MARVEL and GOMEZ: Tools for the Comparison of Molecular Profiling Experiments

Luiz Miguel Camargo[1], Timothy P. Bonnert[2]
[1]miguel_camargo@merck.com, Merck Sharp & Dohme; [2]tim_bonnert@merck.com, Merck Sharp & Dohme

To permit the analysis and comparison of data from multiple experiments and from different platforms such as QPCR, microarray, and proteomics, we have designed both an interface for result summarisation and combination, MARVEL, and a classification visualisation tool GOMEZ. Together, these tools provide oblique views into multiple experimental data sets.

A-39

Optimization of MHC Class I Prediction

Rahee Shaban[1], Clare Sansom[2], David Moss[3]
[1]r.shaban@mail.cryst.bbk.ac.uk, School Of Crystallography, Birkbeck College; [2]c.sansom@mail.cryst.bbk.ac.uk, School Of Crystallography, Birkbeck College; [3]d.moss@mail.cryst.bbk.ac.uk, School Of Crystallography, Birkbeck College

CombiPRED is a matrix-based MHC Class I prediction tool combining MHC allele matrices from three MHC prediction programs - nHLAPred, BIMAS and SYFPEITHI - that are available in the public domain. We used the ROC method to pick the best performing matrices from these programs, and combined them to form CombiPRED

A-40

A Proteomics Software Platform Integrating Validation of Experimental Data with Subsequent Categorization, Prioritization and Data Mining in a Rich Biological Context

Soeren Schandorff[1], Jan Christian Brønd[2], Dan Bach Kristensen, Christian H. Ahrens, Hans Jespersen, Keiryn L. Bennett, Alexandre V. Podtelejnikov, Kenneth Budin, Jesper Matthiesen, Peter Venø, Søren Larsen, Brian Ramsgaard, Peder Thusgaard Ruhoff
[1]Schandorff@mdsdenmark.com, MDS Denmark; [2]jbrond.mdsdenmark.com, MDS Denmark

We present a protein identification software platform for liquid tandem mass spectrometry (LC MS/MS) of complex mixtures. The platform bridges the process from validation of experimental results to the categorization, prioritization and Data Mining of the biological results, through comparative data analysis of peptides and proteins in a bioinformatics context

A-41

goCluster: Identifying Functionally Relevant Clusters in Microarray Expression Data

Gunnar Wrobel[1], Michael Primig[2], Monica Boselli, Angelika Amon
[1]gunnar.wrobel@unibas.ch, Swiss Bioinformatics Institute; [2]michael.primig@unibas.ch, Swiss Bioinformatics Institute

goCluster employs clustering to partition an expression dataset into groups of genes with similar expression. Each cluster is subsequently searched for genes with similar annotation using statistical testing. goCluster rapidly identifies functional areas with significant expression changes. The program has been implemented in R and is based on Bioconductor.

A-42

Biomax Literature Mining Services

Victoria Penide-Lopez[1], dr. Karsten Wenger[2], eduardo torres, chantal ackermann
[1]victoria.penide-lopez@biomax.com, Biomax Informatics AG; [2]karsten.wenger@biomax.com, Biomax Informatics AG

The wealth of data for the bio-sciences requires the implementation of systems with powerful linguistical/statistical tools. Access to the data is simplified by summarizing information in a formal conceptual framework. Relevant base entities are related to each other in Databases for further processing using domain specific background-knowledge and linguistics.

A-43

A Dynamic Data Structure for Mining the Protein Structure Database

Alex Pooley[1], Andrew Turpin[2], Simon Puglisi and Steven Bottomley
[1]pooleyam@computing.edu.au, Curtin University of Technology; [2]andrew@computing.edu.au, Curtin University of Technology

A dynamic inverted file index was used to mine information on phi, psi angle distributions in known protein structures within the Molecular Modeling Database (MMDB). The method returns phi, psi angle distributions for input query sequences in under a second on a standard PC and currently only requires 15Mb of memory.

A-44

Gene-Ontology Analysis Reveals Association of Tissue-Specific 5′ CpG-island Genes with Development and Embryogenesis

Peter N. Robinson[1], Ulrike Böhme[2], Rodrigo Lopez, Stefan Mundlos, Peter Nürnberg
[1]peter.robinson@charite.de, Charité University Hospital, Institute of Medical Genetics;
[2]ulrike.boehme@charite.de, Charité University Hospital, Institute of Medical Genetics

We analyzed associations between Gene-Ontology terms and 5′-CpG islands in human RefSeq genes. We found a highly significant overrepresentation of 5′ CpG island genes annotated to development and related terms compared to RefSeqs without CpG islands, which became more significant when tissue-specific RefSeqs were analyzed separately.

A-45

Mining the PDB for Information Leading to Successful Protein Crystallization Experimental Design

David S Dougall[1], Vanathi Gopalakrishnan[2], Dan Hennessey, John Rosenberg
[1]dougalld@cbmi.pitt.edu, University of Pittsburgh;
[2]vanathi@cbmi.pitt.edu, University of Pittsburgh

Primary sequences with valid experimental pH values were extracted from the PDB and theoretical titration curves were calculated. These curves were then clustered using a self-organizing map. Significant differences were found among several variables, including experimental pH. The observed differences suggest experimental regions for the design of crystallization experiments.

A-46

MotifMiner++: A High Performance Data Mining Toolkit for Mining Biomolecular Structures

Keith Marsolo[1], Srini Parthasarathy[2], Chao Wang, Hongyuan Li, Dmitrii Polshakov,
[1]marsolo@cis.ohio-state.edu, OSU; [2]srini@cis.ohio-state.edu, OSU

Discovering important structures in molecular datasets has been the focus of many recent research efforts in biological and chemical informatics. In this poster we present a domain extensible high performance toolkit for examining such datasets. Specifically here we examine extensions that allow for mining and analyzing protein data.

A-47

An Algorithm for Biological Network Inference Based on DE Models

Ruye Wang[1], Zhaohua Tang[2]
[1] ruye_wang@hmc.edu, Harvey Mudd College;
[2]ztang@jsd.claremont,edu, Joint Science Department, Claremont Colleges

A network inference method is presented to estimate the parameters of a differential equation model of a biological network containing a large number of components (genes, proteins, neurons, etc.). The method decomposes the problem into a set of sub-problems each finding the parameters associated with a single component by solving an algebraic problem in semi-closed form.

A-48

Comparisons of SVM and ICA for Distinctive Gene Expression Patterns

Tyesia Pompey [1], Cranos William [2], Jung Kim and Winser Alexander
[1]tp992605@ncat.edu, North Carolina A&T State University ; [2]cmwilli5@ncsu.edu, North Carolina State University

The rapid advancements in microarray technology have enabled high-throughputs of genome–wide measurements of gene transcription level. We apply independent component analysis (ICA) as a versatile unsupervised approach for microarray analysis, and evaluate its performance against other leading unsupervised methods such as clustering and support vector machines (SVM).

A-49

A Method of Inferring Molecular Relationships from a Metabolic Database

Daniel McShan[1]
[1]Daniel.McShan@uchsc.edu, University of Colorado School of Medicine

This poster will discuss the automated inference of a molecular ontology from the KEGG Ligand database. We infer compositional (has-a) relationships not explicity annotated in any molecular database. While this current investigation is limited in scope, it reveals several important insights into the process of extracting molecular relationships from molecular data.

A-50

Soft Clustering, Feature Selection, and Network Inference Using Continuous Models and Gaussian Processes

Christoph Best[1], Joannis Apostolaki[2], Ralf Zimmer
[1]christoph.best@ifi.lmu.de, LMU Bioinformatics; [2]joannis.apostolakis@ifi.lmu.de, LMU Bioinformatics

We investigate methods based on continuous models and Gaussian processes for the identification of regulatory motifs from gene expression arrarys and other data. A low-dimensional mapping representation is combined with stochastic Markov network inference to extract and model regulatory relationships.

A-51

Aiding Text Mining with Context

Robert[1]
[1]kueffner@informatik.uni-muenchen.de, LMU Munich

We aim to annotate the vast amounts of data generated by text mining approaches with contextual information to allow human experts to quickly find relevant results. Distance metrics are defined to compare objects within the scientific literature such as documents, journals and authors.

A-52

GoSurfer: Using Multiple Hypothesis Testing Methods in Finding Associations Between Gene Ontology Terms and Gene Sets

Sheng Zhong[1], Lu Tian, Cheng Li, Wing H Wong
[1]szhong@hsph.harvard.edu, Harvard University

GoSurfer is a visualization tool that maps user defined gene sets onto the Gene Ontology (GO) space and find the GO terms that are significantly enriched in one of the input gene sets. New statistical methodologies considering multiple comparison issues were equipped to GoSurfer. GoSurfer is available at http://www.gosurfer.org

A-53

MeSH Terms for Gene Expression Cluster Validation and Annotation

Andreas Rechtsteiner[1], Luis M Rocha[2]
[1]andreas@lanl.gov, Los Alamos National Lab; [2]rocha@lanl.gov, Los Alamos National Lab

The automated validation and functional annotation of gene expression cluster with MeSH key terms is illustrated. The MeSH term-gene associations of 3 co-expression clusters from herpes infected human fibroblast cell were analyzed with Latent Semantic Analysis Distinct functional themes were found to be associated with the clusters.

A-54

Validation of Biological Hypotheses Using Joint Analysis of Biological Networks And Expression Data

Sabine Trochim[1], Florian Sohler[2], Daniel Hanisch, Ralf Zimmer
[1]sabine.trochim@biosolveit.de, BioSolveIT GmbH, Sankt Augustin, Germany; [2]florian.sohler@ifi.lmu.de, Institut für Informatik, Ludwig-Maximilians-Universität München, Germany

Joint analysis of expression data and biological context information enables researchers to quickly assess the functional context and relevant interaction partners of significantly regulated genes. We present a case study based on the yeast compendium dataset showing how to use the algorithms implemented in ToPNet for validating biological hypotheses.

A-55

Query-Driven Biclustering of Microarray Data by Gibbs Sampling

Qizheng Sheng[1], Yves Moreau[2], Bart De Moor
[1]qizheng.sheng@esat.kuleuven.ac.be, ESAT-SCD, Katholieke Universiteit Leuven; [2]yves.moreau@esat.kuleuven.ac.be, ESAT-SCD, Katholieke Universiteit Leuven

We have extended our previous work on the Gibbs sampling biclustering algorithm to solve the query-driven biclustering problem for microarray data where the task is to recruit patients to a known set of patients who share the same pathological type while identifing the genes that are responsible for the grouping.

A-56

Identification of Human Fusion Gene Transcripts in Expressed Sequence Databases

BK Lee[1], Yoonsoo Hahn[2], Tapan K. Bera and Ira Pastan
[1]bk@nih.gov, NIH; [2]hahny@pop.nci.nih.gov, NIH

Chromosomal rearrangements resulting in gene fusion are frequently involved in carcinogenesis. We developed a semi-automatic procedure for identifying fusion gene transcripts and collected 60 known and 177 new fusions by using public Databases. The predicted IRA1/RGS17 fusion was experimentally confirmed, demonstrating that expressed sequence Databases can be used to discover chromosomal aberrations.

A-57

Using an Ensemble of Neural Network Classifiers for the Recognition of E.Coli Promoters in Gene Sequences

Romesh Ranawana[1], Vasile Palade[2]
[1]romesh.ranawana@comlab.ox.ac.uk, University of Oxford Computing Laboratory;
[2]vasile.palade@comlab.ox.ac.uk, University of Oxford Computing Laboratory

We present a neural network based multi-classifier system for the identification of Escherichia coli (E.Coli) RNA polymerase promoter sequences in strings of DNA. Four neural networks were trained on identical training data encoded differently and the results of the networks were combined using a combinatorial function. The resulting system exhibited a precision of 0.9818.

A-58

Computing an Optimal Set of Domain Classes for a Set of Proteins

Gulriz Aytekin-Kurban[1]
[1]gulriz@cs.uchicago.edu, The University of Chicago

On a multi-domain protein, the domains that are evolutionary units usually have distinct sets of homologous proteins. For a set of multi-domain proteins with homologous domains and their alignments with a large set of non-redundant database proteins, we compute the underlying domain units that are likely to generate the alignments

A-59

Integrating the Gene Ontology into the Unified Medical Language System

Jane lomax[1], Alexa McCray[2]
[1]jane@ebi.ac.uk, EMBL-EBI; [2]mccray@nlm.nih.gov, National Library of Medicine

The Gene Ontology (GO), a set of controlled vocabularies used for gene product annotation, has recently been integrated into the National Library of Medicine's Unified Medical Language System (UMLS). All GO terms now have a corresponding UMLS concept, and the mapping is available through the UMLS Knowledge Source Server.

A-60

A Fuzzy Clustering Module For SCIpath: Exploring Syn-Expression

Amos Folarin[1], Sylvia Nagl[2]
[1]a.folarin@medsch.ucl.ac.uk, UCL;
[2]nagl@pat.biochem.ucl.ac.uk , UCL

Partitioner is a java based application which implements the Fuzzy k-means algorithm. The output of Partitioner can subsequently be displayed in either the ClusterBrowser, an interactive viewer complete with Gene Ontology annotation or using ClusterMapper to visualise the clustered data in the framework of a biological pathway.

A-61

iHOP: a Gene Network for Navigating Through PubMed

Robert Hoffmann[1], Alfonso Valencia[2]
[1]hoffmann@cnb.uam.es, CNB; [2]valencia@cnb.uam.es, CNB

By employing genes and proteins as hyperlinks between sentences and abstracts, the information in PubMed can be converted into one navigable resource. iHOP is an online service that provides this gene guided network as a natural way of accessing the more than ten million abstracts in PubMed and brings all the advantages of the internet to scientific literature research.

A-62

Integrating the Gene Ontology into the Unified Medical Language System

Jane lomax[1], Alexa McCray[2]
[1]jane@ebi.ac.uk, EMBL-EBI; [2]mccray@nlm.nih.gov, National Library of Medicine

The Gene Ontology (GO), a set of controlled vocabularies used for gene product annotation, has recently been integrated into the National Library of Medicine's Unified Medical Language System (UMLS). All GO terms now have a corresponding UMLS concept, and the mapping is available through the UMLS Knowledge Source Server.

A-63

Graph Computation Applied to Cancer Microarrays

Desmond J Higham[1], Keith Vass [2]
[1]djh@maths.strath.ac.uk, University of Strathclyde; [2]k.vass@beatson.gla.ac.uk, The Beatson Institute for Cancer Research

We outline a graph methodology for processing gene expression data coming from a range of experiments The graph structure makes the data amenable to modern computational algorithms that reveal clustering and ordering information Initial results are given for some oral cancer data

A-64

A New Method for Comparing Results from Gene Expression Data Clustering

Aurora Torrente[1], Misha Kapusheski[2], Alvis Brazma
[1]aurora@ebi.ac.uk, EMBL Outstation, EBI; [2]ostolop@ebi.ac.uk, EMBL Outstation, EBI

Application of different clustering methods on the same data often produces different results and it is important to understand how they relate to each other for biological interpretation of the data. We present a clustering comparison method and its implementation that finds correspondences between clusters on both sides, approximates 'true' underlying clusters and restores biologically meaningful clusters.

A-65

A System for Extracting Gene Relationships from Biomedical Literature, Generating Pathways and Output Files in Systems Biology Markup Language (SBML) Format

Mustafa Dameh[1], Jo-Ann Stanton[2], Peter Whigham, David Green
[1]mustafa.dameh@anatomy.otago.ac.nz, University of Otago, Departments of Anatomy and Systems Biology, and Information Science; [2]jo.stanton@anatomy.otago.ac.nz, University of Otago, Departments of Anatomy and Systems Biology

We use text mining techniques to collect biological information from the literature describing interactions between genes. Our system produces graphical representations in real time, which can be presented textually as a Systems Biology Markup Language (SBML). The system has the potential for widespread application in understanding gene product relationships.

A-66

CUREOS: Data-mining for Conserved TFBSs

Suzanne Thomas[1], Tim Beissbarth[2], Joelle Michaud, Tony Papenfuss, Hamish Scott
[1]sthomas@wehi.edu.au, WEHI; [2]beissbarth@wehi.edu.au, WEHI

CUREOS is a database of conserved transcription factor binding sites. It has a user-friendly interface and will be freely available on the web in 2004. CUREOS can be used to find patterns in genetic regulation and therefore aid the researcher in elucidating regulatory genetic networks and hypothesise related physiological networks.

A-68

Discovering Related Conceptual Terms from Scientific Texts: a Methodology to Assist Ontology Creation

M. Chagoyen[1], P. Carmona[2], J.M. Carazo, A. Pascual-Montano
[1]monica.chagoyen@cnb.uam.es, Centro Nacional de Biotecnologia; [2]pcarmona@cnb.uam.es, Centro Nacional de Biotecnologia

We propose the use of the non-negative matrix factorization algorithm to the analysis of relevant scientific texts in order to automatically discover related terms and concepts. The methodology is specially indicated in those domains where concepts should be described in a wide context or across traditionally independent disciplines.

A-69

Annotating Microarray Reporter Sequences to Well-Described Genes

Chris Evelo[1], Stan Gaj[2], Joris Korbeeck, Willem Ligtenberg, Eduard ter Voert, Rachel van Haaften
[1]chris.evelo@bigcat.unimaas.nl, BiGCaT Bioinformatics; [2]stan.gaj@bigcat.unimaas.nl, BiGCaT Bioinformatics

Microarray reporter sequences are normally annotated through automated BLAST searches against nucleotide Databases. Using a PERL based UNIGENE BLAST output parser combined with a Data Mining procedure through database crosslinks we managed to select high quality alignments with full coding sequences and to relate those to well described proteins wherever possible.

A-70

Mining Value From Gene Expression Data

Davina K Button[1], Kevan M. A. Gartland[2], Leslie D. Ball, Louis Natanson, Jill S. Gartland, Peter Ghazal4, Ishbel Duncan5, Adrian C. Newton, Bruce Marshall and Gary D. Lyon
[1]davina.button@abertay.ac.uk, University of Abertay Dundee; [2]k.gartland@abertay.ac.uk, University of Abertay Dundee

DRASTIC is a relational database for gene expression developed by University of Abertay Dundee and Scottish Crop Research Institute to record molecular plant responses to treatments. The INSIGHT Project is focused on developing a suite of tools to intelligently mine the database. Techniques used include Data Mining , clustering and visualisation. INSIGHT tools are available at http://www.drastic.org.uk

A-71

Classification from Full Text: A Comparison of Canonical Sections of Scientific Papers

Gail Sinclair[1], Bonnie Webber[2]
[1]csincla1@inf.ed.ac.uk, School of Informatics, University of Edinburgh; [2]bonnie@inf.ed.ac.uk, School of Informatics, University of Edinburgh

Accelerating growth in biomedical literature has stimulated activity on automated classification from this literature. This work attempts to improve on an earlier study associating biological articles to GO codes. It demonstrates the need for more access to full text articles and for the use of Part-of-Speech tagging.

A-72

Whatizit - A Server Pipeline on a Linux Cluster for Text Annotation in the Biomedical Domain

Harald Kirsch[1], Sylvain Gaudan[2], Dietrich Rebholz-Schuhmann
[1]kirsch@ebi.ac.uk, EMBL-EBI; [2]gaudan@ebi.ac.uk, EMBL-EBI

Whatizit is a set of modules distributed on several Linux servers for instantaneous annotation of biomedical text. Whatizit identifies biological facts in scientific text such as: UniProt concepts, mutations and protein-protein interactions. Wherever possible concepts are automatically linked to the original source.

A-73

Integrating and Visualising Text Mining Services in Biomedical Workflow Environments

Rob Gaizauskas[1], Neil Davis[2], George Demetriou, Yikun Guo and Ian Roberts
[1]R.Gaizauskas@dcs.shef.ac.uk, Department of Computer Science, University of Sheffield; [2]ndavis@dcs.shef.ac.uk, Department of Computer Science, University of Sheffield

Biomedical workflow environments allow researchers to carry out complex tasks in order to collect, analyse and mine information from a variety of disparate and heterogeneous resources. We describe how text mining services are integrated in a workflow designed to support in-silico experiments relating to the genetic bases of Graves' Disease and Williams Syndrome

A-74

The LANL BioCreAtIvE Task 2 Submission

Karin Verspoor[1], Judith Cohn[2], Cliff Joslyn, Sue Mniszewski, Andreas Rechtsteiner, Luis M. Rocha, Tiago Simas
[1]verspoor@lanl.gov, Los Alamos National Laboratory; [2]jcohn@lanl.gov, Los Alamos National Laboratory

We describe our submission to BioCreAtIvE, addressing automatic annotation of a protein into the Gene Ontology based on a given publication. We categorized terms derived from the document neighborhood of the protein into nodes in GO based on overlaps with terms on GO nodes and additional terms identified as related.

A-75

Graph-based Clustering of Biological Data

Aynur Dayanik[1], Craig G. Nevill-Manning[2]
[1]aynur@cs.rutgers.edu, Rutgers University;
[2]craignm@google.com, Google

We describe a novel method for clustering biological data by exploiting the interlinked structure of biological Databases. First, we construct a network of sequence-structure-literature with pairwise relationships. Then we apply graph partitioning techniques to infer clusters of related articles, sequences and structures. We found that the resulting clusters exhibit strong topicality

A-76

HIV Protease Cleavage Specificity: Model Complexity, Prediction Accuracy and Biological Rule Extraction

Liwen You[1], Thorsteinn Rögnvaldsson[2]
[1]liwen.you@ide.hh.se, Intelligent Systems Lab., School of Information Science, Computer and Electrical Engineering, Halmstad University; [2]denni@ide.hh.se, Intelligent Systems Lab., School of Information Science, Computer and Electrical Engineering, Halmstad University

The HIV protease cleavage prediction problem was studied on a new data set with 691 octamers. We found that a linear support vector machine and a simple perceptron outperforms other algorithms on this problem. We show the prediction performance from different models and extract cleavage specificity rules from the data set.

A-77

Historically Assessing the Validity of Molecular Interaction Prediction Algorithms

Rajan M. Lukose[1], Eytan Adar[2], Main-San Chan
[1]lukose@hpl.hp.com, HP Labs; [2]eytan@hpl.hp.com, HP Labs

For the yeast protein interaction network, we construct a novel database representing knowledge of the network through time. We test new algorithms designed to predict interactions solely from information known at the time. These algorithms work well, suggesting that the algorithms and methodology are useful for biological prediction in general.

A-78

Bioinformatic Analyses of Differentially Expressed Novel Psychosis-inducing Genes

Catherine Winchester[1], Anne Marquet[2], Lucy O'Donovan, Pawel Herzyk, Judy Pratt and Brian Morris
[1]cwinches@udcf.gla.ac.uk, University of Glasgow;
[2]0311554m@student.gla.ac.uk, University of Glasgow

Differentially expressed novel psychosis-inducing genes have been identified from a genomic screen utilising our rat model of schizophrenia and GeneChips from Affymetrix. We present our strategy for confirming the identity of the Affymetrix rat probe sets, identifying mouse and human orthologues, genomic locations and functional information using open source software

A-79

Dynamic Time Warping for Aligning Chromatograms in a Quantitative Proteomics Experiment

P.D. Moerland[1], J.W. Back[2], D. Speijer and A.H.C van Kampen
[1]p.d.moerland@amc.uva.nl, Bioinformatics Laboratory, Academic Medical Center, University of Amsterdam;
[2]jwback@science.uva.nl, Swammerdam Institue for Life Sciences, Mass Spectrometry Group, University of Amsterdam

Quantification of protein abundance by inverse labeling consists of two converse labeling experiments. Resulting chromatograms are generally shifted and stretched with respect to each other. We describe a method for aligning chromatograms such that replicate estimates of protein abundance can be matched. Alignments are validated in a large-scale proteomics experiment.

B-1

Sorting Points Into Neighborhoods(SPIN)

Ilan Tsafrir[1], Dafna Tsafrir[2], Liat Ein-Dor, Or Zuk and Eytan Domany
[1]tsafriri@wisemail.weizmann.ac.il, Weizmann Institute of Science; [2]Fedafna@wisemail.weizmann.ac.il, Weizmann Institute of Science

We present SPIN, a novel unsupervised method for organization and visualization of data. SPIN utilizes traits of distance matrices to sort objects in a natural ordering that highlights the underlying structure of the original, multidimensional data. We show the high performance of SPIN in the analysis of microarray experiments.

B-2

VistaClara: An Interactive Information Visualization Approach to Text Mining Microarray Gene Annotations

Robert Kincaid[1], Aditya Vailaya[2]
[1]robert_kincaid@agilent.com, Agilent Laboratories; [2]aditya_vailaya@agilent.com, Agilent Laboratories

VistaClara is an interactive visualization for exploratory microarray analysis that includes text mining annotations of user-selected sets of genes. Phrase or word-based statistical analysis can be applied to any text annotation (Gene Ontology, LocusLink summaries, pathway membership, etc.). Statistically over/under-represented entries provide biological insight within a highly interactive exploratory environment

B-3

PfamAlyzer: Exploring Pfam Domain Architectures

Volker Hollich[1], Erik L.L. Sonnhammer[2]
[1]Volker.Hollich@cgb.ki.se, CGB, Karolinska Institutet; [2]Erik.Sonnhammer@cgb.ki.se, CGB, Karolinska Institutet

PfamAlyzer provides means for seeking specific domain architectures within Pfam. Arbitrary domains can be combined freely, optionally with gaps. The query may be limited to taxonomic groups. PfamAlyzer also provides innovative homology searching by first predicting Pfam domains from the sequence and subsequently domain-querying SWISS-PROT. PfamAlyzer is available at http://pfam.cgb.ki.se/pfamalyzer

B-4

TraceScan: A Novel Tool to Assist in the Discovery of Mutation Sites and Base Miscalls

Marc Cooper[1], Patricia Evans[2], Martin Lague, Rebecca Griffiths, David De Koeyer, Sharon Regan, Virendra Bhavsar, Barry Flinn
[1]marc.cooper@unb.ca, University of New Brunswick; [2]pevans@unb.ca, University of New Brunswick

TraceScan is a visualization tool for sequence chromatograms that analyzes the impact of substituting base calls within tracefiles where overlapping peaks exist for called bases. TraceScan assists in the discovery of candidate mutation sites and erroneous base calls using local alignments of homologous sequences and the visualized trace.

B-5

WAViS Server for Visualisation of Alignments of Biological Sequences

Radek Zika[1], Jan Paces[2]
[1]zikar@img.cas.cz, Institute of Molecular Genetics; [2]hpaces@img.cas.cz, Institute of Molecular Genetics

Web Alignment Visualisation Server contains a set of web-tools designed for quick generation of publication-quality color figures of multiple alignments of nucleotide or amino acids sequences. It can be used for identification of conserved regions and gaps within many sequences using only common web browser

B-6

SABIO: a System to Support the Analysis of Biochemical Pathways

Isabel Rojas[1], Ulrike Wittig[2], Renate Kania, Esther Ratsch, Andreas Eberhart, Olga Krebs
[1]Isabel.Rojas@eml-r.villa-bosch.de, EML Research gGmbH; [2]Ulrike.Wittig@eml-r.villa-bosch.de, EML Research gGmbH

The SABIO (system for the analysis of biochemical pathways) was developed to support researchers in the analysis of biochemical reactions. SABIO was conceived to function as a workbench for the storage and querying of information related to metabolic reactions, general information as well as organism specific and experimental.

B-7

WebLab: A User-Friendly Bioinformatics Analysis Platform for Work Groups

Jianmin Wu[1], Ge Gao[2], Di Liu, Lei Kong, Zhe Li
[1]*wujm@mail.cbi.pku.edu.cn, CBI, Peking Univ;*
[2]*gaog@mail.cbi.pku.edu.cn, CBI, Peking Univ*

WebLab, a protocol-based analysis platform with a user-friendly interface was developed using XML for program parameters and workflows. The applications are invoked with a job manager, and the data and analysis results are stored in a database and could be shared by users in the same group.

B-8

iCompare: a Genome Browser for Displaying Comparative Genomics Annotation

Syam S Tatineni[1], S Ribrioux[2], Adrian Bruengger, Markus John.
[1]*syam_s.tatineni@pharma.novartis.com, Novartis Pharmaceuticals;*
[2]*sebastien.ribrioux@pharma.novartis.com, Novartis Pharmaceuticals*

The iCompare genome browser is a web-based visualization tool displaying annotation tracks from different sources and species based on coordinate, keyword, and sequence searches. Its focus is on displaying comparative genomics annotations. iCompare accesses a generic Oracle database holding the annotation tracks and utilizes scalable vector graphics (SVG) for visualization.

B-9

CRAVE: Visualization of Phenotype Ontologies

G.V.Gkoutos[1], E.Green[2], A. Blake, S. Greenaway, A.-M. Mallon, J.M. Hancock, D. Davidson
[1]*g.gkoutos@har.mrc.ac.uk, Medical Research Council ;*
[2]*e.green@har.mrc.ac.uk, Medical Research Council*

Concept Relation Assay Value Explorer (CRAVE) is a visualization tool for viewing Phenotype Ontologies created according to recent proposals. CRAVE is available online at: www.mgu.har.mrc.ac.uk/servlet/browser.frameset

B-10

UVWORD-G: A Graphical Tool for the Visualization of Word Counts in DNA Sequences

José Vicente Martínez[1], Vicente Arnau[2], Ignacio Marín
[1]*, Universidad de Valencia;* [2]*, Universidad de Valencia*

We describe a JAVA program that determines all words of up to 14 nucleotides present in DNA sequences. It allows a varied repertoire of analyses, from descriptive studies of a sequence to comparative analyses between sequences. The program has a variety of graphical outputs that allow to comfortably visualize the results.

B-11

Feature Plots: Visualizing System Dynamics in Cyclic Processes

Anders Fausbøll[1], Thomas Skøt Jensen[2], Ulrik de Lichtenberg, Soren Brunak
[1]*fausboll@cbs.dtu.dk, Center for Biological Sequence Analysis, Technical University of Denmark;*
[2]*skot@cbs.dtu.dk, Center for Biological Sequence Analysis, Technical University of Denmark*

Networks of thousands of genes and proteins are becoming available. Software exists for laying out static network graphs, but less has been done for visualizing dynamics in a system-wide manner. We introduce protein feature plots producing graphical views of system dynamics that can be readily inspected by the human eye.

B-12

ArrayXPath: Mapping and Visualizing Microarray Gene Expression Data with Integrated Biological Pathway Resources using Scalable Vector Graphics

Hee-Joon Chung[1], Min goo Kim[2], Chan Hee Park, Jihoon Kim, Ju Han Kim
[1]*joonny96@snu.ac.kr, SNUBI: SNUBiomedical Informatics ;* [2]*satyrs1@snu.ac.kr, SNUBI: SNUBiomedical Informatics*

ArrayXPath is a web-based service for mapping and visualizing microarray gene expression data for integrated biological pathway resources using Scalable Vector Graphics (SVG). By integrating major bio-Databases and searching pathway resources, ArrayXPath automatically maps different types of identifiers from microarray probes and pathway elements. ArrayXPath is available at http://www.snubi.org/software/ArrayXPath

B-13

Comparative Genomics: Viewing the Whole Ensembl

K. Cara Woodwark[1], James Smith[2], James Stalker, Jessica Severin, Brian Gibbons, Roger Pettett, Abel Ureta-Vidal
[1]cara@ebi.ac.uk, E.B.I.; [2]js5@sanger.ac.uk, Sanger Institute

The new Ensembl Comparative genomics viewer will be presented, as well as extensions of the current genome specific views. These will include multiple sequence alignments, expanded whole genome comparisons and pairwise species genome views combined with putative protein orthologues.

B-14

Managing and Exploring Image Based Data

Kort, Alexander[1], Kreutter, Stefan[2], Arnold, Jungmann, Vogt, Schwarz
[1]alexander.kort@fit.fraunhofer.de, Fraunhofer FIT; [2]stefan.kreuttert@fit.fraunhofer.de, Fraunhofer FIT

We have realised a prototype for managing heterogeneous, semi-structured image-based experiment data using XML and web services. The prototype integrates visual Data Mining, automatic view generation, advances image inspection techniques and annotation management. Integrated into a lab workflow this prototype is used for several high content screening applications.

B-15

A Perl Library for Generating Hierarchical, Tree-based Browser Interfaces for SCOP Database Searches

Thomas P. Walsh[1], Geoffrey J. Barton[2]
[1]tom@compbio.dundee.ac.uk, School of Life Sciences, University of Dundee, Scotland, UK; [2]geoff@compbio.dundee.ac.uk, School of Life Sciences, University of Dundee, Scotland, UK

A Perl library has been developed that creates interfaces for displaying the results of SCOP database searches in a tree-based hierarchical interface that reflects the SCOP classification. The library is designed to be easily extended to generate interfaces for fold recognition or structure comparison programs.

B-16

ZigZag for BioInformatics

Adam Moore[1], Ted Nelson[2], Tim Brailsford, Helen Ashman
[1]axm@cs.nott.ac.uk, Web Technologies Group, School of Computer Science, University of Nottingham; [2]tandm@xanadu.net, Oxford Internet Institute, University of Oxford

Bioinformatics data comes from a variety of sources, in a plethora of formats and conventions. ZigZag is a unique hyperstructural space, allowing data to exist simultaneously in many different contexts. A preliminary ZigZag structure has been created, demonstrating a space that simultaneously contains chemical, structural, bibliographic and reaction pathway data.

B-17

Attaching Biology to the Genome Using Ontologies

Norberto B. Delacruz[1], Jedidiah Mathis[2], Mary Shimoyama, Dean Pasko, Susan Bromberg, Simon Twigger, Jeff Nie, Anne Kwitek, Victoria Petri, Weiye Wang, Peter Tonellato, Howard Jacob
[1]noriede@mcw.edu, Rat Genome Database, MCW; [2]jmathis@mcw.edu, Rat Genome Database, MCW

The focus of biological research will increasingly turn to understanding the biological implications of genomic data. RGD developed a system of ontology annotations laid down as Genome Browser tracks. This will help biologists visualize and integrate of biological phenomena related to particular stretches of the genome.

B-18

Self-organizing Maps for Visualizing Codon Usage Data

Jose A. Neme[1], Pedro Miramontes[2]
[1]neme@uxmcc2.iimas.unam.mx, IIMAS, UNAM; [2], Fac. Ciencias, UNAM

Distributions obtained by self-organizing maps (SOM) show that codon bias may be explaided by CpG levels. Also, more evidence is given about the fact that G+C content and biological domain are causes of codon bias. SOM is a better multidimensinal data analysis tool because it preserves local topology and visualization obtained by it is more intrepretable.

B-19

Analysis of DNA Microarray Data by using Spherical Self-Organizing Maps

Daisuke Nakatsuka[1], Tomoyuki Kato[2], Heizo Tokutaka, Shinya Urase, Naruto Ohta, Yasaburo Matsuura, Kikuo Fujimura, Yasushi Kawata, Masaaki Ohkita,
[1]tokutaka@ele.tottori-u.ac.jp , Tottori University; [2]tonann815@yahoo.co.jp, Tottori University

By the Self-Organizing Maps (SOM), various multi-dimensional data can be clustered. Moreover, by making the component maps, it can be known what kind of influence any arbitrary dimensional component can give for clustering. Here, a new analytical technique which uses a spherical SOM is proposed and used for DNA microarray data.

B-20

Variable Selection from Principal Components

Christopher Bowman[1], Richard Baumgartner[2], Ray Somorjai
[1]Christopher.Bowman@nrc-cnrc.gc.ca, National Research Council; [2]Richard.Baumgartner@nrc-cnrc.gc.ca, National Research Council

We apply to microarray data a method for feature selection based on principal component analysis. The k features selected by the algorithm span a subspace that matches as closely as possible the space spanned by the first few principal components. The method is applied and validated on publicly availably microarray data.

B-21

MAGIC Gene Discovery: Tools for EST/Genomic Sequencing and Expression Analysis

Marie-Michele Cordonnier-Pratt[1], Lee Pratt[2], Haiming Wang, Chun Liang, Feng Sun, Dmitri Kolychev, Robert M. Freeman, Jr.
[1]mmpratt@uga.edu, University of Georgia; [2]leepratt@uga.edu, University of Georgia

MAGIC Gene Discovery is the DNA-sequencing half of a modular approach to a genomic, integrated and comprehensive resource for gene discovery and expression. Focusing on expressed sequence tags, SNPs and microsatellites, it also handles genomic and full-length cDNA sequencing and automated annotation via BLAST. Additional information is at http://fungen.org.

B-22

ScanView: a PROSITE Motif Hits Visualizer that Maps Functional and/or Structural Residues

Edouard de Castro[1], Christian J.A. Sigrist[2], Nicolas Hulo, Alexandre Gattiker, Elisabeth Gasteiger, Amos Bairoch
[1]ecastro@isb-sib.ch, Swiss Institute of Bioinformatics; [2]Christian.Sigrist@isb-sib.ch, Swiss Institute of Bioinformatics

Protein sequences can be scanned for the presence of specific PROSITE motifs using the ExPASy web tool ScanProsite. Here we present the updated scanprosite results viewer: ScanView that provides an interactive hit viewer and feature detector/viewer in both text and graphical forms, allowing easier analysis of ScanProsite results.

B-23

Microbial Genome Viewer

Robert Kerkhoven[1,1], Frank H.J. van Enckevort[1,2,3 ,2], Jos Boekhorst[1], Douwe Molenaar[2,3], and Roland J. Siezen[1,2,3]
[1]R.Kerkhoven@cmbi.kun.nl, [1]Centre for Molecular and Biomolecular Informatics, University of Nijmegen, The Netherlands; [2]Frank.van.Enckevort@cmbi.kun.nl, [2]NIZO food research, Ede, The Netherlands; [3]Wageningen Centre for Food Sciences, Wageningen; The Netherlands.

We present a web-based visualization tool, the Microbial Genome Viewer (www.cmbi.kun.nl/MGV), which enables interactive generation of chromosome wheels and linear genome maps from Genome Annotation data. The high-quality images are in Scalable Vector Graphics format. Gene-related data such as transcriptome and time-course microarray experiments can be superimposed on the maps.

B-24

HeatMapper and GO-StarTree: New Visualisations for the Understanding of Large Biological Datasets

Michael Moorhouse[1], George[2], Peter Valk, Roel Verhaak, Ruud Delwel, Peter van der Spek
[1]m.moorhouse@erasmusmc.nl, Erasmus MC; [2]g.garinis@erasmusmc.nl, Garinis

HeatMapper and GO-StarTree are tools that help biologists visualise large datasets. This first presents the correlation of genotypic data (e.g. microarray experiments) next to phenotypic data (e.g. molecular/ clinical data). The second is an interactive Gene Ontology viewer that preserves the context of each GO term. See also: www.erasmusmc.nl/bioinformatics/ mjmoorhouse/ismb2004index.html

B-25

Ongoing Development of GOPArc: Mapping Proteome and Expression Data onto Metabolic Pathways

Sebastian Oehm[1], Daniela Bartels[2], Alexander Goesmann, Joern Kalinowski, Folker Meyer
[1]Sebastian.Oehm@CeBiTec.Uni-Bielefeld.DE, Center for Biotechnology, University Bielefeld; [2]Daniela.Bartels@CeBiTec.Uni-Bielefeld.DE, Center for Biotechnology, University Bielefeld

Genome Annotation and expression data analysis is greatly facilitated by views integrating different kinds of related genome data. We have extended our GOPArc software to enable mapping of proteome and expression data onto metabolic pathways. Sample mappings of expression data from Corynebacterium glutamicum grown on glucose and acetat are shown.

B-26

Genome Visualization Support for Analysis of Diversity and Vaccine Development for HIV-1

Anelda Boardman[1], Lincoln Stein[2], Winston Hide
[1]anelda@sanbi.ac.za, South African National Bioinformatics Institute, University of the Western Cape, Cape Town, South Africa; [2]lstein@cshl.org, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, USA

Genome analysis through visualization will contribute to the understanding of HIV biology and epidemiology and speed up the discovery of vaccines for which a dire need exist. We have evaluated existing genome browsers for their applicability to visualize HIV genomes and annotations and developed a HIV Genome Browser prototype based on GBrowse.

B-27

Web Services Based Comparative Genomics

Mark Hoebeke[1]
[1]Mark.Hoebeke@jouy.inra.fr, INRA - MIA

In the field of comparative genomics, both sequence data and analysis tools accumulate at an alarming rate. We propose an extensible architecture giving acces to both of these resources through a Web services interface. We also added a module to our genome exploration client enabling it to access our repository through the Web services layer.

B-28

Ontoglyphs: A Language for Protein Function, Binding and Localization

Brigitte A.Tuekam[1], Christopher Hogue[2], Greg Pintilie, Rod Gonzaga.
[1]btuekam@blueprint.org, Blueprint Initiative; [2]chogue@blueprint.org, Blueprint Initiative

We have developed the OntoGlyphs, a computational biology visualization system. OntoGlyphs are graphical entities conveying information about a protein via attributes such as molecular function, biological process, and cellular localization as specified by Gene Ontology terms. The collection of glyphs consists of 34 functional, 25 binding and 24 location categories

B-29

Hyperbolic Visualization and Navigation of Gene Ontology Hierarchies

Prabhakara V. Choudary[1], Prashanti R. Srinivas[2], Fredric A. Gorin, and Edward G. Jones
[1]pvchoudary@ucdavis.edu, Department of Psychiatry & Behavioral Science;[2]Department of Neurology, Center for Neuroscience, University of California

We describe here a framework for simultaneous visualization of all three branches of gene ontology (GO), i.e., Biological Process, Molecular Function, and Cellular Component, as well as supporting evidence as icons/thumbnails, in hyperbolic space, and for intuitive navigation of the Focus+Context without losing orientation.

C-1

PIRSF Protein Classification System

Raja Mazumder[1], Winona C. Barker[2], Sehee Chung, Zhang-Zhi Hu, Darren Natale, Anastasia Nikolskaya, C. R. Vinayaka, Lai-Su L. Yeh, Cathy H. Wu
[1]rm285@georgetown.edu, Protein Information Resource, Georgetown University; [2]wb8@georgetown.edu, Protein Information Resource, Georgetown University

The PIR-SF (SuperFamily) classification system provides classification of whole proteins and domains to reflect their evolutionary relationships. The PIRSF database is central to PIR/UniProt functional annotation of proteins and is accessible at http://pir.georgetown.edu/pirsf for report retrieval and sequence classification.

C-2

An Open-Source Gene-Centric Genomics Warehouse Platform

Hilmar Lapp[1], Serge Batalov[2], Keith Ching, David Block, John Walker, Andrew Su, John Hogenesch
[1]hlapp@gnf.org, Genomics Institute of the Novartis Research Foundation; [2]sbatalov@gnf.org, Genomics Institute of the Novartis Research Foundation

We present SymGene and SymAtlas (http://symatlas.gnf.org/), a devotedly gene-centric genomics warehouse that integrates annotation and genome location from various sources and presents a unified view on annotation alongside with gene-related functional datasets. The system uses meta-information to interpret and collapse redundant annotation, and allows one to quickly locate favorite genes.

C-3

SOI and Database Query

Shengqiang Shu[1], Chris Mungall[2], Suzanna Lewis, Gerry Rubin
[1]sshu@fruitfly.org, BDGP; [2]cjm@fruitfly.org, BDGP

Sequence ontology instantiation (SOI) can be created from features and their relationships in a chado database. SQL query using SOI is uniform and fast. Two lightweight SOI modules can construct any data model from a chado using same one SQL statement. They have excellent query performance.

C-4

ISV db: Integrated Splicing Variants Database

Gloria Fu[1], Ueng-Cheng Yang[2]
[1]g39103016@ym.edu.tw, Institute of Biochemistry, Bioinformatics Program, National Yang-Ming University, Taipei, Taiwan; [2]yang@ym.edu.tw, Institute of Bioinformatics, Bioinformatics Research Center, National Yang-Ming University, Taipei, Taiwan

ISVdb integrates splicing variants information with gene features. The users may thus explore the effects of gene features, such as SNP, on alternative splicing. Alternatively, the users may explore the effects of alternative splicing on gene features, such as protein domains. A user-friendly interface was implemented for such purpose.

C-5

LIMSTILL Platform: Managing High-throughput Reverse-genetics Screens in Model Organisms

Victor Guryev[1], Eugene Berezikov[2], Ronald Plasterk, Edwin Cuppen
[1]guryev@niob.knaw.nl, Hubrecht Laboratorium; [2]berezikov@niob.knaw.nl, Hubrecht Laboratorium

LIMSTILL (LIMS for Identification of Mutations by Sequencing and Tilling) is an open-source software designed to streamline the informatics and management parts of the Hubrecht Laboratory facility for high-throughput screening for mutants in model organisms (http://limstill.niob.knaw.nl). It includes steps for amplicon selection, primer design, sequence analysis, and annotation of mutations.

C-6

dbERGE II: Database of Experimental Results on Gene Expression

Prachi Shah[1], Belinda Giardine[2], Laura Elnitski, Cathy Riemer, Yi Zhang, Webb Miller and Ross Hardison
[1]pss11@psu.edu, The Pennsylvania State University; [2]giardine@bio.cse.psu.edu, The Pennsylvania State University

dbERGE II (http://www.bx.psu.edu) records experimental results from various gene expression assays. It features a user friendly, web-based query interface and is designed with a scalable, streamlined conceptual model. Moreover, it allows integration with the GALA database and the UCSC Genome Browser.

C-7

Mouse SAGE Site: Database of Mouse Digital Gene Expression Data

Petr Divina[1], Jiri Forejt[2]
[1]divina@biomed.cas.cz, Institute of Molecular Genetics, Academy of Sciences of the Czech republic, Centre for Integrated Genomics, Videnska 1083, CZ 142 20, Prague, Czech Republic; [2],

Mouse SAGE Site is a database of public available SAGE libraries generated from mouse tissues and cell lines The database aims to provide mouse geneticists with easy-to-use web tools to explore and analyze the mouse SAGE data with reliable tag-to-gene identification Mouse SAGE Site is freely accessible at http://mouse.biomed.cas.cz/sage/.

C-8

MATS-Pro (Management, Analysis, and Tracking System for Proteomics): An integrated Web-based Database System for Storage and Analysis of Proteomics Data

Amarendra Yavatkar[1], Catherine Campbell[2], Yang Fann
[1]yavatka@ninds.nih.gov, NINDS NIH and SRA International; [2]campbelc@ninds.nih.gov, NINDS NIH and SRA International

We have developed a robust, web based, relational database, MATS-Pro (Management, Analysis, and Tracking System for Proteomics) to store and analyze SELDI-TOF and MS-MS data. The system also has an analysis toolset to provide various statistical, Data Mining, and clustering methods allowing differentiation of protein profiles among groups.

C-9

Enabling Research with the Mouse Genome Informatics (MGI) System

Judith A Blake[1], Mary E. Dolan[2], James A. Kadin, Joel E. Richardson, David P. Hill, Janan T. Eppig, Harold J. Drabkin, Li Ni, Josh Winslow, Lori Corbani, Alex Diehl, Ben L. King, Cynthia Smith, Carroll W. Goldsmith, Carol J. Bult, Martin Ringwald and the Mouse Genome Informatics Group
[1]jblake@informatics.jax.org, The Jackson Laboratory; [2]mdolan@informatics.jax.org, The Jackson Laboratory

The Mouse Genome Informatics (MGI www.informatics.jax.org) system provides a comprehensive public resource about the laboratory mouse and includes multiple bio-ontologies that facilitate knowledge representations. We now provide access to software tools to facilitate the use of MGI resources in a research context particularly in respect to Gene Ontology (GO) annotations.

C-10

GCA, Gene Expression Database of DC Responses

Andrea Splendiani[1], Marco Brandizi[2], O. Beretta, M. Capozzoli, N. Pavelka, M. Pelizzola, C. Vizzardelli, F. Granucci, P. Castagnoli
[1]andrea.splendiani@unimib.it, Università di Milano Bicocca; [2]marco.brandizi@unimib.it, Università di Milano Bicocca

Genopolis database collects information on dendritic cells (DC, very important for immune response regulation) gene expression. Experiments may be annotated (MIAME web interface) in a supervised pipeline, by properly role-distinguished users. A flexible and pluggable search and graphical presentation facility allows for data retrieval and visualization. An advanced knowledge system for achieved results annotation is being designed.

C-11

Applications of SeqHound - a Biological Sequence and Structure Database Programming Platform

Katerina Michalickova[1], Renan Cavero[2], Zhe Wang, Michael Matan, Elizabeth Burgess, Kai Zheng, Victor Gu, Rachel Farrall, Hao Lieu, Rong Yao, Ian M. Donaldson, Christopher W.V. Hogue
[1]kmichali@blueprint.org, Blueprint, Samuel Lunenfeld Research Institute and Department of Biochemistry, University of Toronto; [2]rcavero@blueprint.org, Blueprint, Samuel Lunenfeld Research Institute

In the past, we introduced SeqHound (http://seqhound.blueprint.org) as an open source programming resource containing daily updated Entrez Databases, 3-D structural data and functional annotations. This resource is accessible remotely via APIs in C, C++, Java and BioPerl or locally when hosted in-house. We are presenting new additions and several published applications as examples of SeqHound use.

C-12

SigPep : A Database For Signal Peptides

Khar Heng Choo[1], Shoba Ranganathan[2], Tin Wee Tan
[1]justin@bic.nus.edu.sg, National University of Singapore; [2]shoba@bic.nus.edu.sg, Macquarie University

SigPep is a meta-database of information on prokaryotic and eukaryotic signal peptides found in the public Databases. The SigPep database provides one-click access to information on the signal peptide. SigPep is available at http://proline.bic.nus.edu.sg/sigpep

C-13

The Bioinformatics Knowledge Base

Adam Shlien[1], Sylvain Foisy[2], Maxime Caron, Jean-Philippe Laverdure, Enrique Madrid, Tien Duc Nguyen, Edith Ribourtout, Gertraud Burger
[1]adam.shlien@bioneq.qc.ca, BioneQ;
[2]sylvain.foisy@bioneq.qc.ca, BioneQ

The Bioinformatics Knowledge Base is a collaborative repository of bioinformatics ideas and practical know-how. Users can freely access the tool to gather new knowledge and add their own. http://apps.bioneq.qc.ca/twiki/bin/view/Knowledgebase

C-14

Drosophila melanogaster Exon Database

Bernett Teck Kwong Lee[1], Shoba Ranganathan[2], Tin Wee Tan
[1]bernett@bic.nus.edu.sg, National University of Singapore; [2]shoba@els.mq.edu.au, Macquarie University

DEDB is a database containing gene structure information derived from Genome Annotations available at the Berkeley Drosophila Genome Project. The gene structure information is transformed into splicing graphs, which forms the basis for alternative splicing classification as well as the visualization. DEDB is available at http://proline.bic.nus.edu.sg/dedb/index.html.

C-15

Online Predicted Human Interaction Database (OPHID): Exploring the Human Interactome

Kevin R. Brown[1], Igor Jurisica[2]
[1]kbrown@uhnres.utoronto.ca, Dept. of Medical Biophysics, University of Toronto; [2]juris@cs.toronto.edu, Dept of Medical Biophysics and Dept. of Computer Science, University of Toronto

OPHID is the Online Predicted Human Interaction Database, and contains over 42,000 human protein interactions (known and predicted). OPHID extends the known human interactome using model organism data, and is highly useable for both small- or large-scale analysis. OPHID is accessible at http://ophid.utoronto.ca.

C-16

PRECISE: PREdicted and Consensus Interaction Sites in Enzymes

Melissa R. Landon[1], Sandor Vajda[2], Shu-Hsien Sheu, David Lancia, Karl H. Clodfelter
[1]mlandon@bu.edu, BU Bioinformatics Program;
[2]vajda@bu.edu, BU Dept. of Biomedical Engineering

Recent computational efforts in the studies of protein-ligand interactions have focused on gaining insight into the relationship between enzyme structure and ligand recognition. To help address this problem, we have developed PRECISE, a public database of predicted and consensus interaction sites in enzymes, which displays in a graphical format the ligand interactions of a group of functionally-related enzymes.

C-17

Bridging Life Science Databases using Grid Technology for Genome-based Drug Discovery

Masato Kitajima[1], Takahiro Kosaka[2], Kazuto Yamazaki, Reiji Teramoto, Gen Kawamura, Susumu Date, Shinji Shimojo, Hideo Matsuda
[1]kita@ist.osaka-u.ac.jp, Osaka University;
[2]tak-k@ais.cmc.osaka-u.ac.jp, Osaka University

The completion of the Human Genome Project fueled an increase in the number of widely distributed life science Databases. Bringing all these together is of great value to genome-based drug discovery. Our approach of integrating these Databases using data grid technology proved to be successful in application to lead identification

C-18

GenSpector-LSDB(TM): Locus-Specific Database for a Genotyping Microarray

Taejoon Kwon[1]
[1]taejoon_kwon@samsung.com, Samsung Advanced Institute of Technology

GenSpector-LSDB is a locus-specific database that can be used in the analysis of genotyping microarray. It supports the data querying and retrieval with semantic web technology, and user-friendly web interface. To control the quality of data, it refers to a published data. In addition to the human loci, the pathogen loci can be stored in this system.

C-19

EMMA 2 - A MAGE-compliant System for Storage and Analysis of Microarray Data

Michael Dondrup[1], Alexander Goesmann[2]
[1]Michael.Dondrup@CeBiTec.Uni-Bielefeld.DE, CeBiTec; [2]Alexander.Goesmann@CeBiTec.Uni-Bielefeld.de, CeBiTec

EMMA is a web-based software system for management and analysis of transcriptomic data. EMMA 2 now supports the complete MAGE-ML format for data import and export. EMMA allows mapping of gene expression data onto proteome data or pathways and vice versa and provides extensible analysis and visualization Plug-Ins via the R-language.

C-20

The Biological Data Converter for Sharing Information among Heterogeneous Flat Files

Young Jin Jung[1], Hyo Sung Cha[2], Young Uk Kim, Keun Ho Ryu
[1]yjjeong@dblab.cbu.ac.kr, Chungbuk National University; [2]kkido@dblab.cbu.ac.kr, Chungbuk National University

In order to solve the problems about exchanging the biological information among heterogeneous flat files, we describe the BSML based transformation method utilizing XSL. When original data are updated, it need not modify a source program but alter the only mapping information and related XSL among various file formats.

C-21

PIA LIMS: a Laboratory Information Management System for High Throughput Projects Based on PLUK Technology

Sergey Vasiliev[1], Yannis Kalaidzidis[2], Hanjo Hennemann
[1]vasil@caesar.de, research center caesar; [2]yannis@computer.org, Belozersky Institute of Physico-Chemical Biology, MSU

PIA LIMS is an object oriented laboratory information management system based on PLUK software technology. We aim at the creation of a flexible and expandable system for high throughput applications involving the simultaneous tracking of thousands of objects (clones, plates etc.) and the control of various robotic devices.

C-22

SOPdb: An SOP Resource for the European Comprehensive First-line Phenotyping Protocol

E.C.J. Green[1], G.V.Gkoutos[2], J. Weekes, A.-M. Mallon, J.M. Hancock, S.D.M. Brown
[1]e.green@har.mrc.ac.uk, Medical Research Council; [2]g.gkoutos@har.mrc.ac.uk, Medical Research Council

EUMORPHIA aims at the development of new approaches in phenotyping, mutagenesis and informatics. Its focus is on the development, standardisation and dissemination of primary and secondary phenotyping protocols for all body systems in the mouse. SOPdb is an XML resource for accessing and storing European Comprehensive First-Line Phenotyping Protocols (ECFLPs) for mice.

C-23

POBS: A Database of Potential Protein Binding Sites in Proteins of Known Structure

Timothy E. Reddy[1], Charles Delisi[2]
[1]treddy@bu.edu, Boston University Bioinformatics Program; [2]delisi@bu.edu, Boston University Bioinformatics Program

The POBS database (http://bioinfo.bu.edu/pobs) is a large repository of computationally determined solvent accessible oligopeptide sequences for a wide range of proteins found in the Protein Data Bank (PDB). Among other potential uses, these sequences serve as a starting point in the determination of potential protein binding sites for the design of protein Microarrays.

C-24

UniProt: the Universal Protein Resource

Elisabeth Gasteiger[1], UniProt Consortium[2]
[1]Elisabeth.Gasteiger@isb-sib.ch, Swiss Instiitute of Bioinformatics; [2]SIB/EBI/PIR

UniProt (Universal Protein Resource) is the most comprehensive catalog of information on proteins, created by EBI, SIB and PIR. Its 3 components are optimized for different uses The UniProt Knowledgebase (Swiss-Prot and TrEMBL), the UniRef clusters of knowledgebase sequences sharing 100, 90 and 50% similarity, the UniParc archive, a comprehensive sequence repository

C-25

Human Endogenous Retroviral Families: HERV Database

Jan Paces[1], Radek Zika[2], Adam Pavlicek, Vaclav Paces
[1]hpaces@img.cas.cz, Institute of Molecular Genetics;
[2]zikar@img.cas.cz, Institute of Molecular Genetics

HERVd (Human Endogenous Retrovirus database) is a comprehensive database of distinct families of human endogenous retroviruses identified in the human genome. The database is accessible via WWW at http://herv.img.cas.cz and can be used for search according to many criterias.

C-26

BioinfoTools: a Kinase-oriented Database Integrated with a Collection of Tools for Data Manipulation

Luca Sartori[1], Roberta Bosotti[2], Marco Carreras, Eleonora Gianti, Emanuela Scacheri, Giorgio Ukmar, Simon Plyte, Antonella Isacchi
[1]luca.sartori@pharmacia.com, Pharmacia, Pfizer Group;
[2]roberta.bosotti@pharmacia.com, Pharmacia, Pfizer Group

We have developed an Oracle kinase-oriented database and interfaced it with a powerful search engine core, written in Borland Delphi It contains information on kinase sequences, classification, orthologs, chromosomal location, disease association, expression, IC50 data and also an inventory of reagents It has been integrated with classical Bioinformatic tools.

C-27

Mouse Promoter Database: a Database for the in Silico Analysis of Promoter

sooyoung cho[1], myungguen chung[2], younseek lee, sangsoo kim
[1]singylu@ihanyang.ac.kr, hanyang university;
[2]aobo@hanyang.ac.kr, hanyang university

We construct transcription factor binding site database for mouse genes. Transcription factor binding sites in 5′ upstream region of all mouse known genes, were obtained by using TRNASFAC. False positive is removed by simple method and CpG islands information were included in the database

C-28

BRENDA, the Enzyme Database

Christian Ebeling[1], Gregor Huhn[2], Antje Chang, Marion Gremse, Christian Heldt, Ida Schomburg, Dietmar Schomburg
[1]c.ebeling@uni-koeln.de, Institute for Biochemistry, University of Cologne; [2]Gregor.Huhn@uni-koeln.de, Institute for Biochemistry, University of Cologne

BRENDA represents a comprehensive collection of enzyme and metabolic information, based on primary literature. BRENDA includes biochemical and molecular information on classification and nomenclature, reaction and specificity, functional parameters, occurrence, enzyme structure, application, engineering, stability, disease, isolation and preparation, links and literature references.

C-29

Database and Tools for Analyzing Gene Regulation (BiGeR)

Hedi Peterson[1], Jaak Vilo[2]
[1]peterson@egeen.ee, University of Tartu, Riia 23, Tartu 51010, Estonia; [2]vilo@egeen.ee, Estonian Biocentre, Riia 23, Tartu 51010, Estonia

We are developing the database and tools to support the gene regulation studies. The aim is to combine previously known information from the literature, as well as data from various experiments or in silico predictions. Focus is on supporting in silico studies and to allow comparisons between different data sources.

C-30

The LCB Data Warehouse

Adam Ameur[1], Ola Spjuth[2], Jakub Orzechowski Westholm, Vladimir Yankovski, Jan Komorowski
[1]Adam.Ameur@lcb.uu.se, The Linnaeus Centre for Bioinformatics, Uppsala University; [2], Department of Pharmaceutical Biosciences, Uppsala University

The LCBDWH (The Linnaeus Centre for Bioinformatics Data Warehouse) at Uppsala University in Sweden is a web-based infrastructure for the collection and analysis of microarray expression data that provides an on-line service for the scientific community. The core of the system is BASE (Saal et al, 2002), a widely-used open source Laboratory Information Management System.

C-31

EMBL-Align: a Public Multiple Sequence Alignment Database

Robert J. Vaughan[1], Vincent Lombard[2], Alexandra E. van den Broek, Weimin Zhu, Rolf Apweiler
[1]vaughan@ebi.ac.uk, European Bioinformatics Institute; [2]lombard@ebi.ac.uk, European Bioinformatics Institute

The EMBL-Align database stores nucleotide and protein sequence alignments, providing alignment data published in scientific literature in a computer and human readable format for further analysis and peer review. Webin-Align is a dedicated web-based submission tool for submission to the EMBL-Align database http://www.ebi.ac.uk/embl/Submission/align_top.html.

C-32

InterPro3D

Ujjwal Das[1], Nicola Mulder[2], Rolf Apweiler and The InterPro Consortium
[1]ujjwal@ebi.ac.uk, EBI/EMBL; [2]nicky@ebi.ac.uk, EBI/EMBL

InterPro3D is a protein resource for the mapping of SCOP and CATH structural domains on UniProt protein sequences. InterPro3D uses the InterPro graphical interface showing the location of structural domains on a sequence through the mapping between the PDB/EMSD chain(s) and the protein sequence by EMSD. InterPro3D is a part of InterPro database and is available at http://www.ebi.ac.uk/interpro.

C-33

MELDB: A Clustered Protein Database for Microbial Esterases and Lipases

Ho-Young Kang[1], Cheol-Goo Hur[2], Eun Kyoung Jeon, Hyeon Su Ro, Jeong-Kee Lee, Tae- Kwang Oh, Jihyun F. Kim
[1]kangho@kribb.re.kr, Genome Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB); [2]hurlee@kribb.re.kr, Genome Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB)

MELDB is a comprehensive clustered protein database of microbial lipases and esterases which are hydrolytic enzymes very important in the modern industry. The proteins of MELDB are clustered into groups according to their sequence similarities based on local pairwise alignment and a graph clustering algorithm. The result is available at http://meldb.gem.re.kr.

C-34

XML Format of the UniProt Knowledgbase

Allyson Williams[1], Kai Runte[2], The UniProt Consortium
[1]allyson@ebi.ac.uk, European Bioinformatics Institute; [2]krunte@ebi.ac.uk, European Bioinformatics Institute

The UniProt Knowledgebase, composed of UniProt/Swiss-Prot and UniProt/TrEMBL, is the world's most comprehensive protein database UniProt XML provides an easily parsable database view, with the order and naming of elements based on the flat-file format. The Knowledgebase is updated fortnightly with ftp files produced in flat-file, FASTA, and XML format.

C-35

The Usefulness of Integrated Databases: a Case Study of COLUMBA

Stefan Günther[1], Kristian Rother[2], Silke Trissl
[1]stefanxguenther@web.de, Institute of Biochemistry, Charitè, Berlin, Germany; [2]kristian.rother@charite.de, Institute of Biochemistry, Charitè, Berlin, Germany

In our poster we want to show possible applications of COLUMBA by highlighting several aspects of the linkage quality in biological Databases. Moreover we used the database to generate sets of PDB entries containing Protein-DNA complexes and their homologues to conduct research on motives of DNA-binding sides in proteins.

C-36

COLUMBA - A Database of Annotated Protein Structures

Silke Trissl[1], Kristian Rother[2], Ulf Leser
[1]trissl@informatik.hu-berlin.de, Institute of Informatics, Humboldt-Universität zu Berlin, Germany; [2]kristian.rother@charite.de, Institute of Biochemistry, Charitè, Berlin, Germany

We present COLUMBA, a database of information on protein structures that integrates data from twelve different biological Databases, including ENZYME, KEGG, SCOP, CATH, DSSP, and SwissProt. COLUMBA allows for the quick computation of sets of protein structures that share interesting properties according to the different data sources.

C-37

TRANSPRO: A Repository of Gene Regulatory Sequences in Genomes

Klaus Hornischer[1], Alexander Kel[2], Olga Kel-Margoulis, Birgit Lewicki-Potapov, Volker Matys and Edgar Wingender
[1]klh@biobase.de, BIOBASE GmbH; [2]ake@biobase.de, BIOBASE GmbH

TRANSPRO is a database containing the nucleotide sequences of gene regulatory regions including extensive annotation. The database integrates data from TRANSGENOME, a genomic database which is filled and maintained by the extraction of various external Databases, with gene regulatory data from the TRANSFAC, TRANSCompel and SMARt Databases

C-38

ASD: the Alternative Splicing Database

T.A. Thanaraj[1], Stefan Stamm[2], Francis Clark, Jean-Jack Riethoven, Vincent Le Texier & Juha Muilu
[1]thanaraj@ebi.ac.uk, European Bioinformatics Institute, UK; [2]stefan@stamms-lab.net, University of Erlangen, Germany

ASD presents data on alternative splicing in human and other species as derived through (i) computational delineation using available transcript sequences; and (ii) collecting experimentally determined splice events from peer-reviewed journals. The reported splice events and isoform transcripts are annotated for various biological features. The data is available at http://www.ebi.ac.uk/asd/

C-39

A Knowledge Base for the Protein Localizome

Mitsuteru Nakao[1], Keun-Joon Park[2], Paul Horton
[1]nakao-mitsuteru@aist.go.jp, Computational Biology Research Center, National Institute of AdvanceIndustrial Science and Technology; [2]park-kj@aist.go.jp, Computational Biology Research Center, National Institute of Advance Industrial Science and Technology

We are constructing a novel knowledge base (CBRC Localizome Knowledge Base: CLKB), containing proteins with protein localization information confirmed by experiments and predicted by computational methods expressed in the GO, and allows users to select proteins with annotations fulfilling a desired evidence-level.

C-40

Cross-species Mapping between Anatomical Ontologies Based on Lexico-syntactic Properties

Sarah Luger[1], Stuart Aitken[2]
[1]sluger@inf.ed.ac.uk, XSPAN/University of Edinburgh; [2]stuart@inf.ed.ac.uk, XSPAN/University of Edinburgh

We establish alignments between model anatomical ontologies for different species (C.elegans, drosophila, zebrafish and mouse), using both structural and lexical clues. The analysis of both terminology and structure is pivotal to the task of aligning these anatomical ontologies (and their related GO codes). We propose novel, automated alignment methods.

C-41

Development of Public Literature Services at EBI

Peter Stoehr[1], Lichun Wang[2], Rodrigo Lopez, Weimin Zhu
[1]stoehr@ebi.ac.uk, EMBL-EBI; [2]lcwang@ebi.ac.uk, EMBL-EBI

Literature is a relatively new addition to the public bioinformatics services of the EBI. We describe existing resources based on MEDLINE and patent applications, and links to biological Databases. Future developments will include access to full-text, advanced text search facilities, and identification and mark-up of biological concepts in text.

C-42

PfamRDB: A Freely Available MySQL Relational Database for Protein Bioinformatics

Mhairi Marshall[1], Rob Finn[2], Alex Bateman
[1]mhairi.marshall@sanger.ac.uk, Wellcome Trust Sanger Institute; [2]rdf@sanger.ac.uk, Wellcome Trust Sanger Institute

Pfam is a large collection of protein domains and families. The website is run from a freely available MySQL relational database (PfamRDB). PfamRDB contains a wealth of information, including domains, motifs, active sites, structural and interaction data. It can be downloaded from the ftp site: ftp://ftp.sanger.ac.uk/pub/Databases/Pfam/database_files/ for local use.

C-43

Fetchprot

Kristofer Franzén[1], Patrik Hassel[2], Pär Lannerö, Björn Ursing
[1]franzen@sics.se, SICS; [2]Patrik.Hassel@metamatrix.se, Metamatrix AB

Fetchprot is a project with the goal to automatically find, assess, and gather information about proteins with experimentally verified functions, from scientific texts within the areas of molecular biology and bio-chemistry. This will be carried out by developing and applying language technology, and to build infrastructure and knowledge bases.

C-44

Storing and Retrieving Biological Instances with the Instance Store

Daniele Turi[1], Phillip Lord[2], Michael Bada, Robert Stevens
[1]dturi@cs.man.ac.uk, University of Manchester; [2]p.lord@russet.org.uk, University of Manchester

The instance store is a software component that addresses the difficulty of reasoning with instances in description logics. We have used it to represent and query the gene-product annotations of the GO database and to support a tool that aims to guide users in manual GO-term annotation of gene products.

C-45

KGML (KEGG Markup Language) for Exchanging the KEGG Graph Objects

Susumu Goto[1], Shuichi Kawashima[2], Toshiaki Katayama, Mika Hirakawa, Minoru Kanehisa
[1]goto@kuicr.kyoto-u.ac.jp, Kyoto University; [2]shuichi@kuicr.kyoto-u.ac.jp, Kyoto University

The KEGG Markup Language (KGML) is an exchange format of the KEGG graph objects, especially the KEGG pathway maps that are manually drawn and updated. KGML enables automatic drawing of KEGG pathways and provides facilities for computational analysis and modeling of protein networks and chemical networks.

C-46

An SRS Implementation of Biological Resource Centers Catalogues Available for Download

D. Marra[1], R. Lopez[2], F. Piersigilli, A. Manniello, P. Romano
[1]domen ico.marra@istge.it, National Cancer Research Institute; [2]rls@ebi.ac.uk, EBI

The SRS based CABRI "one-stop-shop" for biological materials (http://www.cabri.org/) includes information on more than 120.000 items from 28 collections. Information is periodically extracted and made available to interested public SRS services. An implementation of CABRI extracted Databases has been made available at the EBI SRS service (European Bioinformatics Institute, http://srs.ebi.ac.uk/).

C-47

GOnet: a Database of Protein-protein Interaction and Gene Ontology

Yimeng Dou[1], Pierre Baldi[2], Suzanne Sandmeyer
[1]ydou@ics.uci.edu, Univ of California, Irvine; [2]pfbaldi@ics.uci.edu, Univ of California, Irvine

We introduce the GOnet database and its related tools that incorporate current protein-protein interaction data, GO terms, level of GO term relatedness and the option of choosing the most reliable subset of interactions confirmed by various experiment methods. GOnet is available at http://contact14.ics.uci.edu/sgdindex or through the IGB website at http://www.igb.uci.edu/servers/Databases.html.

C-48

Rfam: Annotating Non-coding RNAs in Genomes

Simon Moxon[1], Sam Griffiths-Jones[2], Alex Bateman
[1]sm3@sanger.ac.uk, The Sanger Institute; [2]sgj@sanger.ac.uk, The Sanger Institute

Rfam is a collection of non-coding RNA families represented by multiple sequence alignments and covariance models. Release 5.0 contains 176 families. These families can be used to annotate sequence, including whole genomes, for the presence of non-coding RNA genes. Rfam is available at http://www.sanger.ac.uk/Software/Rfam/ and http://rfam.wustl.edu/.

C-49

BIOZON: a System for Unification, Management and Analysis of Heterogeneous Biological Data

Aaron Birkland[1], Golan Yona
[1]apb18@cornell.edu, Cornell University

Biozon is a unified biological database that integrates heterogeneous data types and the relationships between them into a single extensive schema. This schema allows one to see each data instance in its full biological context. More importantly it allows for complex searches that span multiple data types and for computations on that data.

C-50

PEPdbPub - a Database of Protist EST Sequences

Emmet A. O'Brien[1], Yue Zhang[2], Liusoing Yang, Eric Wang, B. Franz Lang and Gertraud Burger
[1]eobrien@bch.umontreal.ca, Universite de Montreal;
[2]yzhang@bch.umontreal.ca, Universite de Montreal

PEPdbPub is a database of EST sequences from a taxonomically wide range of protists, many not previously well characterised. We provide EST sequences and functional annotation. New organisms are added to the public dataset as the sequencing projects are completed.

C-51

The aMAZE Database Goes Public

Christian Lemer[1], Hassan Anerhour[2], Jesintha Mary Maniraja, Olivier Sand, Jean Richelle, Shoshana Wodak
[1]chris@amaze.ulb.ac.be, SCMBB;
[2]hassan@amaze.ulb.ac.be, SCMBB

The goal of the aMAZE project is to develop a workbench for the representation and analysis of networks of molecular interactions and cellular processes, genetic expression and regulation, enzymatic transformations and regulation, metabolic pathways, signal transduction, etc...

C-52

An Integrated Data Management System for Major Biological Databases

Hyeweon Nam[1], Daesang Lee[2], Pan-Gyu Kim, Kiejung Park
[1]hwnamyh@yahoo.co.kr, Chungnam National University; [2]dslee@smallsoft.co.kr, SmallSoft Co., Ltd.

We have developed WinBioDB with Windows interfaces, which includes importing modules and searching interfaces for 10 major public Databases such as GenBank, PIR, SwissProt, KEGG, EPD, ENZYME, REBASE, Prosite, Blocks, and Pfam. User Databases can be constructed with searching results of queries and their entries can be edited.

C-53

The Snow System, a Tool for Representation and Analysis of Networks

Christian Lemer[1], Erick Antezana[2], Fabian Couche, Simon De Keyzer, Frédéric Fays, Olivier Hubaut, Jean Richelle, Shoshana Wodak
[1]chris@amaze.ulb.ac.be, SCMBB;
[2]erick@amaze.ulb.ac.be, SCMBB

The Snow system is mainly constituted of a workbench (Snow) and a database management system (Igloo). It has been developed to represent and analyze networks in the context of the aMAZE project for the representation and the analysis of biological processes.

C-54

XTeCT, Using XML Technology to Enhance the Search for Treatment Targets, a Data Integration Project

Evangelos Pafilis[1], Ela Hunt[2], John N. Wilson
[1]vags@dcs.gla.ac.uk, University of Glasgow;
[2]ela@dcs.gla.ac.uk, University of Glasgow

XTeCT project aims to automate biological data integration. The motivation is to prepare a digest of publicly available knowledge. The underlying idea is to exploit both structural and content information of XML documents and the means to materialise this is the combined use of indexing, relational database and Data Mining technique (http://xtect.cis.strath.ac.uk/).

C-55

Using EchoBASE to Predicts Gene Function in Escherichia coli K-12

Raju Misra[1], Gavin Thomas[2], Richars Horlers, Wolfganf Reindl
[1]rvm102@york.ac.uk, University of York;
[2]ght2@york.ac.uk, University of York

EchoBASE is a database that integrates information from post-genomic experiments onto a genome database to predicts functions to uncharacterised gene products. It is available at http://www.ecoli-york.org/

C-56

A Protocol Linking Bioinformatics Resources to the Functional Analysis of a Gene

Joan C. Bartlett[1], Elaine G. Toms[2]
[1]bartlett@fis.utoronto.ca, University of Toronto;
[2]etoms@dal.ca, Dalhousie University

We present a protocol describing the application of bioinformatics analysis to the functional analysis of a gene sequence. The fourteen step protocol with three parallel pathways provides detail on each analytical step, linking multiple bioinformatics analyses together to address a biological problem, and providing the foundation for a semi-automated expert system.

C-57

INOH: A Pathway Database of Biological Events

Ken Ichiro Fukuda[1], Naotaka Ono[2], Satoko Yamamoto, Tatsuya Kushida, Yuki Yamagata, Toshihisa Takagi
[1]fukuda-cbrc@aist.go.jp, CBRC, AIST; [2]onon@hgc.jp, BIRD, JST

INOH is a pathway database of model organisms including human, mouse, rat and others. In INOH, the term pathway refers to higher order functional knowledge such as relationships among multiple bio-molecules that constitute signal transduction pathways or biological events in general. The system can be accessed at http://www.inoh.org.

C-58

Modelisation and Representation of Biological Data in Ascidians: The Ciona Virtual Embryo

Tassy[1], Lemaire[2]
[1]tassy@ibdm.univ-mrs.fr, CNRS IBDM;
[2]lemaire@ibdm.univ-mrs.fr, CNRS IBDM

The ANISEED system offers a representation of developing Ascidian embryos in terms of anatomy, proteome annotation and gene expression. Results can be accessed at http://aniseed-ibdm.univ-mrs.fr. In addition, ANISEED provides a "virtual embryo" tool that allows manipulating 3D models of Ascidian embryos and mines the data in a graphical way.

C-59

GpiIS: Towards an Integrated Information System Around Plant Genomes

Samson[1], Legeai[2], Albini, Chetouani, Karsenty, Thomas, Kimmel, Rouille, Grando, Luyten, Joets, Falque, Chataigner, Kerboul, Perche, Sajot, Bouttes, Hotelier, Sapet, Alaux, Gigonzac, Scala, James, Duclert
[1]delphine.samson@infobiogen.fr, INRA URGI;
[2]flegeai@infobiogen.fr, INRA URGI

Genoplante-info information system is a web based system (http://genoplante-info.infobiogen.fr), composed of several applications (in Java and Perl) built above a relational database (Oracle) that includes integrated schemas for sequence data (EST, mRNA), map data (genetic maps, QTL maps), transcriptome data, proteomic data and soon SNP data.

C-60

A Multi-relational Ontology for Understanding Patterns in Free Text

Michelle Maxwell[1], Julie Barnes[2], Nick Tilford, Juergen Harter
[1]michelle.maxwell@biowisdom.com, BioWisdom;
[2]BioWisdom

A multi-relational ontology for understanding patterns in free text. An ontology represents all available knowledge in a consistent and accessible manner. Using linguistic analysis on text we have built a multi-relational ontology in a clinical domain. Patterns in data using trend identification software can provide a snapshot of a domain and identify new insights that can aid drug discovery

C-61

Drug Targets and Target Genes

Kotoko Nakata[1], Yoshitomo Tanaka[2], Hiroki Momose, Hiroshi Tanaka, Tsuguchika Kaminuma
[1]nakata@nihs.go.jp, National Institute of Health Sciences; [2]ytanaka@bioinfo.tmd.ac.jp, Tokyo Med. & Dent. Univ.

Directing attention to the protein-chemical interaction and the signal pathway, we have developed a series of platform Databases for pharmacogenomics, and linked these Databases in a system of a target database. Besides including target gene information, we tried to elucidate target genes activated or repressed by the receptor in silico

C-62

Integrative Analysis of Cancer-related Data Using CAP

Pierre Dönnes[1], Annette Höglund[2], Marc Sturm, Nicole Comtesse, Christina Backes, Eckart Meese, Oliver Kohlbacher, Hans-Peter Lenhof
[1]doennes@informatik.uni-tuebingen.de, Department for Simulation of Biological Systems, Eberhard Karls University, Tübingen, Germany; Center for Bioinformatics, Saarland University, Saarbrücken, Germany; [2]hoeglund@informatik.uni-tuebingen.de, Department for Simulation of Biological Systems, Eberhard Karls University, Tübingen, Germany; Center for Bioinformatics, Saarland University, Saarbrücken, Germany

CAP is a flexible and extensible system for the integrative analysis of heterogeneous cancer-related data. Several aspects of cancer such as cancer genetics and immunology are brought together to address the fundamental causes of cancer. CAP is freely available at the following web site: http://www.bioinf.uni-sb.de/CAP/

C-63

The Stanford Microarray Database

Tina Hernandez-Boussard[1], Ihab Awad[2], Janos Demeter, Jeremy Gollub, Joan Hebert, Michael Nitzberg, Farrell Wymore, Pat Brown, Cathrine A. Ball, and Gavin Sherlock
[1]boussard@genome.stanford.edu, Stanford University; [2]ihab@genome.stanford.edu, Stanford University

DNA Microarrays are widely used for genome-scale studies. The Stanford Microarray Database (SMD; http://smd.stanford.edu /) is a relational database with an integrated collection of web-based application tools. Recent developments include incorporation of various ontologies, accepting and producing data in MAGE-ML, and accepting data generated from Agilent and Affymetrix Microarrays.

C-64

GERON Clinical: A Clinical Data Repository Framework

J. Raphael Gibbs[1], Sourav Bandyopadhyay[2], Amanda Singleton, Melissa Hanson, John Werner, Andrew Singleton, Katrina Gwinn-Hardy and Jaime Duckworth
[1]gibbsr@mail.nih.gov, Laboratory of Neurogenetics, NIA, NIH; [2]bandys@mail.nih.gov, Laboratory of Neurogenetics, NIA, NIH

In an effort to improve the availability and efficiency in the collection and mining of clinical data associated with disease, we have implemented a framework for creating clinical data repositories. GERON Clinical manages data collection, security, storage and retrieval in various aspects of clinical genetics research. System available at http://neurogenetics.nia.nih.gov.

C-65

GERMINATE: A Plant Data Management System

Jennifer Lee[1], Guy Davenport[2], David Marshall, T.H. Noel Ellis, Michael J. Ambrose, Jo Dicks, Theo J. L. van Hintum and Andrew J. Flavell
[1]jlee@scri.sari.ac.uk, University of Dundee; [2]guy.davenport@bbsrc.ac.uk, John Innes Centre

GERMINATE is a generic plant data management system implemented in PostgreSQL and designed to hold passport data and a range of additional data types including molecular markers. Its use of a top level Accessions entry ID allows association and querying between disparate sets of data associated with an accession

C-66

Building a Bioinformatics Platform for Post-Genomics Research

Michael Watson[1]
[1]michael.watson@bbsrc.ac.uk, Institute for Animal Health

We present how freely available open source projects and software can be integrated to produce a sophisticated bioinformatics platform for molecular biology.

C-67

Protein World & Orthology Benchmarking

Tim Hulsen[1], Peter Groenen[2], Wilco Fleuren
[1]T.Hulsen@cmbi.kun.nl, CMBI; [2]peter.groenen@organon.com, NV Organon

Protein World is a database in which all of the currently known and predicted proteins have been compared through the Smith-Waterman algorithm with calculation of Z-values. We use this dataset to define a strategy to test the quality of ortholog identification methods.

C-68

The Reengineered RCSB Protein Data Bank: New Means of Data Query and Distribution

Wolfgang F Bluhm[1], "The RCSB PDB Team"
[1]wbluhm@sdsc.edu, RCSB Protein Data Bank

The Protein Data Bank (http://www.pdb.org/, PDB), maintained by the Research Collaboratory for Systems Biology (RCSB), is completely reengineering its Web site and Databases We present two areas of new functionality, browsing of structures, and data query and distribution through Web Services.

C-69

Genome Knowledgebase: A Database of Biological Pathways

Geeta Joshi-Tope[1], Imre Vastrik[2], Peter D'Eustachio, Gopal Gopinathrao, Lisa Matthews, Marc Gillespie, Guanming Wu, Adrian Arva, Esther Schmidt, Bijay Jassal, Bernard de Bono, Suzanna Lewis, Ewan birney, Lincoln Stein
[1]joshi@cshl.edu, Cold Spring Harbor Laboratory ; [2]vastrik@ebi.ac.uk, European Bioinformatics Institute

Genome Knowledgebase is a curated resource of core pathways and reactions in human biology. It is a rich repository for Data Mining and for didactic purposes. GK is located at www.genomeknowledge.org

C-70

Exploiting Terrestrial Sentinel and Model Species to Integrate Tiers of Biological Response to Pollution

B.A. Hedley[1], R. Schmid, A. Anthony, J. Wasmuth, M. Blaxter
[1]ann.hedley@ed.ac.uk, University of Edinburgh

LumbrBASE: A web-based database resource for Lumbricus rubellus and Caenorhabditis elegans ESTs and associated annotations. ESTs are generated from worms exposed to five different environmental pollutants or from a control environment. We aim to use bioinformatic analysis techniques to identify the changes and trends between treatments and species.

C-71

A Personalized Biological Database System based on XML

Sung Hee Park[1], Kwang Su Jung[2], Young Jin Jung, Hyo Sung Cha, Young Uk Kim, Keun Ho Ryu
[1]shpark@dblab.chungbuk.ac.kr, Chungbuk National University; [2]ksjung@dblab.chungbuk.ac.kr, Chungbuk National University

We apply XML technology to a tool that handles biological data for biologist. We introduce a personalized genomic sequence management system based on XML. Our system consists of following components: Flat file parsers, format conversion based on XML, MyPage editor, edition for sequence and annotation, and storage manager for a personalized database.

C-72

A Search Facility for the UniProt (Universal Protein Resource) Resource

Sam Patient[1]
[1]spatient@ebi.ac.uk, EBI - European Bioinformatics Institute

UniProt (http://www.uniprot.org) is a centralized resource for protein sequences and functional information. The UniProt user base will be large and diverse and so to facilitate their access to the data the UniProt Search Facility has been created. This will provide full text, line type searching, filtering, downloading, and much more.

C-73

Algorithms and Database for the Analysis and Rapid Update of Human Immunoglobulin Sequences

Jin Lu[1], William Peter Long[2], Margaret Baker, Bernard Amegadzie
[1]jlu9@cntus.jnj.com, Centocor; [2]plong3@cntus.jnj.com, Centocor

New algorithms and methods have been developed for the rapid collection, analysis, and classification of human antibody sequences into a useful database. Multiple data displays are available for viewing data in the database.

C-74

VSDB - Vertebrate Secretome Database

Eric W. Klee[1], Liz Saftalov[2], Stephen C. Ekker, Lynda B. M. Ellis
[1]klee0025@tc.umn.edu, University of Minnesota; [2]esaftalo@purdue.edu, Purdue University

The VSDB contains secreted proteins for human, fugu, mouse, swine, and zebrafish. Protein and nucleotide sequences from disparate sources are analyzed by localization prediction programs. When available, experimental annotations for secretion are also stored. The database can return single proteins or large sequence sets.

C-75

hp-DPI: Helicobacter Pylori Database of Protein Interactomes, A Combined Experimental and Inferring Interactions

Chung-Yen Lin[1], Chia-Ling Chen[2], Chi-Shiang Cho, Li-Ming Wang, Chia-Ming Chang, Pao-Yang Chen, Chen-Zen Lo and, Chao A. Hsiung
[1]cylin@nhri.org.tw, National Health Research Institutes; [2]chialing@nhri.org.tw, National Health Research Institutes

The motivation of hp-DPI (Helicobacter pylori Database of Protein Interactomes) is to construct a human gastric pathogen H. pylori protein interaction database from a high-quality experimental dataset. This system is combining the inferred associations and experimental data with integrated annotations in web-based interface. Hp-DPI can be accessed freely at http://dpi.nhri.org.tw/hp/

C-76

Biological Pathway Research Support: The cBio Pathway Information Resource (cPath)

Gary D. Bader[1], Ethan Cerami[2], Rob Sheridan, Chris Sander
[1]bader@cbio.mskcc.org, Memorial Sloan-Kettering Cancer Center; [2]cerami@cbio.mskcc.org, Memorial Sloan-Kettering Cancer Center

cPath is open-source pathway database software that eases data integration from multiple sources. It currently supports the PSI-MI protein interaction standard and is being extended to support the BioPAX pathway exchange format. It is available for local installation and can connect with the Cytoscape network visualization and analysis tool.

C-77

InterMine - an Open-source Object Data Warehouse System

Andrew Varley[1], Richard Smith[2], Matthew Wakellng, Mark Woodbridge, Kim Rutherford, Gos Micklem
[1]ajv12@cam.ac.uk, University of Cambridge; [2]richard@flymine.org, University of Cambridge

InterMine is a generic object data warehouse system. Using one or more data models in many different formats (UML, OWL, DAG, XML Schema) as input, the software builds a complete system for efficiently integrating, storing and querying objects together with a user customisable web front end and webservice. See www.intermine.org.

C-78

IntAct - An Extensible Open Source Framework for Molecular Interactions

S. Kerrien[1], H. Hermjakob[2], L. Montecchi-Palazzi, C. Lewington, S. Mudali, S. Orchard, M. Vingron, B. Roechert, P. Roepstorff, A. Valencia, H. Margalit, J. Armstrong, A. Bairoch, G. Cesareni, D. Sherman, R. Apweiler
[1]skerrien@ebi.ac.uk, European Bioinformatics Institute; [2]hhe@ebi.ac.uk, European Bioinformatics Institute

IntAct is an extensible open source framework for molecular interactions. The project aims at defining a standard for the representation and annotation of interaction data and providing a public repository populated with experimental data from project partners and curated literature data. It contains more than 25.000 interactions and is available at http://www.ebi.ac.uk/intact

C-79

Mod-Prot, a Companion Database of Swiss-Prot for the Protein Modifications

Séverine Duvaud[1], Nathalie Farriol-Mathis[2], Elisabeth Gasteiger and Amos Bairoch
[1]Severine.Duvaud@isb-sib.ch, Swiss Institute of Bioinformatics; [2]Nathalie.Farriol-Mathis@isb-sib.ch, Swiss Institute of Bioinformatics

Mod-Prot is a new companion database of Swiss-Prot, devoted to the description of biological aspects of protein modifications. Manually annotated by Swiss-Prot curators, Mod-Prot helps to enforce PTM annotation standardization, complement Swiss-Prot user documentation and is useful to maintainers of MS-related identification tools. Mod-Prot will be available on ExPASy: http://www.expasy.org

C-80

The FlyMine Project

R Lyne[1], A Varley[2], Guillier F, Rana D, Rutherford K, Smith R, Wakeling M, Woodbridge M, Lilley K, Mizuguchi K, Russell S, Ashburner M, Micklem G
[1]rachel@flymine.org, University of Cambridge; [2]andy@flymine.org, University of Cambridge

The FlyMine project is an open-source project to build an integrated database of genomic, gene expression and proteomics data for Drosophila, Anopheles and other insects. A powerful and flexible query interface to these data enables users to perform arbitrary and complex queries across the data. Progress in the FlyMine project will be presented.

C-81

The HUPO Proteomics Standards Initiative

Henning Hermjakob[1], Chris Taylor, Weimin Zhu, Rolf Apweiler
[1]hhe@ebi.ac.uk, European Bioinformatics Institute

The HUPO Proteomics Standards Initiative (PSI) aims to define community standards for data representation in proteomics to facilitate data comparison, exchange and verification

C-82

ChEBI: A Dictionary of Chemical Compounds

P. de Matos[1], M. Ennis[2], K. Degtyarenko, M. Darsow, M. Guedj, H. Hermjakob, M. Rijnbeek, E. Kretschmann, D. Binns, R. Apweiler
[1]pmatos@ebi.ac.uk, EBI; [2]mennis@ebi.ac.uk, EBI

ChEBI is a freely available dictionary of chemical compounds, with IUPAC and NC-IUBMB endorsed terminology. ChEBI's focus has been on nomenclature but also incorporates a chemical ontology and cross-references to UniProt. ChEBI is available at http://www.ebi.ac.uk/chebi/

C-83

An Ontology for the Japanese Medaka Fish to Describe Anatomy Developmental Stages and Mutant Phenotypes

Thorsten Henrich[1], Mirana Ramialison[2], Monte Westerfield, Joachim Wittbrodt
[1]henrich@embl.de, EMBL-Heidelberg; [2]ramialis@embl.de, EMBL-Heidelberg

For the description of gene expression patterns and mutants we generated ontologies describing our model organism Medaka (Oryzias latipes) in terms of developmental stages, anatomy and phenotypes. The ontologies can be downloaded in DAGedit format at Open Biological Ontologies OBO (http://obo.sourceforge.net/)

C-84

TmaDB: A Repository for Tissue Microarray Database

Archana Sharma-Oates[1], David R. Westhead[2], Phil Quirke
[1]a.sharmaoates@leeds.ac.uk, Academic Unit of Pathology, School of Medicine, University of Leeds; [2]westhead@bmb.leeds.ac.uk, School of Biochemistry and Microbiology, University of Leeds

The use of tissue microarray (TMA) technology generates a vast quantity of data that requires a systematic approach to storage and analysis. A relational database (known as TmaDB) has been designed and implemented to archive and aid the analysis of all aspects of data relating to TMAs.

C-85

The InterPro Database

David Lonsdale[1], Maria Krestyaninova[2], Jennifer McDowall, Nicola Mulder, Rolf Apweiler, and the InterPro Consortium
[1]davidl@ebi.ac.uk, EMBL-EBI; [2]mariak@ebi.ac.uk, EMBL-EBI

InterPro (http://www.ebi.ac.uk/interpro) is an integrated protein classification resource combining PROSITE, PRINTS, Pfam, ProDom, SMART, TIGRFAMs, PIR SuperFamily and SUPERFAMILY, into a unified database. Currently InterPro contains &gt;10,400 entries representing 86% of all proteins in UniProt. InterPro entries provide functional and structural annotation for the proteins, as well as their taxonomy.

C-86

Storing and Exploring Microbial Proteomic Data

Klaus-P. Pleissner[1], Alexander Krah[2], Sven Buettner, Peter Schmelzer, Ralf Wessel, Frank Schmidt, Wolfgang Wehrl, Peter R. Jungblut
[1]pleissner@mpiib-berlin.mpg.de, Max Planck Institute for Infection Biology; [2]krah@mpiib-berlin.mpg.de, Max Planck Institute for Infection Biology

The proteome database system 2D-PAGE comprises four Databases for microbial research. Using the R language data analysis can be dynamically performed. Differentially regulated proteins detected in comparative proteomics are visualized via internet. To elucidate protein spot distribution patterns the multivariate correspondence analysis has been applied for peptide mass fingerprints.

C-87

Development of An Automatic Update System for Protein Sequences Based on XML

Kwang Su Jung[1], Sung Hee Park[2], Keun Ho Ryu
[1]ksjung@dblab.cbu.ac.kr, Database Laboratory, Chungbuk National University; [2]shpark@dblab.cbu.ac.kr, Database Laboratory, Chungbuk National University

We described a method for updating protein sequences by detecting difference between existing entries in a sequence database and new released sequence data in XML documents. Especially, the history of update and error log has been recorded in order to prove the accuracy of updating operation. The triggers are used to generate log files.

C-88

ODD-Genes - a Practical GRID Demonstrator to Enable Post-genomic Discovery

Thorsten Forster[1], Peter Ghazal[2], Arthur Trew, Malcolm Atkinson, Ratnadeep Abrol, Andrew Murdoch, Thomas Seed, Sean McGeever, Charaka Palansuriya, Muriel Mewissen
[1]Thorsten.Forster@ed.ac.uk, Scottish Centre for Genomic Technology (GTI); [2]Peter.Ghazal@ed.ac.uk, Scottish Centre for Genomic Technology (GTI)

ODD-Genes is a collaborative project resulting in a demonstrator web application. It uses GRID and related technologies (Globus Toolkit, SunDCG TOG, OGSA-DAI), enabling biological researchers efficiently analyse microarray data, which is then combined with relevant information from other, independently managed, Databases on the GRID.

C-89

An Approach for Integrating Genomic and Proteomic Expression Data

Lena Hansson[1], Peter Ghazal[2], J Douglas Armstrong, Graeme Grimes, Muriel Mewissen
[1]Lena.Hansson@ed.ac.uk, UoE; [2]p.ghazal@ed.ac.uk, Scottich Centre for Genomic Technology and Informatics

We present a method for integrating high-throughput genomic and proteomic data, based on a mixture of three different approaches: link, view integration and a warehouse. The integration uses the UniGene identifer and is tested on experimental data describing gene and protein responses to worm expulsion in mice.

C-90

DBQgen: A Flexible and Generic SQL-driven Graphical Query System

Stephan Steigele[1], Marc Roettig[2], Kay Nieselt
[1]steigele@informatik.uni-tuebingen.de, ZBIT - Center for Bioinformatics Tuebingen; [2]roettig@informatik.uni-tuebingen.de, ZBIT - Center for Bioinformatics Tuebingen

DBQgen is a generic query system realizing graphical access to existing Databases (e.g. PostgreSQL, MySQL) without any knowledge of the SQL-language. SQL-restrictions and projections can be directly formulated by using DBQgen's dialog based query system. As illustration, a database of genomewide natural-antisense-transcripts can be accessed at http://www.zbit.uni-tuebingen.de/pas/DBQgen.

C-91

Gene Expression Analysis of Neuroblastomas with Signal Transduction Networks

Rainer König[1], Frank Westermann[2], Alla Bulashevka, Benedikt Brors, Roland Eils
[1]r.koenig@dkfz.de, DKFZ; [2]f.westermann@dkfz.de, DKFZ

A signal transduction network is reconstructed from the literature-based and manually curated database Transpath (Biobase, Germany) and analysed with gene expression profiling data of pediatric neuroblastomas with good and bad prognosis. Gene-expression defines interaction strengths of neighbouring molecules to study and compare the topology of the networks.

C-92

EMBL-Bank Genome Resources

Paul Browne[1], Tamara Kulikove[2], Guy Cochrane, Nadeem Faruque, Carola Kanz, Vincent Lombard, Mary Ann Tuli, Rolf Apweiler.
[1]browne@ebi.ac.uk, EBI; [2]kulikova@ebi.ac.uk, EBI

Increases in the quantity of genome sequence data submitted to the International Sequence Database Collaboration have necessitated improvements in information acquisition and presentation by EMBL-Bank. The genomes server accesses a large set of complete genomes while the Third Party Annotation (TPA) dataset collects re-annotated genome and transcriptome sequences.

C-93

The GenomeMatrix Information Retrieval System

Hewelt[1], Dong[2], Narang, Meyer, Ben Kahla, Hennig, Himmelbauer, Yildirimman, Zehetner, Krause, Haas, Vingron, Yaspo, Lehrach
[1]hewelt@rzpd.de, RZPD; [2]siwei@rzpd.de, RZPD

GenomeMatrix (http://www.genome-matrix.org) is a multi-species database/interface system which integrates a wide range of information resources relevant to determine gene function. The modular architecture (organism specific data modules combined by gene-based orthology relationships) allows multi-species queries to compare data of orthologous genes of several organisms.

C-94

Proteome Analysis and Related Resources at the EBI

Manuela Pruess[1], Paul Kersey[2], Lawrence Bower, Ujjwal Das, Jorge Duarte, Alexander Kanapin, Youla Karavidopoulou, Virginie Mittard, Rolf Apweiler
[1]mpr@ebi.ac.uk, EMBL Outstation - The European Bioinformatics Institute (EBI); [2]pkersey@ebi.ac.uk, EMBL Outstation - The European Bioinformatics Institute (EBI)

The EBI's Proteome Analysis Database (http://www.ebi.ac.uk/proteome/) provides a tool for the in silico analysis of proteins and whole proteomes. Comprehensive statistical analyses of the predicted proteomes of fully sequenced organisms are carried out, using resources like InterPro and CluSTr, proteome sets can be downloaded and further proteome-related links are available.

C-95

Development of a Database Dedicated to the Treatment and Analysis of Long Lists of Genes

Vu Manh TP[1], Roder L[2], Piovant
[1]vumanh@ibdm.univ-mrs.fr, CNRS; [2]roder@ibdm.univ-mrs.fr, CNRS

We use Drosophila melanogaster as a model organism to identify and characterize gene networks involved in human diseases. We chose a multigenic approach, more efficient for this high throughput study. The analysis of genome-scale datasets will be facilitated with an automatical tool : an original WWW database offering standardized structural and functional annotations of Drosophila genes.

C-96

The SYSTERS Protein Family Database

Antje Krause[1], Thomas Meinel[2], Hannes Luz, Eike Staub, Martin Vingron
[1]akrause@igw.tfh-wildau.de, TFH Wildau;
[2]meinel@molgen.mpg.de, MPI for Molecular Genetics

The SYSTERS protein family database consists of the grouping of all publicly available protein sequences from the Swiss-Prot and TrEMBL Databases as well as of the predicted protein sequence sets of several completely sequenced organisms into disjoint protein family and superfamily clusters. SYSTERS is available at systers.molgen.mpg.de

C-97

PRODORIC: A Comprehensive Database of Molecular Networks in Prokaroytes

Andreas Grote[1], Richard Münch[2], Maurice Scheer, Karsten Hiller and Dieter Jahn
[1]andreas.grote@tu-bs.de, Technical University Braunschweig / Institute of microbiology;
[2]r.muench@tu-bs.de, Technical University Braunschweig / Institute of microbiology

PRODORIC (Prokaryotic Database of Gene Regulation) systematically organizes information about prokaryotic gene expression, and integrates it into regulatory networks. Clickable graphic representations of operon, gene and promoter structures including regulator binding sites, transcriptional and translational start sites, supplemented with information about regulatory proteins, are available at various detailed levels.

C-98

Mouse Genome Information Management System (GIMS) and its Application to Study Parasitic Infection

Cornelia Hedeler[1], Renu Datta[2], Norman W. Paton, Kathryn J. Else
[1]chedeler@cs.man.ac.uk, Department of Computer Science, University of Manchester;
[2]renu.datta@man.ac.uk, School of Biological Sciences, University of Manchester

Tight integration of different kinds of biological data allow experimental data analyses to be interrelated and placed in biological context Mouse GIMS has been used to analyse transcriptome data to study immune response of mouse to infection by the intestinal nematode Trichuris muris and interrelate the experimental data to pathways.

C-99

GPX: An Environment for the Storage and On-line Analysis of Microarray Data

Graeme Grimes[1], Peter Ghazal[2], Stuart Moodie, Muriel Mewissen, Sean McGeever, John Beattie, Thorsten Forster
[1]Graeme.Grimes@ed.ac.uk, The Scottish Centre for Genomic Technology and Informatics ;
[2]p.ghazal@ed.ac.uk, The Scottish Centre for Genomic Technology and Informatics

GPX provides web based access to a database of MIAME compliant experiments. GPX contains various experiment types using Affymetrix and custom microarray platforms designed at the GTI. Biologists can use GPX to find experiments related to their field of interest and analyse measurements across experiments. Using BioLink Matchmaker the biologist can identify probes for equivalent genes across different array platforms.

C-100

Sequences Meet Biology: Integration of Sequences as Database Objects in Mouse Genome Informatics

James A. Kadin[1], Richard M. Baldarelli, Benjamin L. King, Lori E. Corbani, Sharon Cousins, Jonathan S. Beal, Jill Lewis, David B. Miers, Michael B. Walker, Joel E. Richardson, Judith A. Blake, Martin Ringwald, Janan T. Eppig, Carol J. Bult, and the Mouse Genome Informatics Group
[1]jak@informatics.jax.org, The Jackson Laboratory, Bar Harbor, ME, USA 04609

The Mouse Genome Informatics (MGI) Database has radically improved support for sequence data. We now store virtually all mouse sequences and integrate those sequences with our highly curated information on mouse genes, expression, function, phenotypes, strains, and orthology. Queries integrating the above information with sequence attributes are supported.
http://www.informatics.jax.org

C-101

EDIT: A Set of Easy-to-use Bio-Data Extraction, Integration and Refreshing Tools

Tianyun Ni[1], Stanley Su[2], Lei Zhou
[1]tni@cise.ufl.edu, 1Database Systems R&D Center, College of Engineering, University of Florida;
[2]su@cise.ufl.edu, 1Database Systems R&D Center, College of Engineering, University of Florida

EDIT aimed at allowing bio-researchers with minimum computer proficiency to define the structure of a desired bio-entity, identify the data sources for its data field, and then automatically extract data over the internet and output the integrated data in predefined structure and format. An Event-Trigger-Rule Server is used to support data refreshing and notification.

C-102

MMDBBIND - Macromolecular Interactions with Atomic Level Detail

Howard Feldman[1], Kevin Snyder[2], Christopher Hogue
[1]feldman@mshri.on.ca, The Blueprint Initiative;
[2]ksnyder@blueprint.org, The Blueprint Initiative

MMDBBIND provides over 18,000 high quality fully annotated, searchable, atomic-level detail molecular interaction records to BIND (Biomolecular Interaction Network Database). Interactions can be viewed with the Cn3D structure viewing software, with interacting residues highlighted. These could be used, for example, to generate empirical protein-protein docking potentials.

C-103

The Metaclinic Database System

Thomas S. Caldwell[1], Mary E. Edgerton[2]
[1]tom.caldwell@vanderbilt.edu, Vanderbilt University Medical Center ; [2]mary.edgerton@vanderbilt.edu, Vanderbilt University Medical Center

Our Metaclinic system can rapidly create HIPAA compliant, patient-centric, normalized Databases with web-based input and reporting from metadata stored in Oracle or MSAccess Databases. Current applications include lung, breast, GI, head & neck, and gyn-onc research with extended access to external tissue microarray and proteomics systems.

C-104

Graph Data Management for Biological Networks

Barbara Eckman[1], Paul Brown[2], Julia Rice
[1]baeckman@us.ibm.com, IBM; [2]pbrown1@us.ibm.com, IBM

The management of graph datatypes representing biological networks was highlighted as an open research problem in the recent NSF/NLM/DOE-sponsored workshop on Data Management for Molecular and Cell Biology (www.lbl.gov/~olken/wdmbio). We describe a prototype from IBM Research that enables graph operations over biological networks to be executed from within SQL queries.

C-105

A Genome-wide and Non-redundant Mouse Transcription Factor Database

Harukazu Suzuki[1], Mutsumi Kanamori[2], Hideaki Konno, Naoki Osato, Jun Kawai, Yoshihide Hayashizaki
[1]harukazu@gsc.riken.jp, RIKEN Genomic Sciences Center (GSC); [2]kanamori@gsc.riken.jp, RIKEN Genomic Sciences Center (GSC)

We have developed a genome-wide and non-redundant mouse transcription factor database that has been systematically constructed on the basis of the Locus ID and the GO annotation, and that has several unique features. The database, together with its viewer, may be highly useful especially for the experimental biologists.

C-106

The Make2D-DB II Tool: Towards a Large Federated 2-DE Database Network

Khaled Mostaguir[1], Christine Hoogland[2], Ron D. Appel
[1]Khaled.Mostaguir@isb-sib.ch, Swiss Institute of Bioinformatics - Proteome Informatics Group; [2]Christine.Hoogland@isb-sib.ch, Swiss Institute of Bioinformatics - Proteome Informatics Group

Make2D-DB II is a tool to manage 2-DE Databases, with a strong emphasis on consistency and controlled vocabularies. It allows the creation and conversion of existing text-based federated 2-DE Databases into reliable relational schema. Features include automatic integration of external data sources and interconnection of similar remote 2-DE Databases. It is available at http://www.expasy.org/ch2d/make2ddb.html.

C-107

NMR Manager: Metabolomics Database Application for Interpreting Complex NMR

Robert Stones[1], Adrian Charlton[2], John Godward
[1]r.stones@csl.gov.uk, Central Science Laboratory; [2]adrian.charlton@csl.gov.uk, Central Science Laboratory

NMR Manager is a metabolite profiling database for the management and visualisation of Nuclear Magnetic Resonance (NMR) spectra. Search engines enable parameter and spectral peak searching and retrieval of spectra in a graphical interface. Allowing rapid comparisons between database spectra against spectra generated from complex mixtures e.g. blood.

C-108

Classification and Computer Representation of Enzyme Reactions: the MACiE Database

Gemma Holliday[1], Gail Bartlett[2], Peter Murray-Rust, Janet M. Thornton, and John B. O. Mitchell
[1]glh29@cam.ac.uk, Unilever Centre for Molecular Science; [2]g.bartlett@imperial.ac.uk, EBI

MACiE is a unique database of enzyme reaction mechanisms, focusing on well characterised enzymes with PDB entries and relatively well understood mechanisms. We hope to demonstrate its versatility and its use to explore the chemistry of enzyme reactions and address the classification of enzymes.

C-109

Crop Plant EST Information System CR EST

The Crop EST Database (CR EST) is an information system for storing and analysing crop EST data from multiple organisms including various EST clustering projects, BLASTX searches against the NRPEP database, functional annotations, assignments of metabolic pathways, and tools for the construction of complex search queries, interactive data exploration, and visualisation.

C-110

An Electronic Platform for Integrating and Analysing Data for Plant Research

Stephan Weise[1], Andreas Stephanik, Christian Kuenne, Thomas Thiel, Thomas Funke, Ivo Grosse
[1]weise@ipk-gastersleben.de, Instituteof Plant Genetics and Crop Plant Research (IPK)

We present a data warehouse solution for the integration of genotypic, phenotypic, taxonomic, and expression data from both IPK-internal and publicly-available data sources together with tools for data analysis and visualization.

C-111

A Grid Infrastructure Supporting the Usage of Secure Federated, Distributed Biomedical Data

Dr Richard Sinnott [1], Dr Micha Bayer[1], Derek Houghton, Dr Dave Berry, Prof Malcolm Atkinson, Magnus Ferrier, Prof David Gilbert, Dr Ela Hunt, Dr Neil Hanlon
[1]National e-Science Centre - University of Glasgow

The Cardiovascular Functional Genomics consortium (CFG) uses rat models of hypertension to study genes responsible for hypertension in the human. Experimental work requires data sharing between consortium members as well as access to many external Databases. The BRIDGES project is developing GRID technologies to enable data sharing and integration.

D-1

Comparison of De Novo Pattern Finding Algorithms for Identifying Regulatory Motifs in the Parasite Cryptosporidium Parvum

Nandita Mullapudi[1], Jessica C Kissinger[2]
[1]nandita@uga.edu, Univ of GA, Athens USA; [2]jkissing@uga.edu, Univ of GA, Athens USA

We have evaluated the performance, usefulness and suitability of various pattern-finding algorithms for their ability to identify de novo patterns in the genomes of unicellular eukaryotes. This study is aimed at identifying conserved sequence motifs upstream of genes that may be involved in gene regulation.

D-2

INCA: Synonymous Codon Usage Analysis Software

Fran Supek[1], Kristian Vlahovicek[2],
[1]fsupek@public.srce.hr, Department of Molecular Biology, Division of Biology, Faculty of Science, Zagreb University, Zagreb, Croatia; [2]kristian@icgeb.org, Department of Molecular Biology, Division of Biology, Faculty of Science, Zagreb University, Zagreb, Croatia AND Protein Structure and Bioinformatics, International Centre for Genetic Engineering and Biotechnology, Trieste, Italy

INCA (INteractive Codon usage Analysis) provides an array of features useful in analysis of synonymous codon usage in whole genomes. In addition to computing common codon usage indices, INCA offers numerous options for interactive graphical display, and clusters genes using a neural network algorithm, the self-organizing map (SOM). Available at http://www.bioinfo-hr.org

D-3

The Interaction of DNA with Clusters of Aminoacids in Proteins

Sathyapriya Rajagopal[1], Saraswathi Vishveshwara[2]
[1]spriya@mbu.iisc.ernet.in, MBU,IISc; [2]sv@mbu.iisc.ernet.in, MBU,IISc

A Graph Spectral method is used to obtain amino acid clusters from the side chains of proteins of the protein-DNA complexes. The geometry of interaction and residue conservation in these clusters are characterised for families of DNA binding proteins. Such clusters are possible recognition motifs involved the recognition of bases in the DNA.

D-4

Genome Scale Organization and Systematic Analysis of Nuclear Receptor Responsive Elements

Hidemasa Bono[1]
[1]bono@saitama-med.ac.jp, Research Center for Genomic Medicine, Saitama Medical School

In order to study gene expression regulation network by nuclear receptors, nuclear receptor target sequence elements (NREs) in genomes were computationally explored. Based on this search result, statistical features for the distribution of NREs in mammalian genomes will be presented in conjunction with the functional feature of target genes.

D-5

Importance of Small Heat Shock Proteins in Engineering Recombinant Protein Production as Revealed by Proteome Profiling

Mee-Jung Han[1], Si Jae Park[2], Tae Jung Park, Dong-Yup Lee, Sang Yup Lee
[1]mjhan@kaist.ac.kr, Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical & Biomolecular Engineering and BioProcess Engineering Research Center, Korea Advanced Institute of Science and Technology; [2]park-sj@kaist.ac.kr, Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical & Biomolecular Engineering and BioProcess Engineering Research Center, Korea Advanced Institute of Science and Technology

Escherichia coli IbpA and IbpB play important roles in protecting recombinant proteins from degradation by cytoplasmic proteases. This suggests that manipulating IbpA and/or IbpB levels may be a good strategy for the qualitative and quantitative control of recombinant protein production in E. coli.

D-6

Identification and Characterization of the Novel Virulence Determinants of Pseudomonas Aeruginosa TB by Signature Tagged Mutagenesis

Prabhakar Salunkhe[1], Lutz Wiehlmann[2], Jörg Lauber Jan Buer and Burkhard Tümmler
[1]pbsalunkhe@yahoo.com, Klin. For. OE 6711; [2]wiehlmann.lutz@mh-hannover.de, Klin. For. OE 6711

We have exploited Signature Tagged Mutagenesis (STM) together with PAO1 GeneChip expression analysis to identify and characterize genes required for the intracellular survival of a highly virulent cystic fibrosis P. aeruginosa isolate. Most of the identified genes with known functions are related to oxidative stress response, flagella and type IV pili biogenesis or have a regulatory function.

D-7

ProTeus - An Archive of Functional Signatures in Protein Termini

Iris Bahir[1], Noam Kaplan[2], Michal Linial
[1]irisba@cs.huji.ac.il, Hebrew University;
[2]kaplann@cc.huji.ac.il, Hebrew University

Most methods that detect protein signatures rely on sequence similarity without considering any positional information. Herein we present a positional-based method for revealing short signatures in both termini among ~130,000 proteins from Swissport. Many of the signatures were previously overlooked. The results are presented as an interactive website - http://www.proteus.cs.huji.ac.il

D-8

Detecting DNA-binding Proteins Using Structural Motifs and Electrostatics

Hugh P. Shanahan[1], Susan Jones[2], Carles Ferrer, Janet M. Thornton
[1]shanahan@ebi.ac.uk, European Bioinformatics Institute; [2]S.Jones@sussex.ac.uk, University of Sussex

We present a method for detecting DNA-binding proteins by a structural motif search and a positive electrostatic potential in that region. We demonstrate that this approach has a true positive rate that is comparable to more complicated structural methods of detecting DNA-binding proteins.

D-9

JVirGel: Calculation and Visualization of Two-Dimensional Protein Gels

Karsten Hiller[1], Andreas Grote[2], Richard Münch, Max Schobert, Dieter Jahn
[1]k.hiller@tu-bs.de, Technical University of Bfraunschweig; [2]andreas.grote@tu-bs.de, Technical University of Bfraunschweig

JVirGel is a Java based software that determines the theoretical isoelectric points and the calculated molecular weights of proteins and visualizes these as a virtual two-dimensional protein map. One feature of JVirGel is the comparison of the virtual protein map with an experimental 2D protein gel by overlaying both.

D-10

High-Throughput Phylogenetic Profiling for Prokaryotic Genomes

Shreedhar Natarajan[1], Mao-Feng Ger [2], Eric Jakobsson
[1]natarajn@ncsa.uiuc.edu, NCSA, Beckman Institute, Center for Biophysics and Computational Biology,University of Illinois, Urbana - Champaign IL, USA; [2]ger@uiuc.edu, NCSA, Beckman Institute, Center for Biophysics and Computational Biology,University of Illinois, Urbana- Champaign IL, USA

We report on the construction of a high-throughput phylogenetic profiling tool kit and its application to transporter proteins with an emphasis on ion channels. More specifically, we are seeking to develop automatic methods for defining homology subclasses to improve the precision of network inference from profiling, as applied to specific transport proteins

D-11

A Computational Method to Identify Human Transcription Factor Binding Sites

Davide Cora[1], Paolo Provero[2], C. Herrmann, C. Dieterich, F. Di Cunto and M. Caselle.
[1]cora@to.infn.it, University of Torino;
[2]provero@to.infn.it, Fondazione per le Biotecnologie

We developed a novel computational approach to identify human transcription factor binding sites. Starting with the catalogue of human-mouse conserved upstream sequences collected in the CORG database, we studied common overrepresented motifs shared by set og these upstreams, with respect to functional annotation and expression profiles data.

D-13

Comparative Mouse Genomics Center Consortium Genotype Database

Jesse C Wiley[1], Manjula Prattipati[2], Warren Ladiges
[1]jwiley@u.washington.edu, University of Washington;
[2]manjup@u.washington.edu, University of Washington

The Comparative Mouse Genomics Center Consortium (CMGCC) Genotype database is a portal system into the mouse models generated by the CMGCC. The database integrates information about specific mouse models, genetic approaches used to generate the animals, and their availability with public domain bioinformatics resources to maximize knowledge sharing.

D-14

The Norwegian Bioinformatics Platform

Nathalie Reuter[1], Pål Puntervoll[2], Eivind Hovig, Hans Krokan, Bjørn Alsberg, Finn Drabløs, Inge Jonassen
[1]reuter@cbu.uib.no, CBU, BCCS, University of Bergen; [2]pal@cbu.uib.no, CBU, BCCS, University of Bergen

The Norwegian bioinformatics platform, funded by the Norwegian Functional Genomics (FUGE) program, focuses on strengthening bioinformatics research and build new expertise, as well as to provide services to national Functional Genomics research projects.

D-15

Functional Genomic Analysis of a Hypertension QTL on Rat Chromosome 1

Dr. Richard Dixon[1], Dr. Steve Haines[2], Laurence Hall, Prof. Nilesh Samani
[1]rd67@le.ac.uk, Leicester University; [2]Leicester University

A region (QTL) on Rat chromosome 1 is known to affect blood pressure. Analysis of whole genome microarray gene expression data yields a candidate gene within the QTL and suggests functionally linked genes across the genome. A comparative genomics approach was used to investigate conserved regulatory sequences that control expression of candidate genes

D-16

Joint Prediction of Sigma Factors and Associated Transcription Factors in Bacillus Subtilis

Yuko Makita[1], Michiel J. L. de Hoon[2], Satoru Miyano and Kenta Nakai
[1]makita@hgc.jp, Human Genome Center, Institute of Medical Science, University of Tokyo; [2]mdehoon@ims.u-tokyo.ac.jp, Human Genome Center, Institute of Medical Science, University of Tokyo

We used binding motif information and expression data for a genome-wide prediction of sigma factor regulation in B.subtilis. We confirm the sigma factor prediction by scanning the region upstream of the predicted transcription start site for binding sites of transcription factors known to be associated with the sigma factor.

D-17

UVCLUSTER: Hierarchical Cluster Analysis of Protein Interaction Data

Vicente Arnau[1], Sergio Mars[2], Ignacio Marín
[1]Vicente.Arnau@uv.es, Departamento de Informática, Universidad de Valencia. Spain; [2]Ignacio.Marin@uv.es, Departamento de Genética, Universidad de Valencia. Spain

We have developed a new program, called UVCLUSTER, that uses iterative hierarchical clustering to analyze protein-protein interaction data. It allows the analysis of hundreds of interacting proteins, generating dendrograms that allows simple visualization of the degree of proximity among them.

D-18

Mechanism of Redox Regulation of TRPC3 Channels

Christian Rosker[1], Michael Lukas[2], Klaus Groschner
[1]cramsen@hotmail.cim, University of Graz; [2]michael.lukas@uni-graz.at , University of Graz

TRP channels have been suggested to serve as cellular redox sensors. We investigated the effects of oxidative stimuli on cellular localization and function of TRPC3. Our results indicate that TRPC3 is able to integrate redox and PLC-mediated signals, and interacts with membrane sterols as the basis of its redox sensitivity.

D-19

Locating Cis-regulatory Modules Genome Wide by Comparative Genomics

Kimmo Palin[1], Jussi Taipale[2]
[1]kimmo.palin@helsinki.fi, University of Helsinki; [2]jussi.taipale@helsinki.fi, University of Helsinki

We have developed a new method and a computational tool to discover cis-regulatory modules and functional transcription factor binding sites in genomic DNA. The method uses comparative genomics and binding site clustering to distinguish between functional and nonfunctional binding sites. The sofware is usable in genome-wide studies

D-20

Alternative Splicing in Eukaryotic Cells - Influence of SR Proteins on Splice Site Selection

Ralf H. Bortfeldt[1], Michael Hiller[2]
[1]bortfeldt@inf.uni-jena.de, FSU Jena; [2]hiller@inf.uni-jena.de, FSU Jena

Splicing is a process that is tightly regulated in higher eukaryotes. Alternative decisions underlie the control of specific splicing regulatory factors (SR proteins). We demonstrate a method to summon evidence for alternatively spliced introns, being spliced under influence of SR proteins.

D-21

Systematic Discovery of Potential Drug Targets by Automated Cell-based Proliferation Assays

Dorit Arlt[1], Wolfgang Huber[2], Christian Schmidt, Urban Liebel, Heiko Rosenfelder, Stephanie Bechtel, Alexander Mehrle, Detlev Bannasch, Ingo Schupp, Markus Seiler, Jeremy Simpson, Mamatha Sauermann, Meher Majety, Ruth Wellenreuther, Rainer Pepperkok, Holger S&uuml;ltmann, Annemarie Poustka, Stefan Wiemann
[1]d.arlt@dkfz.de, Division of Molecular Genome Analysis, German Cancer Research Center, Im Neuenheimer Feld 580, D-69120, Heidelberg, Germany; [2] w.huber@dkfz.de, Division of Molecular Genome Analysis, German Cancer Research Center, Im Neuenheimer Feld 580, D-69120, Heidelberg, Germany

We devised a fully-automated high-throughput cell-based assay to examine the effect of over-expression of 103 novel proteins on the passage of cells through the S-phase of the cell cycle. By integrating expression profiling data and domain annotations, we identified potential targets for cancer drug discovery

D-22

Associations and Hierarchies in the Human Regulatory Network

Thomas Manke[1], Christoph Dieterich[2], Martin Vingron
[1]manke@molgen.mpg.de, MPI-MG; [2]dieteric@molgen.mpg.de, MPI-MG

We report on an in-silico approach to find putative modules of human transcription factors based on binding data in conserved promoter regions. We study associations and hierarchies in the human regulatory network.

D-23

Network Motifs in Integrated Cellular Networks of Transcription Regulation and Protein-protein Interaction

Esti Yeger-Lotem[1], Shmuel Sattath[2], Nadav Kashtan, Shalev Itzkovitz, Ron Milo, Ron Y. Pinter, Uri Alon, Hanah Margalit
[1]estiy@cs.technion.ac.il, The Hebrew University; [2]ssattath@pilatmedia.com, The Hebrew University

We developed algorithms for detecting motifs in networks with two or more types of interactions, and applied them to an integrated dataset of protein-protein interactions and transcription regulation in Saccharomyces cerevisiae. We found a two-protein mixed-feedback loop motif, three-protein motifs exhibiting co-regulation and complex formation, and many four-protein motifs.

D-24

Looking for Correlated Expression in Human Fragile Sites

Angela Re[1], Isabella Sbrana[2], A. Puliti, M. Caselle
[1]angelare@to.infn.it, University of Torino; [2]i.sbrana@geog.unipi.it, University of Pisa

We evaluated the correlators between pairs of expressed fragile sites and clustered the results looking for families of coexpressed sites. We found two&nbsp; sharply separated families which we tried then to characterize at the genomic level looking for anomalies in GC content and in interspersed repeats, & nbsp; tRNA and rRNA densities

D-25

Clustering Pattern Analysis of Transcription Factors for MAPK Signaling Pathway Genes via Knowngene Promoter Database

Kyung Man You[1], Younghee Cho[2], Soo-Ik Chang
[1]kmyou@kistep.re.kr, KISTEP; [2]yhcho@kistep.re.kr, KISTEP

We investigated transcription factors that regulate gene expression in the pathway from Knowngene Promoter Database, and analyzed the clustering pattern of transcription factors for MAPK signaling pathway genes. Our results suggested that two transcription factors, SRF and MAZ, are essential for regulation of MAPK signaling pathway.

D-26

A Map of SNPs to ESTs and its Application to the Detection of Allele-specific Transcript Isoforms

Victoria P. Nembaware[1], Cathal Seoighe[2]
[1]victoria@sanbi.ac.za, SANBI, University of Western Cape; [2]cathal@sanbi.ac.za, SANBI, University of Western Cape

We describe a procedure that we have used to map SNPs from dbSNP to ESTs from dbEST, using EST and SNP locations on the human genome available from the UCSC genome project [1]. We illustrate how this can be used to test whether alternative transcript isoforms are allele-specific.

D-27

Exhaustive Assembly and Functional Classification of Available Plant ESTs Using a New Clustering Algorithm

Andrea Hansen[1], Jean Hani[2], Korbinian Schneeberger, Na Yin, Xiao Liu, Florian Buettner, Klaus Heumann
[1]andrea.hansen@biomax.com, Biomax Informatics AG; [2]jean.hani@biomax.com, Biomax Informatics AG

We have analyzed all ESTs of several plants taken from the dbEST database. We clustered them using a new clustering algorithm based on a lookup table. Subsequently we mapped the clusters onto the Arabidopsis proteome, transfered functional classifications and analysed the resulting distributions in terms of biochemical pathway coverage by using different Biomax software products

D-28

An Effective Approach for the Interpretation of Tandem Mass Spectra to Identify Peptide Modifications

Sangtae Kim[1], Heejin Park[2], Eunok Paek
[1]stkim@kma.ac.kr, Dept. of Computer Science, Korea Military Academy; [2]hjpark@hanyang.ac.kr, College of Information and Communications, Hanyang University

We present an effective method to interpret a tandem mass spectrum of a peptide with more than a few PTMs. Our method is a polynomial time algorithm and consists of three steps that make a combined use of de novo peptide sequencing, parent mass filtering, database searching, and spectral alignment.

D-29

Computational Discovery of Transcription Factors Regulating Human Genes Induced by Cigarette Smoke

Erik Borgström[1], Krzysztof Pawlowski[2], Per Broberg
[1]Erik.Borgstrom@astrazeneca.com, AstraZeneca; [2]Krzysztof.Pawlowski@astrazeneca.com, AstraZeneca

Transcription factor binding sites (TFBS) overrepresented in promoters of genes induced in human lungs by smoking were identified. Multivariate analysis was used to study relations between gene expression patterns, clinical data, and presence of TFBS. Some TF found, (AP-1, AhR), are known to be involved in response to smoking stress.

D-30

Function Specific Expression Profiles in Angiogenesis

Jakub Orzechowski Westholm[1], Adam Ameur[2], Sofie Mellberg, Emma Rennel, Lena Claesson-Welsh, Michael Cross, Jan Komorowski
[1]jakub@lcb.uu.se, The Linnaeus Centre for Bioinformatics, Uppsala University; [2]mada@lcb.uu.se, The Linnaeus Centre for Bioinformatics, Uppsala University

Parallel time profiles from endothelial cells treated with VEGF and growing on different matrices are investigated. The aim is to find genes that are controlled by VEGF on each of the two matrices, and find characteristic expression profiles for genes involved in various biological processes, using the rough set framework for rule mining.

D-31

Detecting 5′ UTR Exons in Human Using Comparative Genomics

Loredana Martignetti[1], Enrico Curiotto[2], Michele Caselle
[1]martigne@to.infn.it, Department of Theoretical Physics of the University of Torino; [2]curiotto@to.infn.it, Department of Theoretical Physics of the University of Torino

CORG is a database of human-mouse conserved upstream sequences. The size distribution of these sequences shows a clear power law decay and an unexpectedly large number of long matches. We give evidence to support the guess&nbsp; that most of these matches correspond to exons in the 5′ UTR regions.

D-32

Genome-Wide Analyses of Cooperative Transcriptional Regulations with Context-Dependent Probabilistic Models

Yuji Kawada[1], Yasubumi Sakakibara[2]
[1]yuji@dna.bio.keio.ac.jp, Department of Biosciences and Informatics, Faculty of Science and Technology, Keio University; [2]yasu@bio.keio.ac.jp , Department of Biosciences and Informatics, Faculty of Science and Technology, Keio University

We explore genome-wide analyses for cooperative binding activities in transcriptional regulations in S. cerevisiae. We build a probabilistic model which focuses on context-dependency according to the distances between transcription factors. The experimental results show that the context-dependent models significantly decrease false-positive rate and hence increase prediction accuracies for binding sites.

D-33

Conservation of Sense-antisense Transcription Between Human and Mouse

Par Engstrom[1], Boris Lenhard[2]
[1]par.engstrom@cgb.ki.se, Center for Genomics and Bioinformatics, Karolinska Institutet, Stockholm, Sweden; [2]boris.lenhard@cgb.ki.se, Center for Genomics and Bioinformatics, Karolinska Institutet, Stockholm, Sweden

We developed a computational procedure to detect pairs of overlapping and oppositely directed genes from mRNA and EST data. Even with stringent criteria for transcript sequence quality and orientation, we detect >3000 sense-antisense pairs in human and mouse transcriptomes. A significant proportion of the pairs have been conserved through evolution.

D-34

Improved Preprocessing of 1H NMR Metabolic Fingerprint Data by Adaptive Binning

Mattias Rantalainen[1], Leo Caves[2], Julie Wilson, Adrian Charlton
[1]mattias@ysbl.york.ac.uk, York Systems Biology Laboratory, Department of Chemistry, University of York; [2]lsdc1@york.ac.uk, Department of Biology, University of York

We suggest an adaptive binning methodology for preprocessing of NMR metabolic fingerprints as an alternative to the commonly used uniform binning methodology. Wavelet transformation of the NMR signal allows us to detect peak areas prior to binning and we are thereby able to avoid limitations in uniform binning

D-35

Integrating and Visualizing Proteomics Data from Web Resources in an End-User Application for 2D Gel Analysis

Nataliya Sklyar[1], Jörg Bernhardt[2], Matthias Berth
[1]sklyar@decodon.com, DECODON GmbH; [2]Joerg.Bernhardt@uni-greifswald.de, University of Greifswald

The tight integration of external annotations in image analysis software effectively assists end-users in interpretation, grouping and classification of expression profiles. This contribution presents a plugin architecture for the gel image analysis software, Delta2D, which enables attaching arbitrary annotation data from web resources to protein spots on 2D gel images.

D-36

A Functional Classification for Drosophila Melanogaster Proteins Based on the Protein-protein Interaction Network

Brun, C[1], Baudot, A[2], Martin, D. and Jacq, B.
[1]brun@lgpd.univ-mrs.fr, LGPD IBDM; [2]baudot@lgpd.univ-mrs.fr, LGPD IBDM

Very recently, two drafts of the drosophila interactome were released [1, 2]. Using PRODISTIN, a computational method allowing the functional clustering of proteins on the basis of interactions [3], we provide the first integrated view of the cellular processes of a metazoan organism as a classification tree

D-37

Integrated Analysis System for ESTs and Proteome of the Young Antlers of the Deer

Mi Ra Roh[1], Hee Jin Park[2], Ji-Hoon Jun, YongWook Kim, Seung-Moak Kim, Byoung Chul Park, Moonkyu Kang, Hyunsu Bae, Youngho Moon, Jae Jong Kim and Yong-ho In
[1]miraroh@idrtech.com, IDRTech, Inc. Research Center.; [2]parkbc@kribb.re.kr, Systemic Proteomics Research Center, KRIBB.

To understand functional relationships between sequence and expression information, we combine hypothetical PMF from ESTs and real PMF from Proteome in one system. This integrated gene function analysis system brings up the evidence of gene function from EST annotation and proteome analysis results at the same time.

D-38

Transcription Factors Involved in Mouse Embryonic Stem Cell Differentiation

hyojeong ban[1], sunghun back[2], yongseong lee, youngseek lee
[1]ban@ihanyang.ac.kr, hanyang university;
[2]bsh96@ihanyagn.ac.kr, hanyang university

The cDNA microarray data were analyzed for finding genes related to differentiation from mouse embryonic stem cell (MESC) to nervous system. In results, we predicted transcription factors and genes that involved in cell differentiation of MESC

D-39

Sequence Conservation and Co-evolution in Multiprotein Complexes

Nicolas Goffard[1], Pascal Durrens[2], Antoine de Daruvar
[1]nicolas.goffard@pmtg.u-bordeaux2.fr, Centre de Bioinformatique Bordeaux (CBiB);
[2]pascal.durrens@pmtg.u-bordeaux2.fr, Centre de Bioinformatique Bordeaux (CBiB)

By extrapolating the experimentally-defined multiprotein complexes in S. cerevisiae to yeast genomes newly sequenced and annotated by the GDR Genolevures, we demonstrate that, even inside a phylogenetic group, the measure of selective pressure for the structure conservation of interacting proteins and for the co-evolution of these proteins is possible.

D-40

A Study of Functional Protein Interactions in Signaling and Degradation

Cheryl Wolting[1], Chris Hogue[2], Jane McGlade
[1]cheryl.wolting@utoronto.ca, Hospital for Sick Children;
[2]chogue@blueprint.org, Blueprint, Samuel Lunenfeld Research Institute

The Signaling and protein Degradation Network of Toronto (SIDNET) proposes to perform high-throughput functional interaction and enzymatic activity assays on selected gene families of the human proteome. In this project, we propose to develop bioinformatics tools to address the problem of practical and efficient data storage and analysis for SIDNET.

D-41

Computational Gene Prioritization

Bert Coessens[1], Stein Aerts[2], Patrick Glenisson, Diether Lambrechts, Bart De Moor
[1]bert.coessens@esat.kuleuven.ac.be, K.U.Leuven;
[2]stein.aerts@esat.kuleuven.ac.be, K.U.Leuven

The results of experimental or computational analyses in the post-genomic era often consist of long lists of candidate genes. A systematic approach is presented for in silico gene prioritization that combines multiple genomic information sources. The combined re-ranking is done with order statistics.

D-42

The Investigation of Tumour Gene Expression Through the Analysis of Expressed Sequence Tag Libraries

James Prendergast[1], Colin Semple[2]
[1]jamesp@hgu.mrc.ac.uk, University of Edinburgh;
[2]colins@hgu.mrc.ac.uk, MRC Human Genetics Unit

EST data is a valuable resource for the investigation of tumour gene expression. In order to further characterise the genetic basis of various cancers we comprehensively examined EST libraries from various normal and cancerous tissues. Results obtained include the identification of gene coexpression as well as differentially expressed genes and genomic regions.

D-43

Computational Analysis Of Saccharomyces Cerevisiae Identifies A New Modality Of Post-Transcriptional Control Of Gene Expression By Programmed Ribosomal Frameshifting

Jonathan L. Jacobs[1], Jonathan D. Dinman[2], Ewan P. Plant
[1]jacobsjo@umd.edu, The University of Maryland;
[2]dinman@umd.edu, The University of Maryland

Programmed -1 ribosomal frameshift (-1 PRF) signals are predicted to be widespread throughout the yeast genome. Testing at the bench confirmed these signals promote efficient levels of -1 PRF and act as mRNA destabilizing elements, suggesting a new modality for post-transcriptional regulation of gene expression.

D-44

Path Finding in Metabolic Networks: Looking for Functional Relationships Between Enzymes

Didier Croes[1], Fabian Couche[2], Jacques van Helden, Shoshana Wodak
[1]didier@scmbb.ulb.ac.be, SCMBB;
[2]fcouche@scmbb.ulb.ac.be, aMaze

Our Metabolic Path Finding Tool finds the shortest path in a metabolic graph. We validate our methodology by rediscovering known annotated pathways. We have defined a new metric to mesure the metabolic distance between enzymes. We calculated metabolic distance between pairs of interacting, fused-genes and operons enzymes. Web site: http://www.scmbb.ulb.ac.be/pathfinding/

D-45

PartiGene - From ESTs to Partial Genomes and Beyond

Ralf Schmid[1], Alasdair Anthony[2], John Parkinson, James Wasmuth, Ann Hedley, Mark Blaxter
[1]R.Schmid@ed.ac.uk, School of Biological Sciences, University of Edinburgh, UK; [2]al.anthony@ed.ac.uk, School of Biological Sciences, University of Edinburgh, UK

PartiGene is an integrated sequence analysis suite which uses publicly available software tools to process and organise EST-datasets. PartiGene takes raw trace chromatograms and performs base calling, vector and low quality sequence removal, clustering into putative genes, consensus sequence prediction, peptide translation and functional annotation of the predicted peptides (http://www.nematodes.org/PartiGene).

D-46

Integrated Pig Genome Analysis

Frank Panitz[1], Henrik Hornshøj, Henrik Stengaard, Lone B. Madsen, Per Horn, Jakob Hedegaard, Christian Bendixen
[1]Frank.Panitz@agrsci.dk, Danish Institute of Agricultural Sciences

Porcine genome and EST sequencing analysis provides unique opportunities to study genetic variation in pigs. We combine sequence analysis, annotation, SNP-detection, microarray production and gene-expression analysis in order to establish a Functional Genomics platform for analysing genetic variation and identifying candidate genes that focus on biomedical and agricultural phenotypes.

D-47

Composite Regulatory Modules in ChIP Derived Sequences

Alexander Kel[1], Dmitry Tchekmenev[2], Olga Kel-Margoulis and Edgar Wingender
[1]ake@biobase.de, BIOBASE GmbH; [2]dtc@biobase.de, BIOBASE GmbH

We analyzed the sequences obtained in ChIP on chip experiments using position weight matrices for transcription factors stored in TRANSFAC database, using MATCH tool. We found that p53 sites are often accompanied by sites for E2F and AhR transcription factors, and sites for cMyc by sites for CREB and USF.

D-48

Prediction of mammalian miRNA genes by comparative genome analysis

Ying Sheng[1], Boris Lenhard[2], Jose Sepulveda-Sanches, Björn M. Ursing
[1]Ying.Sheng@cgb.ki.se, Center for Genomics and Bioinformatics, Karolinska Institutet, Sweden; [2]Boris.Lenhard@cgb.ki.se, Center for Genomics and Bioinformatics, Karolinska Institutet, Sweden

MicroRNAs (miRNA) are endogenous small noncoding RNAs that play important regulatory roles in animals and plants. We combined the rules for the observed conservation patterns of known miRNAs with the requirements for a conserved stem loop structure to construct a pre-miRNA prediction pipeline.

D-49

The Yeast Gene Regulation Network: A Comparative Analysis of Genome Wide Data Sets

Thomas Schlitt[1], Kimmo Palin[2], Esko Ukkonen, Alvis Brazma
[1]schlitt@ebi.ac.uk, British Antarctic Survey;
[2]kpalin@cs.helsinki.fi, 2Department of Computer Science, Helsinki

Using a graph-based approach, we compared large-scale experiments (ChIP-on-chip, microarray, predicted binding sites) with respect to intersection, neighbourhood similarity and graph topology. Despite small intersections between the data sets, we identified a network motif, which includes direct and indirect interactions, and predict the function for five previously uncharacterised genes.

D-50

ELM: Knowledge-based Prediction of Short Protein Interaction Motifs and Post-translational Modification Sites

The ELM Consortium[1], presented by: Paal Puntervoll[2]
[1]http://elm.eu.org; [2]pal@cbu.uib.no, CBU, BCCS, Univeristy of Bergen

ELM is a web-resource for predicting functional sites in proteins. Examples of functional sites are short protein interaction motifs and post-translational modification sites. Predictions are based on patterns (regular expressions) describing functional site motifs. The quality of the predictions is improved by applying context-based filters, taking advantage on biological knowledge URL: http://elm.eu.org

D-51

The SOFG Anatomy Entry List (SAEL) for Functional Genomics Applications and as an Entry Point to Existing Anatomy Ontologies

Stuart Aitken[1], Richard Baldock[2], Jonathan Bard, Albert Burger, Duncan Davidson, Terry Hayamizu, Helen Parkinson, Alan Rector, Martin Ringwald, Jeremy Rogers, Cornelius Rosse, Christian J. Stoeckert
[1]stuart@inf.ed.ac.uk, University of Edinburgh; [2]Richard.Baldock@hgu.mrc.ac.uk, MRC-HGU

The SOFG Anatomy Entry List (SAEL) is a freely-available selection of cross-species anatomical terms mapped to existing anatomy ontologies. It thus provides an entry point to these ontologies and a resource for simple annotation and automated information retrieval. The SAEL is a resource for biologists, curators, and software developers

D-52

Comparison of Clustering Methods for the Identification of Co-regulated Gene Groups from Gene Expression Time Series

Martin Hoffmann[1], Stefan Wölfl[2], Stefan Knoth, Sotirios Ziagos, Vladimir Monossov
[1]martin.hoffmann@hki-jena.de, Hans Knöll Institute of Natural Products Research Jena ; [2]stefan@imb-jena.de, 2Friedrich Schiller University Jena

Different approaches for analyzing the responses of human blood cells to biochemical stimuli are compared. The significance of the results was evaluated by established biological knowledge, i.e. the resulting gene groups were checked for functional dependencies by Databases (Kegg, GenMAPP, TransPath, Gene Ontology) and expert knowledge

D-53

Shape Description of Protein Clefts

Rafael Najmanovich[1], Richard Morris[2], Roman Laskowski and Janet Thornton
[1]rafael.najmanovich@ebi.ac.uk, EBI; [2]rjmorris@ebi.ac.uk, EBI

In the present work we describe the utilization of hybrid ellipsoids to model the shape of protein clefts. The small number of parameters describing the hybrid ellipsoid model can be used to assess binding site shape similarities. The method has applications in the prediction of protein function from structure.

D-54

Comparative Experimental Bioinformatics of Marine Cyanobacteria

Ilka M. Axmann[1], Szymon M. Kielbasa[2], Wolfgang R. Hess
[1]ilka.axmann@gmx.de, Humboldt-University, Institute for Biology, Chausseestr. 117, D-10115 Berlin, Germany; [2]s.kielbasa@itb.biologie.hu-berlin.de, Humboldt-University, Institute for Theoretical Biology, Invalidenstrasse 43, D-10115 Berlin, Germany

Using a comparative computational and experimental approach orthologous upstream nucleotide sequences of marine cyanobacterial genes were analysed (phylogenetic footprinting). The core promoter structure and possible TFBS unknown so far were predicted giving new insights into transcriptional regulation for strains adapted to particular ecological niches within the marine ecosystem.

D-55

FUNGUS: A J2EE Application that Manages High-throughput Functional Cell-based Screening

David A Block[1], Christian Zmasek[2], Phillip McClurg, Hilmar Lapp
[1]dblock@gnf.org, GNF; [2]czmasek@gnf.org, GNF

FUNGUS (Functional Genomics Using Screens) is a software application developed to store, analyze, manage, and present data from a high-throughput functional cell-based screening program. Several hard-learned lessons about database design and object-relational mapping in Java will be presented. An enterprise-level software framework was used to provide scalability and extensibility.

D-56

Ionomics of Yeast

Murlidharan[1], Mathias[2], Suzanna Clark, Jeffrey Harper, Mary Lou Guerinot, David Eide, Michael Gribskov
[1]nair@sdsc.edu, Nair; [2]mgehl@scripps.edu, Gehl

Inductively coupled atomic emission spectroscopy (ICP-AES) has been used as a Functional Genomics tool. Profiles of 12 elements have been studied for 5000 knockout lines in yeast. The method is a useful approach to assign function to genes with as yet unknown function.

D-57

Inferring Differential-Equation Models of Genetic Network Dynamics

Theodore J. Perkins[1], Michael T. Hallett[2], Leon Glass
[1]perkins@mcb.mcgill.ca, McGill University;
[2]hallett@mcb.mcgill.ca, McGill University

We describe an new strategy for inferring differential-equation models of genetic network dynamics from spatiotemporal expression data. The approach can capture complex, combinatorial regulatory relationships between genes, and allows the efficient evaluation of all possible network topologies. We apply the approach to data from the gap-gene system in Drosophila Melanogaster

D-58

Digital Expression in Porcine Tissues

Jan Gorodkin[1], Susanna Cirera[2], Milena Sawera,Ami Klein, Annett Frankel, Claus B. Jørgensen,Merete Fredholm
[1]gorodkin@bioinf.kvl.dk, The Royal Veterinary and Agricultural University; [2]scs@kvl.dk, The Royal Veterinary and Agricultural University

Expression profiles for 100 porcine cDNA libraries are compared and we find that several Gene Ontology categories correlate strongly, for example "transcription regulator activity" with "translation regulator activity". We also find that the brain and testes libraries have most different genes.

D-59

Modular Regulation of Mammalian Terminal Differentiation

Sven Nelander[1], Erik Larsson[2], Erik Kristiansson, Olle Nerman, Petter Mostad, Per Lindahl
[1]sven.nelander@medkem.gu.se, Göteborg University;
[2]erik.larsson@medkem.gu.se, Göteborg University

For several reasons, functionally coupled co-regulated gene units (gene batteries) may be the defining units of terminal cell differentiation. We performed a global unsupervised screen for mammalian gene batteries, and identified their cis regulators. The results strongly underline the importance of gene battery-type regulation in terminal differentiation in mammals.

D-60

Upstream Sequence Motif Analysis of Human Ig and TCR Genes

Narayanan Perumal[1]
[1]nperumal@iupui.edu, School of Informatics, Indiana University - Purdue University at Indianapolis

Potential transcriptional control motifs in the upstream regions of human immunoglobulin and T cell receptor genes have been identified, employing an informatic approach to study sterile transcription of their germline genes. These motifs are probably biologically relevant in transcriptional control since some of them have been implicated in experimental regimens.

D-61

A Combined Computational-experimental Approach Predicts Human MicroRNA Targets

Andrei Kouranov[1], Marianthi Kiriakidou, Peter T. Nelson, Petko Fitziev, Costas Bouyioukos, Malik Yousef, Zissimos Mourelatos, Artemis Hatzigeorgiou
[1]akourano@pcbi.upenn.edu, University of Pennsylvania Center for Bioinformatics

We use a combined bioinformatics and experimental approach to identify important rules governing miRNA-MRE (miRNA-recognition elements) recognition that allow prediction of human miRNA targets. We describe a computational program, "DIANA-microT", that identifies mRNA targets for animal miRNAs and predicts mRNA targets, bearing single MREs, for human and mouse miRNAs.

D-62

A Bioinformatics Approach to Find Breast Cancer Susceptibility Genes

Miguel Pujana[1], Jing-Dong J. Han[2], Muneesh Tewari,Jin-Sook Ahn, Nono Ayiviguedehoussou and Marc Vidal
[1]*miguel_pujana@dfci.harvard.edu, Center for Cancer Systems Biology, Dana-Farber Cancer Institute and Department of Genetics, Harvard Medical School;* [2]*jackie.han@research.dfci.harvard.edu, Center for Cancer Systems Biology, Dana-Farber Cancer Institute and Department of Genetics, Harvard Medical School*

We used a bioinformatics approach to search for genes that are associated with known breast cancer susceptibility genes by Data Mining a spectrum of large-scale functional genomic data from different organisms.

D-63

Using Structural Knowledge for Ab Initio Prediction of Transcription Factor Targets

Tommy Kaplan[1], Nir Friedman[2], Hanah Margalit
[1]*tommy@cs.huji.ac.il, Hebrew University;* [2]*nir@cs.huji.ac.il, Hebrew University*

Our approach combines sequence and structure to learn context-specific DNA-recognition preferences and to predict binding sites of novel transcription factors. We apply it to the Cys2His2 zinc finger family, and predict targets in Drosophila in a genome-wide manner. We infer the factors' function and activity using their targets' annotations and expression.

D-64

Computer Analyses of Aerobic-anaerobic Regulation in Gamma-Proteobacteria

A.V. Gerasimova[1], D.A. Ravcheev[2], A.B. Rakhmaninova, M.S. Gelfand
[1]*a_gerasimova@yahoo.com, State Scientific Center "GosNII Genetica", Moscow, Russia;* [2]*ravcheyev@iitp.ru, Moscow State University, Department of Bioengineering and Bioinformatics, Moscow, Russia*

A large number of gamma-proteobacteria genomes was studied by methods of comparative genomics. Putative ArcA, FNR and NarP regulons were described in detail for a number of microorganisms. After that different regulons for each organism were compared. We also confirmed known FNR-binding signal, improved the recognition profile for the ArcA-signal and created a new profile for the NarP-signal.

D-65

SNPlexTM System: A Multiplex High-throughput Genotyping Platform Based on the Oligonucleotide Ligation Assay

Xiaoqing You[1], Nicolas Peyret[2], Ryan T. Koehler, Joseph P. Day, Jason Evans, Lily Xu, Heinz Hemken, Annie Titus, Charles R. Scafe, Eugene Spier, Joanna Curlee, and Francisco M. De La Vega
[1]*youxn@appliedbiosystems.com, Applied Biosystems;* [2]*youxn@appliedbiosystems.com, Applied Biosystems*

The SNPlexTM System is a multiplex high-throughput genotyping technology based on an oligonucleotide ligation assay. At a 48-plex level, the technology can resolve >500,000 genotypes/day with a high level of accuracy. It has concordance rates of 99.2% with TaqManâ probe-based 5'-nuclease genotyping assay and of 98.7% with dideoxy sequencing.

D-66

Construction of Rice cDNA Microarray and its Application

Lee Jung-Sook[1], Kim Yong-Hwan[2], Yoon Ung-Han, Lee Gang-Seob, Hyun Do Yoon, Hahn Jang-Ho and Kim Ho-Il
[1]*jungslee@rda.go.kr, National Institute of Agricultural Biotechnology;* [2]*yghnkim@rda.go.kr, National Institute of Agricultural Biotechnology*

Among ~10,000 EST sequences, about 4,600 clones were successfully amplified and printed in glass slides, making it useful to study the abiotic stress function in rice. 317 genes were responding to cold or salt stress over 3 fold changes in rice.

D-67

Genome-wide Survey of Cancer Related DNA Hypermethylation

Ilana Keshet[1], Shlomit Farkash[2], Howard Cedar, Richard A. Young, Alain Niveleau
[1]*kesheti@md.huji.ac.il, Hebrew University Medical School;* [2]*shlomitf@md.huji.ac.il, Hebrew University Medical School*

We have developed a method for the detection of the genome-wide methylation status. This method is based on methylated DNA immunoprecipitation and on microarray technology. Using this method for studying cancer related DNA hypermethylation revealed that it is a controlled mechanism, derived probably by chromatin structure and DNA binding proteins.

D-68

Cumulative Local Cross-Correlation - an Algorithm for the Decomposition of Sequence

PatternsSimon Kogan[1]
[1]*skogan@research.haifa.ac.il, Institute of Evolution, University of Haifa*

Cumulative Local Cross-Correlation is a novel algorithm for reconstruction of patterns (codes) of DNA sequence repeats. It is capable on working with dispersed repeats as well. The algorithm is based on a cross-correlation procedure applied locally in a cumulative fashion. It could be naturally generalized for multidimensional cases and easily adopted in other signal processing applications

D-69

Profiling the Proteome of Murine Mveloma NSO Cells by Two-dimensional Gel Electrophoresis

Haiza Ahmad[1], Jian Zhang[2], Mark Smales
[1]*na31@kent.ac.uk, University of Kent at Canterbury;* [2]*j.zhang@kent.ac.uk, University of Kent at Canterbury*

We have undertaken a study on profiling the proteome in cell lines with different antibody production rates, through the use of 2D gel electrophoresis Using a weighted hierarchical clustering, we showed that the changes in antibody production in these cell lines correlate with the expressions of these proteomes

D-70

Supervised Partitioning of the Protein Space: An Information - Theoretic Approach

Menachem Fromer[1], Moriah Friedlich[2], Noam Kaplan, Nathan Linial, Michal Linial
[1]*fromer@cs.huji.ac.il, The Hebrew University of Jerusalem;* [2]*fridlim@cs.huji.ac.il, The Hebrew University of Jerusalem*

The BF (Best Front) technique uses an information-theoretic defined distance to automatically determine the optimal level in a hierarchical tree of proteins, relative to a given annotation set. This novel method was tested on the ProtoNet hierarchy using Pfam annotations. The quality of this BF is assessed at www.protonet.cs.huji.ac.il/bestFront/

D-71

Application of Massively Parallel Signature Sequencing to Gene Expression Analysis of Theileria Parva Infected Lymphocytes

Etienne P. de Villiers[1], Trushar Shah[2], Roger Pelle, Vish Nene, Malcolm J. Gardner, Evans Taracha and Richard Bishop
[1]*e.villiers@cgiar.org, International Livestock Research Institute;* [2]*t.shah@cgiar.org, International Livestock Research Institute*

In an effort to discover potential vaccine candidate genes of Theileria parva, an intracellular protozoan parasite responsible for east coast fever, a serious tick-borne cattle disease that transforms bovine lymphocytes to a leukemia-like state, we applied Massively Parallel Signature Sequencing (MPSS) to define the complete transcriptome of the T. parva schizont.

D-72

Conserved Sequence Motifs in Regulatory Regions of Prokaryotic Genes: Riboswitches and Beyond

Cei Abreu-Goodger[1], Nancy Ontiveros, Ricardo Ciria, and Enrique Merino
[1]*cei@ibt.unam.mx, Institute of Biotechnology, UNAM*

We search for conserved regulatory motifs without using any knowledge of regulon or metabolic pathway structure , using only orthologous relationships. We can locate every previously reported riboswitch with our method, as well as many new highly conserved regulatory elements.

D-73

Comparative Genomics of Amino Acids Biosynthesis in Bacteria: a Variety of Regulatory Systems

Dmitry Rodionov[1]
[1]*rodionov@iitp.ru, Institute for Problems of Information Transmission*

The metabolic pathways of amino acids biosynthesis in bacteria were investigated by a set of comparative genomics techniques including positional and genome context analyses complemented by identification of amino acid-specific regulatory elements. It is resulted in complete reconstruction of these metabolic pathways and identification of new members of the regulons.

D-74

Chromosomal Organization is Shaped by the Transcription Regulatory Network

Ruth Hershberg[1], Esti Yeger-Lotem[2], Hanah Margalit
[1]rutih@md.huji.ac.il, The Hebrew University;
[2]estiy@cs.technion.ac.il, The Hebrew University and Technion

We study the relationship between transcription regulation and chromosomal organization using methods for network analysis. Integrated networks representing both types of data are created and searched for network motifs that recur more often than expected at random. Motifs found demonstrate that transcription regulation shaped chromosomal organization in pro- and eukaryotes.

D-75

Optimising in Silico Methods for the Detection of Common Cis-regulatory Modules in Co-regulated Genes

Peter Van Loo[1], Stein Aerts[2], Jan Cools, Bart De Moor, Peter Marynen
[1]petervl@ace.ulyssis.org, Human Genome Laboratory, Department of Human Genetics, Flanders Interuniversity Institute for Biotechnology, University of Leuven, Herestraat 49, 3000 Leuven, Belgium;
[2]Stein.Aerts@esat.kuleuven.ac.be, Department of Electrical Engineering (ESAT-SCD), University of Leuven, Kasteelpark Arenberg 10, 3001 Heverlee, Leuven, Belgium

Previous in silico methods for modelling cis-regulatory modules (CRMs) in co-regulated genes were extended and optimised. A CRM model was created for upregulation during HL-60 differentiation and used to scan the human genome. Real time quantitative PCR revealed high probabilities of expression (>80%) and higher than random probabilities (42%) of upregulation.

D-76

Restricted Site Tags (RST) Libraries: Discriminating Analysis of Complex Mixed Microbial Systems

Alexey Kutsenko[1], Veronika Zabarovska[2], Lev Petrenko, Tore Midtvedt, Ingemar Ernberg, Eugene R. Zabarovsky
[1]Alexey.Kutsenko@mtc.ki.se, Microbiology and Tumor Biology Center, Karolinska Institute;
[2]Veronika.Zabarovska@mtc.ki.se, Microbiology and Tumor Biology Center, Karolinska Institute

We develop a new efficient technique for large scale scanning of genomes in complex multiorganisms mixture with discriminating different species and even strains. For a particular organism we create the restricted site tags set (RST) that represents a unique genomic fingerprint of this specie and that is easy to generate.

D-77

Construction of a Dense Single Nucleotide Polymorphism Map for in Silico Mapping in Mouse

Serge Batalov[1], Tim Wiltshire[2], Mathew T. Pletcher, Philip McClurg, S. Whitney Barnes, Erica Lagler, Deborah Nusskern, Molly Bogue, Richard J. Mural, Beverly Paigen
[1]batalov@gnf.org, GNF; [2]timw@gnf.org, GNF

For in silico mapping of phenotypic trait association with ancestrally inherited polymorphisms, we have selected and genotyped in 48 mouse strains 10,917 evenly spaced SNP loci. Use of this SNP marker set with an appropriate statistical model could successfully map known single gene traits and a QTL gene.

D-78

Codon Bias Is a Factor in Regulating Expression via Translation Efficiency in the Human Genome

Yizhar Lavner[1], Daniel Kotlar[2]
[1]yizhar_l@kyiftah.org.il, Tel-Hai Academic College;
[2]danny-k@actcom.co.il, Tel-Hai Academic College

We present evidence, based on tRNA gene copy numbers, suggesting that selection acts on codon bias, in the human genome, not only to increase elongation rate by favoring optimal codons in highly expressed genes, but also to reduce elongation rate by favoring non-optimal codons, in lowly expressed genes.

D-79

LacplantCyc: a Lactobacillus Plantarum Pathway / Genome Database as a Reference for Lactic Acid Bacteria

Frank H.J. van Enckevort[1,2,1], Christof Francke[1,3,2], Bas Teusink[2,3] and Roland J. Siezen[1,2,3]
[1]Frank.van.Enckevort@nizo.nl, [1]Centre for Molecular and Biomolecular Informatics, University of Nijmegen, The Netherlands; [2]NIZO food research, Ede, The Netherlands.; [2]C.Francke@cmbi.kun.nl, [3]Wageningen Centre for Food Sciences, Wageningen; The Netherlands

Lactobacillus plantarum is a versatile lactic acid bacterium that is encountered in various niches. LacplantCyc is a pathway/genome database derived from the annotated genome sequence of L. plantarum WCFS1 (PNAS 2003;100:1990), using the PathoLogic software from PathwayTools (P.Karp). LacplantCyc is envisioned to become the reference Gram-positive PGDB for LAB (http://www.lacplantcyc.nl/)

D-80

Analysis of Alternatively Spliced Coding Frames

Kouichi Kimura[1], Tetsuo Nishikawa[2], Ai Wakamatsu, Jun'ichi Yamamoto, Jun'ichi Uechi, Tomohiro Yasuda, Ken'ichi Nagai, Takao Isogai
[1]kokimura@crl.hitachi.co.jp, Central Research Laboratory, Hitachi Ltd.; [2]nisikawa@crl.hitachi.co.jp, Central Research Laboratory, Hitachi Ltd.

Splicing variants with adjacent coding exons with mismatched coding frames are studied with our tools. Intris visualizes the genome mapping results of splice variants suppressing the common intron regions; TRins shows the coding potential, similarity to known proteins, and discrepancy between the transcript and the genome in an integrated way.

D-81

PhyloMatrix: a Tool for Phylogenetic Profiling within the SYSTERS Protein Family Web Server

Thomas Meinel (1)[1], Antje Krause (1,2)[2], Eike Staub (1), Martin Vingron (1)
[1]Thomas.Meinel@molgen.mpg.de, 1. Max Planck Institute for Molecular Genetics, Dept. Computational Molecular Biology; [2]Antje.Krause@molgen.mpg.de, 2. TFH Wildau, Dept. Biosystemtechnik/Bioinformatik

We present PhyloMatrix, a web-based front end to phylogenetic profiles that are based on the protein family definitions of the SYSTERS database. Information about 6,461 profiles based on 106 complete genomes can be queried via profile pattern searches or various protein annotations at http://systers.molgen.mpg.de

D-82

Computational Image Analysis for Automatic Measurement of Rice Seedling Growth for the Visible Phenotypic Functional Analysis

Takanari Tanabata[1], Tomoko Shinomura[2], Makoto Takano, Noritoshi Inagaki
[1]ttanaba@rd.hitachi.co.jp, Hitachi Central Research Laboratory; [2]shino@harl.hitachi.co.jp, Hitachi Central Research Laboratory

We constructed large-scale image database for measurement of rice seedling growth. We report the automatic growth measuring software and its application for the expansion profile of a growing rice shoot. Using the software, we measured shoot apex position data from images at high temporal resolution and defined various growth parameters.

D-83

Hierarchical Clustering of Medaka Gene Expression Patterns Based on a Controlled Vocabulary of Anatomical Terms

Ramialison[1], Henrich[2], Wittbrodt
[1]ramialis@embl.de, EMBL; [2]henrich@embl.de, EMBL

"Synexpression group" in eukaryotes characterizes a cluster of genes which share a common complex localization and a same biological processes. Here we applied hierarchical clustering on 880 Medaka gene expression pattern descriptions and could identify several groups of co-expressed genes involved in a similar biological context.

D-84

Comparative Promoter and Gene Analyses of Signaling Protein

Ulrike Gausmann[1], Niels Jahn, Kathrin Reichwald, Matthias Platzer
[1]ugau@imb-jena.de, Department of Genome Analysis, Institute of Molecular Biotechnology, Jena, Germany

Analyzing multifunctional signaling proteins we focus on proteins which physically interact and are potentially co-regulated. Transcript analyses were performed to identify putative promote regions and to reveal alternative transcripts. The 5' upstream gene regions were compared by several programs to identify regulatory elements common for a defined group.

D-85

Detection of Conserved Cross-species Regulatory Elements in Higher Eukaryotes Using Motif Detection Algorithm

Amonida Zadissa[1], John McEwan[2], Chris Brown
[1]amonida@sanger.otago.ac.nz, Otago University;
[2]john.mcewan@agresearch.co.nz, AgResearch

Using a combination of the MEME motif prediction technique and developed algorithms, we aim to identify conserved regulator elements in orthologous promoters of co-regulated cardiac bovine genes in human, mouse and rat genomes. The methodology has identified known and novel elements involved in muscular and cardiac specific gene regulation.

D-86

PupaSNP Finder: a Web Tool for Finding SNPs with Putative Effect at Transcriptional Level

Lucía Conde[1], Juan M. Vaquerizas[2], Javier Santoyo, Fátima Al-Shahrour, Sergio Ruiz-Llorente, Mercedes Robledo and Joaquín Dopazo
[1]lconde@cnio.es, Spanish National Cancer Centre (CNIO); [2]jvaquerizas@cnio.es, Spanish National Cancer Centre (CNIO)

PupaSNP is a web-based tool for high-throughput search of SNPs located at intron/exon boundaries, exonic splicing enhancers or predicted transcription factor binding sites that can have a potential phenotypic effect at transcriptional level. SNPs producing amino-acid changes are also retrieved. PupaSNP is available at http://pupasnp.bioinfo.cnio.es.

D-87

A Multi-facetted Approach to Predict Functional Coupling of Genes in Eukaryotes

Andrey V. Alexeyenko[1], Erik L.L. Sonnhammer[2]
[1]Andrey.Alexeyenko@cgb.ki.se, Karolinska Institutet;
[2]Erik.Sonnhammer@cgb.ki.se, Karolinska Institutet

The phenomenon of chromosomal clustering of functionally coupled genes has been used to predict novel relations between genes. The descriptive and comparative data on genomic locations, gene expression, protein-protein interactions are applied to discover functional coupling in a number of eukaryotic organisms.

D-88

Matreshka: a Computational Genome-wide Comparative Genomics-based Screen for Possible Proteins Produced by Usage of an Alternative Overlapping Reading Frame from known Human, Mouse and Rat mRNAs

Sebastien Ribrioux[1], Syam S Tatineni[2], Adrian Bruengger, Markus R John
[1]sebastien.ribrioux@pharma.novartis.com, Novartis Institutes for BioMedical Research;
[2]syam_s.tatineni@pharma.novartis.com, Novartis Institutes for BioMedical Research

Mammalian genome-wide usage of overlapping reading frames might be a more widespread phenomenon than one would anticipate based on the few known examples. Of 8.350 orthologous triplets from RefSeq datasets, 1.784 (21%) were associated with many Kozak-positive conserved overlapping reading frame triplets, and in 4% demonstrated 28 different Prosite motifs.

D-89

A Case Study in Nutrition Genomics: Identification of Likely Candidate Genes Involved in Serum HDL-Cholesterol Levels

Laurence Parnell[1], Chao-Qiang Lai[2], Xian Adiconis, Yueping Zhu, Qiong Yang, L. Adrienne Cupples, Jose Ordovas
[1]larry.parnell@ars.usda.gov, USDA;
[2]chao.lai@tufts.edu, USDA

Nutrition research is rapidly embracing genomics and bioinformatics technologies. Combining genetics and genomics approaches, we first mapped a locus associatiing with elevated HDL3-Cholesterol and then proceeded to identify a gene with a strong likelihood to have a role in raising HDL3-C and thus potentially offer protection against cardiovascular disease.

D-90

The Frequency of Protein-coding Tandem Repeat Variants Among Proteins with a Role in Host Defence

Colm Ó'Dúshláine[1], Denis Shields[2]
[1]codushlaine@rcsi.ie, Royal College of Surgeons in Ireland;
[2]dshields@rcsi.ie, Royal College of Surgeons in Ireland

Polymorphism in host defence proteins may be selectively advantageous. We present a survey of tandem repeat polymorphism in the coding regions of genes in UniGene. Significant variation was detected. Tandem repeat sequence length variations not affecting the reading frame were overrepresented. The Gene Ontology database was used to functionally classify variants.

D-91

Conservation Profiling of Functionally-coupled Gene Clusters Through Multiple Genome Comparison

Hon-Wei Chen[1], Yo-Cheng Chang[2], Der-Ming Liou, Chuan-Hsiung Chang
[1]g39123015@ym.edu.tw, Bioinformatics Program, Institute of Health Informatics and Decision Making, National Yang-Ming University; [2]ycchang2@ym.edu.tw, Institute of Bioinformatics, National Yang-Ming University

We propose a computational approach to recognize conservation profiles of gene clusters which have similar chromosomal arrangements and are functionally-coupled in pathways shared among multiple organisms. The profiles can be used to investigate genome evolution and improve Genome Annotation by identifying missing genes. Our system will be available at http://gel.ym.edu.tw/CICP/

D-92

Cross-species Detection and Analysis of Novel Alternative Splicing Events in Human and Mouse Genes

Zhengyan Kan[1], Phil Garrett-Engele[2], Nick Tsinoremas, Jason Johnson, John Castle
[1]zhengyan_kan@merck.com, Rosetta Inpharmatics; [2]phil_garrett-engele@merck.com, Rosetta Inpharmatics

We have developed a novel method for the cross-species detection and analysis of alternative splicing, applied it to 7,454 orthologous pairs of human and mouse genes, and showed its validation as a tool for novel isoform discovery.

D-93

Chordate MRP RNase

Michael Woodhams[1]
[1]m.d.woodhams@massey.ac.nz, Allan Wilson Centre, Massey University

The MRP RNase gene has previously been identified in four mammals, frog, two plants and many yeasts. In this research, the MRP gene has been found in two fish and a sea squirt by a computer aided search of the complete genomes. Conserved features of MRP are discussed.

D-94

Non-random Gene Order in the Human Genome

Martin J. Lercher[1], Laurence D. Hurst[2]
*[1]m.j.lercher@bath.ac.uk, University of Bath;
[2]l.d.hurst@bath.ac.uk, University of Bath*

Human housekeeping genes are arranged in clusters, even when controlling for gene density and tandem duplications. This explains clustering of highly expressed genes, and of genes expressed in particular tissues. Housekeeping genes are enriched in high-GC regions, and in certain cytogenetic bands, in agreement with theories of chromatin-level gene regulation.

D-95

Discovering Regulatory Modules by Creating Profiles for Gene Ontology Biological Processes Based on Tissue-specificity Scores

Elisabetta Manduchi[1], Jonathan Schug[2], Christian J. Stoeckert Jr.
[1]manduchi@pcbi.upenn.edu, University of Pennsylvania; [2]jschug@pcbi.upenn.edu, University of Pennsylvania

We illustrate an approach to define profiles associated to gene sets of interest, based on rankings of genes according to a suitable tissue-specificity score or to a suitable regulatory module presence score. These profiles can be employed to discover regulatory modules which are specific to certain tissues and gene sets.

D-96

ECortholog: Finding Homologous Genes Using Comparative Genomics Approach

Namshin Kim[1], Seokmin Shin[2], Sanghyuk Lee
[1]deepreds@hanmail.net, Seoul National University; [2]sshin@snu.ac.kr, Seoul National University

ECortholog is a novel web-based utility for finding homologous genes in the human, mouse and rat genomes using comparative genomics approach. It utilizes the BLASTZ-based Chain/Net alignments publicly available at the UCSC genome site to find the conserved blocks in other genomes. ECortholog is available at http://genome.ewha.ac.kr/ECgene/ECortholog.

D-97

Predicting the Binding Specificity of miRNA to the 3′ Untranslated Region of Target Genes

Bram Slabbinck[1], Wim Van Criekinge[2]
[1]University of Ghent; [2]University of Ghent

MicroRNAs function as regulators by imperfectly binding the 3′ UTR of target genes. We have designed a variation of the Smith-Waterman algorithm that accounts for the imperfect multiple binding. Our algorithm can detect target genes of a given microRNA and can be used to search miRNA on a predefined target.

D-98

Identification and Characterization of Protein Subcomplexes in Yeast

Thomas Wilhelm[1], Andreas Beyer[2], Jens Hollunder
[1]wilhelm@imb-jena.de, Institute of Molecular Biotechnology, Jena, Germany; [2]beyer@imb-jena.de, Institute of Molecular Biotechnology, Jena, Germany

We present an algorithm to identify insightful substructures in protein complexes. We find already well-characterized protein machineries, but also hitherto unknown ones. We show that mRNA and protein abundances vary less in subcomplexes than in full complexes, that subcomplex proteins are less abundant in the cell and that subcomplexes are enriched with essential proteins.

D-99

Identification and Modeling of Transcription Factor Binding Sites in Seed Specific Promoters in Arabidopsis Thaliana

We present an algorithm that computes the P-value for the simultaneous occurrence of multiple motifs without relying on the simplifying assumptions that promoter regions are uncorrelated sequences and that the scores for different motifs are statistically independent, and we use this algorithm for a genome-wide identification of ABI3, FUS3, and LEC1 binding sites in Arabidopsis thaliana.

D-100

SNP2CAPS: a Tool for CAPS Marker Development and SNP and INDEL Analysis

Thomas Thiel[1], Raja Kota, Ivo Grosse, Nils Stein, and Andreas Graner
[1]thiel@ipk-gastersleben.de, Institute of Plant Genetics and Crop Plant Research Gatersleben, Germany

We present a computer program, SNP2CAPS, that facilitates the computational conversion of SNP sites into CAPS makers with the goal of providing a robust and cost-effective SNP assay where applicable.

D-101

The MIAPE Standard, The PSI Object Model and What's Next ...

Chris Taylor[1]
[1]christ@ebi.ac.uk, EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK. CB10 1SD

The volume and complexity of data generated by proteomics continues to increase. A standard model and minimum reporting requirement are both needed to facilitate analysis, dissemination and exchange of proteomics data. We present the MIAPE standard, the PSI object model and ontology, tools and reference implementations (markup language and repository)

D-102

An Analysis and Discovery Framework for the Integration of QTL and Gene Expression Data

Robert Hitzemann[1], Shannon McWeeney[2], Christina A. Harrington, Barry Malmanger,Maureen Lawler and John Belknap
[1]Department of Behavioral Neuroscience, [2]mcweeney@ohsu.edu, Division of Biostatistics, Department of Public Health & Preventive Medicine

We compare 5 different low-level analysis methods and determine the ability of each to detect putative differential expression in the collection of surveyed transcripts. Our assessment of each method takes into account model-based estimates of the proportion of false positives and negatives.
A "consensus" analysis strategy is discussed

D-103

The Functional Genomics Experiment Object Model: Integrated Data Standards

Andrew Jones[1], Angel Pizarro[2], Ela Hunt, Jonathan Wastling and Christian J. Stoeckert Jnr.
[1]jonesa@dcs.gla.ac.uk, Department of Computing Science, University of Glasgow;[2]angel@pcbi.upenn.edu, Center for Bioinformatics, University of Pennsylvania

We have created the Functional Genomics Experiment Object Model (FGE-OM) by integrating the microarray and proteomics data standards (MAGE-ML and PEDRo). FGE-OM provides a core framework for describing all Functional Genomics experiments in a common format, and will contribute to the next version of MAGE and future proteomics standards.

E-1

Proteomics and Bioinformatic Studies of the Newly Sequenced Cyanophage SPM-2

Konstantinos Thalassinos[1], James H. Scrivens[2], Martha Clokie, Susan E. Slade, Nicholas H. Mann
[1]k.thalassinos@warwick.ac.uk, University of Warwick;
[2]j.h.scrivens@warwick.ac.uk, University of Warwick

A proteomics and bioinformatics study was undertaken for the newly sequenced cyanophage SPM-2. The proteome was predicted by two programs and the differences between the two were assessed. Twenty-three structural proteins were identified by mass spectrometry and the data mapped back to the genome.

E-2

Analysis of Leishmania Chagasi Transcriptome Data: From ESTs to Relevant Information in a Pathogenomics Approach

Gregory B. Clark[1], Diana M. Oliveira[2], Nilo B. Diniz, Daniel A. Viana, João J. S. Gouveia, Elton J. R. Vasconcelos, Michely C. Diniz, Marianna C. Albuquerque, Thiago D. Ferreira, Ana Carolina L. Pacheco, Raimundo B. Costa, Adriana R. Tome, Rodrigo Maggioni, Jennifer Tsai, Bhooma Thiruvahindrapuram
[1]greg.clark@utoronto.ca, University of Toronto; [2]nugen-l@lcc.uece.br, Universidade Estadual do Ceara

To analyze an EST sequencing project, transcriptome of protozoan Leishmania chagasi (performed by the Brazilian Genomics Program – PROGENE - http://nugen.lcc.uece.br/progene), we have used a compilation of bioinformatics procedures. Results are shown in terms of relevance to structural and functional identity to probe genes related to parasite pathogenicity during visceral leishmaniasis.

E-3

A myGrid Based Nucleotide Annotation Tool

Tracy Craddock[1], Anil Wipat[2], Peter Li
[1]tracy.craddock@ncl.ac.uk, Computing Science, University of Newcastle; [2]anil.wipat@ncl.ac.uk, Computing Science, University of Newcastle

We have developed a tool that demonstrates the value of myGrid to the biologists user. myGrid technology has been used to create a simple sequence annotation tool with a Web interface that utilises myGrid services, workflows and middleware in order to integrate different applications and data sources to annotate previously uncharacterised nucleotide sequences.

E-4

An Adaptive Sliding Window Technique for Inferring DNA Functionality from Sequence Information

Irenaeus te Boekhorst[1], Irina Abnizova[2], Maria Schilstra, Chrystopher Nehaniv
[1]r.teboekhorst@herts.ac.uk, University of Hertfordshire;
[2]irina.abnizova@mrc-bsu.cam.ac.uk, MRC-BSU, Cambridge

We present a content-based approach to investigate statistical properties of coding regions in eukaryotic DNA. Our method is based on a new optimisation technique applied to rescaled range analysis and entropy measurements. This technique is window size independent, and performs an unsupervised search. Preliminary applications allowed identifying the borders of functional regions in DNA.

E-5

PANDORA: Keyword-based Analysis of Protein Sets by Integration of Annotation Sources

Noam Kaplan[1], Michal Linial[2]
[1]kaplann@cc.huji.ac.il, Hebrew University;
[2]michall@cc.huji.ac.il, Hebrew University

PANDORA is a web-based tool that provides automatic graphical representation of the biological knowledge associated with any set of proteins. PANDORA uses a unique approach of keyword-based graph analysis that focuses on detecting subsets of proteins that share unique biological properties and the intersections of such sets Available at: http://www.pandora.cs.huji.ac.il

E-6

Mapping SAGE Tags to Genes Using EST and Genomic Information

Tim Beissbarth[1], Lavinia Hyde[2], Seong-Seng Tan, Terry Speed, Hamish Scott
[1]beissbarth@wehi.edu.au, WEHI; [2]hyde@wehi.edu.au, WEHI

SAGE is a method for high-throughput screening of gene-expression and identification of transcriptomes without prior knowledge of gene sequences. Here we try to make use of SAGE data, that we generated from various brain libraries, as well as publically available EST and genomic information to annotate and identify genes.

E-7

Structure Based Identification of Protein Family Signatures for Function Annotation

Ruchir Shah[1], Luke Huan[2], Deepak Bandopadhyay, Wei Wang, Alexander Tropsha
[1]ruchir@email.unc.edu, UNC-CH; [2]huan@cs.unc.edu, UNC-CH

We present a novel approach to identifying recurrent structure-sequence motifs common to particular protein families. The approach employs Delaunay tessellation to generate protein graphs and frequent subgraph mining to obtain the motifs. We demonstrate the utility of these motifs for highly accurate annotation of several protein families.

E-8

Marsupial Informatics: a Comprehensive Database System for Lactation Genomics in the Tammar Wallaby

Christophe Lefèvre[1], Yvan Strahm[2], Matthew Digby, Kevin Nicholas
[1]chris.lefevre@med.monash.edu.au, Victorian Bioinformatics Consortium, Monash University; [2]yvan.strahm@med.monash.edu.au, Victorian Bioinformatics Consortium, Monash University

A database system for the annotation and analysis of marsupial lactation genomics data of the tamar wallaby is presented. The pipeline for the predictive and functional annotation of Est data is based on mySQL database, open source software and a custom web interface developed in PHP.

E-9

Supra-domains - Evolutionary Units Larger than Single Protein Domains

Vogel C[1], Berzuini C[2], Bashton M, Gough J, Teichmann SA
[1]cvogel@mrc-lmb.cam.ac.uk, MRC Laboratory of Molecular Biology; [2]carlo.berzuini@mrc-bsu.cam.ac.uk, MRC Biostatistics Unit

Supra-domains are combinations of two or more domains that recur in different protein contexts. They represent evolutionary units larger than single protein domains. We describe the properties of those supra-domains.

E-10

Improved Prediction of Signal Peptides - SignalP 3.0

Jannick D. Bendtsen[1], Henrik Nielsen[2], Anders Krogh, Gunnar von Heijne and Soren Brunak
[1]jannick@cbs.dtu.dk, Center for Biological Sequence analysis, BioCentrum-DTU; [2], Center for Biological Sequence analysis, BioCentrum-DTU

We describe improvements of the popular method for prediction of classically secreted proteins, SignalP. SignalP consists of two different predictors based on neural network and hidden Markob model algorithms, where both components have been updated. This combined with currated data set, have improved the performance of the predictor significantly over SignalP version 2.

E-11

GENCOLORS: a Tool for GENomic COmparative Analysis Based on LOw-Redundant Sequencing

A. Romualdi[1], R. Lehmann[2], G. Glöckner, M. Platzer, J. Sühnel
[1]romualdi@imb-jena.de, Biocomputing Group, Institute of Molecular Biotechnology, Jena Centre for Bioinformatics; [2]rleh@imb-jena.de, Department of Genome Analysis, Institute of Molecular Biotechnology, Jena Centre for Bioinformatics

GENCOLORS is a web portal providing easy access to sequence and annotation data of both bacterial genome projects adopting a low-redundany sequencing strategy and published genomes with an emphasis on genome comparison. A specific tool also accessible in a separate version (www.imb-jena.de/JPGV/) generates circular genome plots with many unique features.

E-12

Multifactorial Regulation in Yeast: Looking for Correlated Bindin Sequences

Michele Caselle[1], Davide Cora[2]
[1]caselle@to.infn.it, Department of Theoretical Physics, University of Torino; [2]cora@to.infn.it, Department of Theoretical Physics, University of Torino

The determination of regulatory pathworks is further complicated by coregulation. We examined all couples of 6,7,8 and 9 basis in 500bp upstream in yeast and we determined those whose mean distance or variance have a Bonferroni corrected P-value<0.01. We screened the resulting couples by using GO annotations.

E-13

Finding Specific Protein Families in Newly Sequenced Genomes

Intikhab Alam[1], Georg Fuellen[2]
[1]intikhab@cebitec.uni-bielefeld.de, International Graduate School in Bioinformatics and Genome Research, University of Bielefeld; [2]fuellen@alum.mit.edu, Medizinische Fakultät, c/o Arbeitsgruppe Bioinformatik

To find out a maximum number of possible homologues to a set of proteins in question, We present a tool called GenCHASE, or Genomic Comparative Homology Agreement Search, that combines extrinsic (similarity based gene finding) and intrinsic (ab initio based gene structure prediction) approaches.

E-14

CHRAB - A Knowledge Base for Chromatin Associated Proteins and Protein Domains

Carsten Helgesen[1], Katharina R. Tufteland[2], Gisle Sælensminde, Pal Puntervoll, Karin Wibrand, Reidunn B. Aalen, Rein Aasland
[1]carsten.helgesen@hib.no, Bergen University College; [2]katharina.tufteland@student.uib.no, University of Bergen

CHRAB is a database for management of facts, knowledge and detailed functional annotation of chromatin-associated proteins and protein domains.The main focus is on protein domains, represented as HMMs, but other types of biological objects like functional sites, structures can easily be added. The CHRAB prototype is freely accessible at http://www.bioinfo.no/tools/CHRAB.

E-15

Regulatory RNA Elements - A Motif Search in Cyanobacterial Genomes

Philip Kensche[1], Ilka[2], Stefan Kohl, Hans-Peter Herzl, Wolfgang R. Hess, Jörg Vogel
[1]p.kensche@biologie.hu-berlin.de, Humboldt University Berlin - Institute for Theoretical Biology Berlin; [2]Ilka.Axmann@gmx.de, Axmann

We report the results of a screening for RNA elements in the intergenic regions of multiple cyanobacterial species. The approach applies RNA secondary structure prediction and covariance information to sequences clusters to identify those with conserved structure.

E-16

Quality Ranked Assignment of Gene Ontology Terms to Gene Products - the GOtcha Method

David M A Martin[1], Matthew Berriman[2], Geoffrey J Barton
[1]d.m.a.martin@dundee.ac.uk, School of Life Sciences, University of Dundee; [2]mb4@sanger.ac.uk, The Sanger Institute

GOtcha is a novel method that employs the heirarchical nature of Gene Ontology (GO) in transitive assignment of GO terms to sequences. Unlike most current methods, GOtcha provides a probability score for each GO term rather than just the leaf terms and shows significantly better specificity and selectivity than taking the top BLAST hit.

E-17

Design of Automatic Data Analysis and Annotation Tools

Coral del Val[1], Agnes Hotz-Wagenblatt[2], Karl-Heinz Glatting, Barbara Pardon, Mechthilde Falkenhahn, Sándor Suhai
[1]c.delval@dkfz.de, German Cancer Research Center; [2]hotz-wagenblatt@dkfz.de, Cancer Research Center

High-throughput approaches need, besides suitable data integration procedures, a high degree of automation. This implies running these analyses in a serial rule-dependent fashion (workflow) and in parallel when possible. In this context we have developed different analysis tools for semi-automatic annotation of sequences. Availability: http://genius.embnet.dkfz-heidelberg.de/menu/biounit/open-husar/

E-18

AstraZeneca GeneCatalogue - A Novel View of the Human Transcriptome and Proteome

Ellen Thomas[1], Nina Mian[2], Pete Edwards, Bob Dewhirst, James Todd, Tekpeki Anim, Mark Downey-Jones, David Tilley, Bo Servenius, Krzysztof Pawlowski, Anne Westcott, Mathew Woodwark, Mike Firth & Ian Dix
[1]ellen.thomas@astrazeneca.com, AstraZeneca; [2]nina.mian@astrazeneca.com, AstraZeneca

AZ GeneCatalogue is AstraZeneca's repository for transcript, protein, EST, SNP and reference expression data. Here we present an overview of the system methodology, a critique of the information available and insight into the interface designed to support drug-discovery programmes in AstraZeneca.

E-19

WESTAP - A Web-based EST Annotation Pipeline

Paulo B. Paiva[1], Marco A.G. Ribeiro[2], Daniel Sigulem[3], Marcelo Avedissian[4], Marcelo R.S. Briones[5]
[1]paiva@compbio.epm.br, Health Informatics Department, Federal University of São Paulo; [2]marco@compbio.epm.br, Health Informatics Department, Federal University of São Paulo; [3]sigulem@dis.epm.br, Health Informatics Department, Federal University of São Paulo; [4]avedissian@ecb.epm.br, Physiology Department, Federal University of São Paulo; [5]marcelo@ecb.epm.br, Microbiology-Immunology-Parasitology Department, Federal University of São Paulo

WESTAP is a customizable tool for processing, data management and pre-annotation of sequence data generated by EST sequencing projects. The tool provides a web-based interface for common steps in EST processing, such as data submission, base-calling, vector screening, read storage, assembly and pre-annotation. WESTAP is available freely at http://compbio.epm.br/westap.

E-20

Large-scale Annotation of Coding nsSNPs Based on Structure Considerations

Rachel Karchin[1], Ursula Pieper[2], Mark Diekhans, Daryl Thomas, Fred Davis, Eswar Narayanan, Marc Marti-Renom, Andrea Rossi, David Haussler, Andrej Sali
[1]rachelk@salilab.org, Biopharmaceutical Sciences, University of California, San Francisco; [2]ursula@salilab.org, Biopharmaceutical Sciences, University of California, San Francisco

Coding region polymorphisms that result in amino acid residue changes [i.e., non-synonymous cSNPs (nsSNPs)] are of critical importance in human disease and drug sensitivities. We present results of a genome-wide mapping of human nsSNPs onto comparative protein structure models and identify nsSNPs likely to have functional impacts on human health.

E-21

Some Statistical Properties of Regulatory DNA Sequences, and How to Use Them to Distinguish Regulatory Regions

Irina Abnizova[1], Walter Gilks[2], Rene te Boekhorst
[1]irina.abnizova"mrc-bsi.cam.ac.uk, MRC-BSU; [2]walter.gilks@mrc-bsu.cam.ac, MRC-BSU

We present a statistical approach to distinguish regulatory regions from DNA sequence information. Thus we introduce a new measure of DNA heterogeneity, and a new statistical test for abundance of similar words in regulatory regions. Our methods are tested on eukaryotic DNA from several species, and can be used as complementary annotation tools.

E-22

BCED: A Base-calling Error Detection System to Improve Annotation Accuracy in Microbial Genome Projects

Hongseok Tae[1], Daesang Lee[2], Hyeweon Nam, Kiejung Park
[1]mbio94@netian.com, Dept. of Computer Engineering, Chungnam National University, Gung-Dong 220, Yusung-Gu, Daejeon, 305-764, South Korea; [2]dslee@smallsoft.co.kr, Information Technology Institute, SmallSoft Co., Ltd., Junmin-Dong 461-71, Yusung-Gu, Daejeon, 305-811, South Korea

We have developed a base-calling error detection system in microbial genome projects (BCED). BCED detects dubious bases candidates in a few aspects and shows the list of the genes or sequences that are suspected to contain base-calling errors. This system may be helpful to improve the quality of Genome Annotation.

E-23

Repetitive DNA in the Complete Genome Sequence of the Heartwater Agent, Ehrlichia Ruminantium

Junita Liebenberg[1], Nicola E. Collins[2], Elmarié Louw, F. Erika Faber, Mirinda van Kleef, Basil A. Allsopp
[1]junita@moon.ovi.ac.za, Onderstepoort Veterinary Institute; [2]nicola.collins@op.up.ac.za, University of Pretoria

The genome sequence of Ehrlichia ruminantium, an intracellular rickettsia which causes heartwater, was completed recently. The most striking feature of the genome is the large number of tandem repeats and dispersed repeat sequences. Some of the repeat regions contain a variable number of repeat motifs and in at least one instance two genes appear to have been duplicated and fused.

E-24

Annotation of the Smallest Eukarytic Genome: Ostreococcus Tauri

Stephane Rombauts[1], Sven Degroeve[2], Stephane Rombauts1, S. Degroeve1, S. Robbens1, C. Ferraz3, E. Derelle2, R. Cooke4, M. Delseny4, J. Demaille3, A. Picard2, P. Rouze1, H. Moreau2 and Y. Van de Peer1
[1]strom@psb.ugent.be, Department of Plant Systems Biology; [2]svgro@psb.ugent.be, Department of Plant Systems Biology

In collaboration with the Laboratoire Arago, Banyuls, France, we are performing the full Genome Annotation of the unicellular green alga Ostreococcus tauri. This alga is the smallest eukaryotic organism described until now (comparable to a bacterium) and has a nuclear genome of about 11.5 Mb, divided over 18 chromosomes

E-25

The Annotation Status of Schizosaccharomyces Pombe: Gene Ontology and Protein Family Coverage

Valerie Wood[1], Martin Aslett[2]
[1]val@sanger.ac.uk, Wellcome Trust Sanger Institute; [2]maa@sanger.ac.uk, Wellcome Trust Sanger Institute

The fission yeast Schizosaccharomyces pombe has been sequenced and annotated. Curation is being carried out as part of the GeneDB project. Fission yeast has already proved to be an excellent model for processes related to cell cycle, replication and recombination.

E-26

Protein DAS, Distributed Annotation System for Proteins

A. Prlic[1], T.A. Down[2], T.J.P. Hubbard
[1]ap3@sanger.ac.uk, The Wellcome Trust Sanger Institute; [2]td2@sanger.ac.uk, The Wellcome Trust Sanger Institute

DAS,the distributed annotation system, provides a communication protocol used to exchange annotations between decentralized data sources. We give an overview of extensions to the original DAS protocol in order to provide a client which displays DNA data like SNPs, or Intron and Exon borders, projected onto protein sequences and structures.

E-27

A Higher Order Assembly of the Fugu Genome

Tanya Vavouri[1], Yvonne JK Edwards[2], Martin Goodson, Greg Elgar
[1]tvavouri@rfcgr.mrc.ac.uk, MRC RFCGR; [2]yjedward@rfcgr.mrc.ac.uk, MRC RFCGR

The genome of the pufferish Fugu rubripes in the latest assembly (v 3.0) was highly fragmented. Following a series of simple heuristic steps, we have constructed a more cohesive assembly. Approximately 245Mb of the genome is now contained in 910 regions containing scaffolds ordered and oriented in relation to each other.

E-28

Vertebrate Genome Annotation

J.L. Ashurst[1], L. Wilming[2], Frankish A, Gilbert J.G.R., Grocock R., Hart E., Keenan S., Laird G., Loveland J., Rajan J., Sehra H., Steward C., Swarbreck D., Hubbard T
[1]jla1@sanger.ac.uk, Wellcome Trust Sanger Institute; [2]lw2@sanger.ac.uk, Wellcome Trust Sanger Institute

The Havana group at the Sanger Institute specialises in the manual annotation of finished genomic sequence, which utilises all methods of automatic gene builds in addition with comparative analysis. VEGA, the vertebrate Genome Annotation browser, is a publicly available browser (http://vega.sanger.ac.uk) that is specifically designed to view manually-curated genome data.

E-29

KEGG DAS: Towards a Comprehensive Community Annotation of the Genome

Toshiaki Katayama[1], Minoru Kanehisa[2]
[1]k@bioruby.org, Human Genome Center, Institute of Medical Science, University of Tokyo, Japan; [2]kanehisa@kuicr.kyoto-u.ac.jp, Bioinformatics Center, Institute for Chemical Research, Kyoto University, Japan

KEGG DAS is a new service to access the KEGG annotation for all organisms with BioDAS and GMOD/GBrowse. Our goal is to provide comprehensive reference database for the community annotation of each genome in a unified manner. The service is available at http://das.hgc.jp.

E-30

GenDB-2.0 - Recent Developments of the Open Source Genome Annotation System

Alexander Goesmann[1], Burkhard Linke[2], Daniela Bartels, Alexander Lenhardt, Alice C. McHardy, Heiko Neuweger, Oliver Rupp, Alfred Pühler, Folker Meyer
[1]Alexander.Goesmann@CeBiTec.Uni-Bielefeld.DE, CeBiTec/BRF; [2]Burkhard.Linke@CeBiTec.Uni-Bielefeld.DE, CeBiTec/BRF

The GenDB Genome Annotation software has been developed during the last four years as an extensible open source system. The current version enhances the capabilities of our system for comparative and functional analyses. Furthermore, we present the novel GenDB web frontend and an optimized strategy for automated annotation of coding regions.

E-31

Prophage Detection in Bacterial Genomes

Gipsi Lima[1], Raphael Leplae[2],Shoshana Wodak and Ariane Toussaint
[1]gipsi@scmbb.ulb.ac.be, ULB;
[2]raphael@scmbb.ulb.ac.be, ULB

ACLAME is a database dedicated to a classification of prokaryotic Mobile Genetic Elements including phages and prophages (Leplae et al. 2004). We developed a system to identify prophages in sequenced bacterial genomes, that includes identification of integrases, prophage boundaries, and evaluation of sequence similarity with phage proteins

E-32

ACLAME: A Database for the Reticulate Classification of the Prokaryotic Mobilome

Leplae[1], Lima Mendez[2], Wodak, Toussaint
[1]raphael@scmbb.ulb.ac.be, ULB;
[2]gipsi@scmbb.ulb.ac.be, ULB

The ACLAME database proposes a classification of prokaryotic Mobile Genetic Elements (MGEs). The MGE proteins are automatically classified and their functional annotation manually curated via the contribution of the scientific community. The database can be accessed via a web interface: http://aclame.ulb.ac.be

E-33

A Microbial Genome Annotation System Including Comparative Genomic Analysis

Hwajung Seo[1], Hyeweon Nam[2], Daesang Lee, Hongseok Tae, Kiejung Park
[1]hjseo@smallsoft.co.kr , SmallSoft Co., Ltd.;
[2]hwnamyh@yahoo.co.kr, Chungnam National University

We have developed a web-based microbial Genome Annotation system, which provides web-based interfaces for gene prediction, homology search, promoter analysis, motif analysis and gene ontology analysis. A database searching module, a linear map browser and gene classification viewer was implemented. We have developed a few comparative genomic analysis programs.

E-34

Genome Curation and Annotation at The Rat Genome Database (RGD)

Dean Pasko[1], Dean Pasko, Simon Twigger, Mary Shimoyama, Susan Bromberg, Jiali Chen, Norberto de la Cruz, Chunyu Fan, Cindy Foote, Glenn Harris, Jed Mathis, Nataliya Nenasheva, Rajni Nigam, Victoria Petri, Dorothy Reilly, Weiye Wang, Lan Zhao, Wenhua Wu, Angela Zuniga-Meyer,Peter Tonellato, Howard Jacob
[1]dpasko@mcw.edu, Rat Genome Database, Medical College of Wisconsin

RGD is developing an infrastructure to annotate the rat genome. Open source tools will be used to annotate and share the data. These tools include the Apollo Genome Annotation and curation tool, the GBrowse genomic browser tool, and the Distributed Annotation System (DAS) allowing other websites to view RGD's annotations.

E-35

The Mechanism of Formation of Chimeric ESTs

Eitan Rubin[1]
[1]erubin@cgr.harvard.edu, Bauer Center for Genomics Research, Harvard University

A common artifacts in ESTs are chimera. Here we present a method for reliable detection of chimeric ESTs, and methodological analysis of their structure. Using genomic DNA, we identified 41,000 putative chimera. Analysis of these chimera supports template switch as a major source of chimeric ESTs.

E-36

Exogean: an Expert on Eukaryotic Gene Annotation

Sarah Djebali[1], Franck Delaplace[2], Hugues Roest Crollius
[1]djebali@ens.fr, Genome Organisation and Dynamics Laboratory, ENS; [2]delapla@lami.univ-evry.fr, Laboratoire de Methodes Informatiques

Automatic gene annotation is a basic requirement to identify the precise location of all genes in a genome. We have developped Exogean, an original approach for automatic gene annotation based on mathematical modelling of human expertise. We present encouraging results given by Exogean on human chromosome 22.

E-37

Using Catalygtic Residue Conversation for Function Assignment

Ruth V Spriggs[1], Richard A George[2], Gail J Bartlett, Alex Gutteridge, Malcolm W Macarthur, Craig T Porter, Bissan Al-Lazikani, Mark B Swindells, Janet M Thornton
[1]rspriggs@Inpharmatica.co.uk, EMBL-EBI; [2]richardg@Inpharmatica.co.uk, EMBL-EBI

Knowledge of residues required for catalysis, determined from structure, is used to increase confidence in function inference through sequence homology. When residue conservation is used, compared to homology alone, we see a 79% increase in the proportion of correctly inferred functions for homologues detected at the third iteration of PSI-BLAST.

E-38

BioSapiens: A European Network for Integrated Genome Annotation

Janet Thornton[1], The BioSapiens Consortium[2]
[1]thornton@ebi.ac.uk, European Bioinformatics Institute; [2],

BioSapiens is a European network for integrated Genome Annotation, funded by the EU's 6th Framework Programme, and made up of bioinformatics researchers from 25 institutions in 14 countries. The objective is to provide a large scale, concerted effort in Genome Annotation, using both informatics tools and input from experimentalists.

E-39

prot4EST: Translating Expressed Sequences from Neglected Genomes

JJames Wasmuth[1], Mark Blaxter[2]
[1]james.wasmuth@ed.ac.uk, University of Edinburgh; [2]mark.blaxter@ed.ac.uk, University of Edinburgh

prot4EST exploits available methods to robustly predict peptides from Expressed Sequence Tags. It overcomes the lack of prior sequence data required to train current methods. Additional quality information is also provided which can be set up as an easy to query database. Available from: http://www.nematodes.org/PartiGene

E-40

Retroviral Sequences of the Human and Chimpanzee Genomes

Tove Airola[1], Göran Sperber[2], Jonas Blomberg
[1]tove.airola@medsci.uu.se, Uppsala University; [2]goran.sperber@neuro.uu.se, Uppsala University

RetroTector is a program that detects and characterizes retroviral sequences in genomic sequence collections. It is based on an explicit data model of retroviruses and attempts to classify genus and reconstruct the viral proteins. Analysis of the human genome yielded 9506 retroviral sequences with an average length of 7334 bp.

E-41

Annotation and Sequence Analysis of a Gene-rich Region Deleted in the Del(13)Svea36H Mouse

A.-M. Mallon[1], L. Wilming[2], J. Weekes, J.G.R. Gilbert, J. Ashurst, S. Peyrefitte, L. Matthews, M. Cadman, R. McKeone, C. Sellick, R. Arkell, M.R.M. Botcherby, M.A. Strivens, R.D. Campbell, S. Gregory, P. Denny, J.M. Hancock, J. Rogers & S.D.M. Brown
[1]a.mallon@har.mrc.ac.uk, MRC Mammalian Genetics Unit; [2]lw2@sanger.ac.uk, Wellcome Trust Sanger Institute

We describe the results of the annotation of 12.7 Mb of finished sequence from mouse chromosome 13. The region shows associations between gene clusters and transposable elements consistent with a role as gene factories. The region also contains numerous non-coding Evolutionarily Conserved Regions, which are candidate regulatory sequences.

E-42

metaSHARK: a Database of Automated Metabolic Reconstructions Derived from Genomic DNA Sequence

John W Pinney[1], David R Westhead[2], Glenn A McConkey
[1]john@bioinformatics.leeds.ac.uk, University of Leeds; [2]d.r.westhead@leeds.ac.uk, University of Leeds

We present the metabolic SearcH And Reconstruction Kit (metaSHARK), an online resource for the visualisation, navigation and analysis of fully automate metabolic reconstructions for a variety of organisms. metaSHARK is available online at http://bioinformatics.leeds.ac.uk/shark/ We welcome requests for the analysis of additional genomes.

E-43

Annotation and Representation of a Transcriptome: an Integrated Approach

Daniel Lang[1], Ralf Reski[2], Stefan A. Rensing
[1]Daniel.Lang@biologie.uni-freiburg.de, University of Freiburg, Plant-Biotechnology; [2]Ralf.Reski@biologie.uni-freiburg.de, University of Freiburg, Plant-Biotechnology

We present an integrated approach to generate a knowledge resource for the transcriptome of Physcomitrella patens. Each transcript is represented as an entity containing full annotation and GO associations. The whole clustering process is modeled and results in 5 datasets representing the annotated Physcomitrella transcriptome with different granularity. http://www.cosmoss.org.

E-44

The Encyclopedia of Life: A New Web Resource for Domain-based Protein Annotation

Gregory B Quinn[1], Mark A Miller[2], Kim Baldridge, Ilya Shindyalov, Wilfred Li, Dmitry Pekurovsky, Robert W. Byrnes, Kristine Briedis, Vicente Reyes; Adam Birnbaum, Coleman Mosley, Yohann Potier, Celine Amoreira, Stella Veretnik, Philip E. Bourne
[1]quinn@sdsc.edu, San Diego Supercomputer Center; [2]mmiler@sdsc.edu, San Diego Supercomputer Center

The Encyclopedia of Life (EOL) project seeks to provide researchers with domain-based annotation for putative protein sequences using a rigorously benchmarked integrated Genome Annotation pipeline (iGAP). Data from the software pipeline is presented through an innovative and highly functional web interface. EOL can be found on the web at http://www.eolproject.info

E-45

Automatic Functional Annotation of Protist ESTs

Liisa Koski[1], Gertraud Burger[2], Franz Lang
[1]lkoski@bch.umontreal.ca, Universite de Montreal; [2]Gertraud.Burger@Umontreal.ca, Universite de Montreal

The Protist EST Program (PEP) has sequenced >20 protist organisms to date. We have developed an Automatic Functional Annotation System to assign consensus EST sequences with the most informative possible description. We will present this automatic annotation pipeline, compare it to manual annotations and discuss/compare the system requirements.

E-46

An Advanced Approach of Combining Gene Ontology with Machine Learning for Gene Function Prediction

A.Vinayagam[1], Rainer König[2], Jutta Moormann, Falk Schubert, Roland Eils, Karl-Heinz Glatting, Sándor Suhai
[1]A.Vinayagam@dkfz-heidelberg.de, Deutsche Krebsforschungszentrum (DKFZ); [2]r.koenig@dkfz.de, Deutsche Krebsforschungszentrum (DKFZ)

We developed an automated annotation system that assigns a confidence value for each prediction. In our approach we efficiently combined the Gene Ontology information with co-operative sets of Support Vector Machines. Our system performance is benchmarked with 13 model organisms and applied to annotate Xenopus laevis contig sequences.

E-47

Eukaryotic Genome Annotation with Gnomon - a Multi-step Combined Gene Prediction Tool

Alexandre Souvorov[1], Tatiana Tatusova[2], David Lipman
[1]souvorov@ncbi.nlm.nih.gov, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, 8600 Rockville Pike, Bethesda, MD 20894, USA; [2]tatiana@ncbi.nlm.nih.gov, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, 8600 Rockville Pike, Bethesda, MD 20894, USA

Gnomon uses a set of heuristics to find the maximal self-consistent set of corresponding transcript and protein alignment data to set the constraints for an Hidden Markov Model(HMM)-based gene prediction and then predicts the gene structure in genomic DNA sequences in a multi-step fashion.

E-48

GO On, Tell Us What You Think: Community Input into the Gene Ontology

Jennifer I Clark[1], Midori Harris[2], Jane Lomax, Amelia Ireland and the Gene Ontology Consortium.
[1]jclark@ebi.ac.uk, EMBL-European Bioinformatics Institute; [2]midori@ebi.ac.uk, EMBL-European Bioinformatics Institute

The utility of our controlled vocabularies to biologists is the primary concern of the Gene Ontology Consortium. We have established extensive user feedback systems to identify the critical biological research needs, and to address these in practical ways

E-49

A Methodology Towards In-depth Annotation of Subtelomeres

Renauld[1], Keely[2], Stringer, Barrell and Hall
[1]hjr@sanger.ac.uk, Sanger Institute;
[2]keelysp@ucmail.uc.edu, University of Cincinnati

Progresses in the development of methods and a methodology for an in-depth annotation of subtelomeres ('junk' DNA) are presented, using sequences from gene arrays found at Pneumocystis carinii chromosomal ends.

E-50

Identification and Characterization of Processed Pseudogenes in Arabidopsis Thaliana

David Benovoy[1], Guy Drouin[2]
[1]davidbenovoy@hotmail.com, University of Ottawa;
[2]gdrouin@science.uOttawa.ca, University of Ottawa

To identify the processed pseudogene population in Arabidopsis thaliana, we developed a search algorithm using features of processed pseudogenes. We then studied the nature of genes giving rise to processed pseudogenes, the distribution, GC content and the mutation patterns of processed pseudogenes. We also have found processed genes that might still be functional

E-51

Gene Prediction and Start Codon Selection in GC-rich Genomes

Michaela Falb[1], Friedhelm Pfeiffer[2], Dieter Oesterhelt
[1]falb@biochem.mpg.de, MPI for Biochemistry;
[2]fpf@biochem.mpg.de, MPI for Biochemistry

The unreliability of automatic gene finders on GC-rich genomes (vast gene overprediction, start codon misassignments) was improved for two halophilic strains by intergenome comparison and integration of proteomic data. Currently tools for gene and start codon selection are developed based on analyses of amino acid compositions and pI values.

E-52

Genome Reviews: Integrated Views of Complete Genome Sequences from the EMBL/GenBank/ DDBJ database

Lorna Morris[1], Paul Kersey[2], Nadeem Faruque, Tamara Kulikova, Eleanor Whitfield, Rolf Apweiler
[1]lmorris@ebi.ac.uk, European Bioinformatics Institute;
[2]pkersey@ebi.ac.uk, European Bioinformatics Institute

Genome Reviews represent curated versions of complete genome entries from the EMBL/GenBank/ DDBJ sequence repository, that include additional annotation imported from other data sources. The Genome Reviews project provides a regularly updated, standardised and comprehensively annotated view of this data in an EMBL-compatible format (http://www.ebi.ac.uk/GenomeReviews).

E-53

Computational Mapping of Barley ESTs Using the Synteny to Rice

Thomas Thiel[1], Andreas Graner[2], Stefan Posch[3], Nils Stein[2], Rajeev K. Varshney[2], Ivo Grosse[2]
[1]thiel@ipk-gastersleben.de, Institute of Plant Genetics and Crop Plant Research Gatersleben, Germany; [2]Institute of Plant Genetics and Crop Plant Research Gatersleben, Germany; [3]Department of Mathematics and Computer Science, Halle University, Germany

We present a pipeline for the computational prediction of barley EST mapping positions on the IPK barley transcript map using the synteny between barley and rice.

F-1

Analysis of a Chemical-Genomic Network by Integrating the Information of Literatures and Microarray Technology

Yasushi Okuno[1], Gozoh Tsujimoto[2]
[1]okuno@pharm.kyoto-u.ac.jp, Graduate School of Pharmaceutical Sciences, Kyoto University; [2]gtsuji@pharm.kyoto-u.ac.jp, Graduate School of Pharmaceutical Sciences, Kyoto University

We have developed a Data Mining method to extract implicit biomedical information by linking gene expression information on microarray to chemical and medicinal compounds. This method can discover the new candidates which are impossible to be detected by microarray experiments, but which can have a causal association with the medicinal compound and the disease.

F-2

Identification of Growth-associated Regulatory Circuit of FadR by Transcriptome Profiling

Jin Hwan Park[1], Byung Hun Kim[2], Jong Hyun Choi, Dong-Yup Lee, Sang Yup Lee
[1]jinhwan@kaist.ac.kr, Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical & Biomolecular Engineering and BioProcess Engineering Research Center, Korea Advanced Institute of Science and Technology; [2]ByungHunKim@kaist.ac.kr, Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical & Biomolecular Engineering and BioProcess Engineering Research Center, Korea Advanced Institute of Science and Technology

The identification of regulatory circuit of global regulators is a prerequisite for the construction of industrially useful strains. In this study we examined the regulatory circuit of FadR, which is one of the global regulators in Escherichia coli, at transcriptional level using DNA microarray experiment.

F-3

In Silico Microdissection of Microarray Data from Heterogeneous Cell Populations

Harri Lähdesmäki[1], Ilya Shmulevich[2], Valerie Dunmire, Olli Yli-Harja, Wei Zhang
[1]harri.lahdesmaki@tut.fi, Institute of Signal Processing, Tampere University of Technology; [2]is@ieee.org, Cancer Genomics Laboratory, The University of Texas M. D. Anderson Cancer Center

We propose a computational framework for removing the effects of sample heterogeneity on gene expression measurements. For the cases where no information about the mixing percentages of different cell types is available, we develop an optimization-based method for joint estimation of the mixing percentages and the expression values of the pure cell samples

F-4

A Fully Bayesian Model to Cluster Gene Expression Profiles

Fatima Sanchez-Cabo[1], Claus Vogl[2], Simon Hubbard, Olaf Wolkenhauer, Zlatko Trajanoski
[1]F.Sanchez-Cabo@postgrad.umist.ac.uk, UMIST, Manchester, UK; [2]claus.vogl@vu-wien.ac.at, Veterinaermedizinische Universitaet Wien, Vienna, Austria

This poster presents a Bayesian probabilistic model to cluster gene expression data from replicate experiments. Concurrently with the other parameters, the number of clusters is inferred and missing values are imputed. The method was applied to the transcriptome of cell-cycle synchronized yeast cells at different time points and of differentiating mouse adipocytes

F-5

Dynamic Model-Based Algorithm for Screening and Genotyping over 100K SNPs on Oligonucleotide Microarrays

Xiaojun Di[1], Hajime Matsuzaki, Teresa A. Webster, Guoying Liu, Shoulian Dong, Dan Bartell, Jing Huang, Richard Chiles, Geoffrey Yang, David Kulp, Giulia C. Kennedy, Rui Mei, Keith W. Jones, Earl Hubbell and Simon Cawley
[1]Xiaojun_di@affymetrix.com, Affymetrix, Inc.

A dynamic model-based algorithm for screening and genotyping over 100K SNPs on oligonucleotide Microarrays, it starts from a probe level dynamic model-based likelihood, ends with a SNP level statistical aggregation to elevate genotyping confidence, it is available in Affymetrix's Genotping Tools software package and soon in GDAS software package.

F-6

Using Order Statistics to Identify Functionally Relevant Gene Neighborhoods in Hematopoietic Stem Cell Regulation

Hongxian He[1], Natalia B. Ivanova[2], Gregory R. Grant, Jason A. Hackney, Lyle Ungar, Ihor R. Lemischka, Christian J. Stoeckert, Jr.
[1]hongxian@pcbi.upenn.edu, Univeristy of Pennsylvania; [2]nivanova@molbio.princeton.edu, Princeton University

We present an approach of using order statistics and coordinated differential expressions to identify functionally relevant genes and gene neighborhoods in hematopoietic stem cells regulation using microarray data. This provides us a list of genes and their interacting neighbors that are potentially regulatory components in hematopoietic stem cell differentiation.

F-7

Adjusting Graphical Models for Reverse Genetic Engineering

Anja Wille[1], Philip Zimmermann[2], Eva Vranova, Andreas Fuerholz, Oliver Laule, Stefan Bleuler, Lars Hennig, Amela Prelic, Peter von Rohr, Lothar Thiele, Eckart Zitzler, Wilhelm Gruissem, Peter Buehlmann
[1]awille@inf.ethz.ch, ETH Zurich; [2]philip.zimmermann@ipw.biol.ethz.ch, ETH Zurich

We present a modified graphical Gaussian modeling approach for reverse engineering of genetic regulatory networks with many genes and few observations. We show in a simulation study that our approach outperforms standard graphical Gaussian modeling. Our method is applied to infer a gene network for isoprenoid biosynthesis in Arabidopsis thaliana

F-8

Selecting Optimal Oligos Using Phylogenetic Information

Alexander Schliep[1], Jonas Heise[2], Sven Rahmann
[1]schliep@molgen.mpg.de, Max Planck Institute for Molecular Genetics; [2]heise@molgen.mpg.de, Max Planck Institute for Molecular Genetics

Microarrays can be used to establish presence or absence of biological agents in a sample. Based on our earlier work (Schliep et. al., 2003, Klau et. al. ISMB 2004) we propose a method for using phylogenetic information in selecting oligonucleotides, which helps to increase robustness and to find novel agents.

F-9

Infering Gene Function by SVM Analysis of Splicing-specific Microarray Data

Yael Mandel-Gutfreund[1], Manny Ares[2], Todd, A. Burckin, Roland J. Nagel, Grant. A Hartzog, William S. Noble
[1]yael@darwin.ucsc.edu, Technion; [2]ares@biology.ucsc.edu, UCSC

The effects of 87 different mutations in genes required for different steps in the yeast gene expression pathway have been examined using splicing specific Microarrays. To probe for gene function at different steps in the pathway, we have performed a machine learning phenotype analysis using Multi-Class Support Vector Machines.

F-10

Spotfinding for Nylon Arrays

Adele Cutler[1], Yi Xie[2], Bart Weimer
[1]adele@math.usu.edu, Utah State University; [2]yxie1@jhmi.edu, Johns Hopkins

We describe a method for detecting spots and extracting intensities for nylon arrays. The method is implemented as an ImageJ plugin. Special capabilities include fast automatic grid finding, detection of very weak spots, and handling saturation of strong spots due to limitations in the dynamic range of the detection device.

F-11

Gene Expression-based, Individualized Outcome Prediction for Surgically Treated Lung Cancer Patients

Shuta Tomida[1], Katsumi Koshikawa[2], Yasushi Yatabe, Tomoko Harano, Nobuhiko Ogura, Tetsuya Mitsudomi, Masato Some, Kiyoshi Yanagisawa, Toshitada Takahashi, Hirotaka Osada, and Takashi Takahashi
[1]stomida@aichi-cc.jp, Divisions of Molecular Oncology, Aichi Cancer Center Research Institute; [2]kkoshika@aichi-cc.jp, Divisions of Molecular Oncology, Aichi Cancer Center Research Institute

Individualized outcome prediction classifiers were successfully constructed based on expression profiling of 50 lung cancer cases. The resultant classifier yielded 82% accuracy for forecasting survival or death 5 years after surgery of a given patient. In addition, more than 90% accuracy was achieved by the histologic type-specific outcome classifiers.

F-12

Explaining Microarray Results Using Abductive Inference

Irene Papatheodorou[1], Antonis Kakas[2], Marek Sergot
[1]ivp@doc.ic.ac.uk, Department of Computing, Imperial College; [2]ack@doc.ic.ac.uk, Department of Computer Science, University of Cyprus

Microarrays can be used to identify gene interactions. We analyse such results by abduction, inference from effects to possible causes. We present an Abductive Logic Program, a simple model of how gene interactions cause changes in gene expression. The input are microarray data and output are possible gene interactions.

F-13

Microarray Analysis on Many Genes Determine a Cell Type

Wataru Fujibuchi[1], Paul Horton[2]
[1]fujibuchi-wataru@aist.go.jp, Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Japan; [2]horton-p@aist.go.jp, Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Japan

We define a cell's state as the rank order of its gene expressions. Using microarray data, we assessed the number of randomly selected genes necessary to identify traditional cell types with rand correlation. We found that 50 genes are enough to discriminate between normal cell types, but more are needed for cancer cells.

F-14

Environmental Sample Processor Detects Microorganisms Remotely, in Real-Time via Molecular Probes

Scholin, C., R.[1], B. Roman[2], Marin III, S. Jensen, J. Feldman, E. Massion
[1]scholin@mbari.org, MBARI; [2]brent@mbari.org, MBARI

The Environmental Sample Processor (ESP) is an autonomous, robotic laboratory for long term, in situ application of molecular probes. It collects discrete water samples, concentrates microorganisms and archives them for later microscopic and toxin analyses while identifying and quantifying species - radioing results in real-time for processing and interpretation.

F-15

Self-repairing System for Biological Knowledge Databases in Estimating Gene Networks with Expression Data

Seiya Imoto[1], Tomoyuki Higuchi[2], Takao Goto and Satoru Miyano
[1]imoto@ims.u-tokyo.ac.jp, University of Tokyo; [2]higuchi@ism.ac.jp, Institute of Statistical Mathematics

We propose a statistical joint learning model for repairing database information and simultaneously for estimating a gene network based on Bayesian networks and gene expression data. We show the high performance of the proposed method through the analysis of the yeast cell cycle data.

F-16

ToMAS: Software Development Toolkits for Microarray Analysis System

Bong-Kyung Chun[1], Pyung-Jun Lee[2], Hee-Jeong Jin, Myung-Jae Jun, Ji-Hyun Yoon, Chul-Jin Jang, Kyung-Sin Lee, Hye-Jung Kim, Hwan-Gue Cho
[1]bkchun@pearl.cs.pusan.ac.kr, Pusan National University ; [2]pjlee@pearl.cs.pusan.ac.kr, Pusan National University

ToMAS is an integrated toolkit(component softwares) for the development of microarray analysis software. It consists of four major components, which are ManagerCOM, ArrayCOM, NormalCOM and ClusterCOM. ToMAS allows us to develop a new microarray analysis system more easily and rapidly. Information about ToMAS is available at http://garnet.cs.pusan.ac.kr/~tomas.

F-17

Principal Component Analysis and the GE-biplot for Exploration of Microarray Data

Susan R Wilson[1], Yvonne Pittelkow[2]
[1]sue.wilson@anu.edu.au, anu; [2]Yvonne.Pittelkow@anu.edu.au, anu

PCA is widely used for statistical analysis of microarray data; note its usefulness depends on initial appropriate standardisation. Another ordination method, the GeneExpression(GE)-biplot, is proving to be extremely useful. It displays genes and chips simultaneously in a way that can be readily interpreted; basic details and examples are in www.bepress.com/sagmb/vol2/iss1/art6/

F-18

Genome Annotation of P.falciparum using Microarray Data

K.M.Simpson[1], T.P.Speed[2]
[1]ksimpson@wehi.edu.au, The Walter and Eliza Hall Institute; [2]terry@stat.berkeley.edu, The Walter and Eliza Hall Institute and U.C. Berkeley

We describe a technique for Genome Annotation using microarray data and sequence analysis tools. The technique is illustrated using custom Affymetrix chips designed for the study of P.falciparum

F-19

Mixture Distribution Modeling of Noise Using the Duplicated Data in the Microarray

Masaru Takeya[1], Takehiro Matsuda[2], Masao Iwamoto, Toshiya Nakaguchi, Norimichi Tsumura and Yoichi Miyake
[1]katu@affrc.go.jp, National Institute of Agrobiological Sciences; [2]matsuda-t@graduate.chiba-u.jp, Chiba University

Plot distribution of duplicated data in the microarray was represented as a mixture distribution modeling of multiple 2-dimensional noise distributions. A relative amount of gene expression was indicated from the mixture distribution. This method was applied to extract a candidate for the cluster element from outer genes of the cluster.

F-20

MicroArray On Line Analysis System (MOLAS): A Web Platform for Array Data Management and Analysis

Wei-Chen Chen[1], Chung-Yen Lin[2], Yueng-Shiang Huang, I-Shou Chang, and Chao A. Hsiung
[1]snoweye@nhri.org.tw, National Health Research Institutes; [2]cylin@nhri.org.tw, National Health Research Institutes

MOLAS is a web-based customizable bioinformatics package designed for manager and analyze massive array data. MOLAS can track the flow of array data, provide the report of quality control, and various tools for data analysis such as statistical filters, novel normalization methods, visualization tools, clustering methods and simple biological interpretation.

F-21

Prediction Model for Pathway Extension Using Microarray Expression Profiles

Tae Su Chung[1], Keewon Kim[2], Juhan Kim
[1]epiai@snu.ac.kr, Human Genome Research Institute; [2]keien1@snu.ac.kr, SNUBI

Biological pathways are collections of known relations or reactions between biological objects i.e. genes or gene products. We proposed a scoring system, introduced by the social network analysis, to extend an incomplete pathway structures.

F-22

Singular Value Decomposition - Based Algorithm for Gene Expression Data Analysis

Jeerayut Chaijaruwanich[1], Asawin Meechai[2], Supapon Cheevadhanarak, Sakarindr Bhumiratana
[1]jeerayut@cs.science.cmu.ac.th, Chiang Mai University; [2]asawin.mee@kmutt.ac.th, King Mongkut's University of Technology Thonburi

We propose a novel method to reorder genes according to their similar expression behaviors under different conditions. Our method, based on SVD, uses the eigenvalue-eigenvector information of the given microarray data. It allows us to better visualize microarray data and identify the most expressed-repressed genes during the time course experiments.

F-23

"Copasetic Analysis": A Framework for the Blind Analysis of Microarrqay Imagery

Paul O'Neill[1], Karl Fraser[2], Zidong Wang and Xiaohui Liu
[1]paul@ida-research.net, Brunel Univeristy; [2]karl@ida-research.net, Brunel Univeristy

Microarray image analysis techniques have focused on the direct analysis of raw image data, failing to utilise the image's full potential, resulting in lab technicians having to guide the analysis algorithms. We present a dynamic framework to automate the process of microarray image analysis using a variety of techniques.

F-24

RMAGEML: Integrating MAGE-ML Format Microarray Data and Bioconductor

Steffen Durinck[1], Joke Allemeersch[2], Vincent J. Carey, Yves Moreau, Bart De Moor
[1]steffen.durinck@esat.kuleuven.ac.be, ESAT-SCD, KULeuven; [2]joke.allemeersch@esat.kuleuven.ac.be, ESAT-SCD, KULeuven

MAGE-ML is an XML standard for describing and exchanging information about microarray experiments. The Bioconductor project provides a framework for statistical analysis of genomic data in R. We describe RMAGEML, a new Bioconductor package that integrates MAGE-ML format data exchange and Bioconductor. RMAGEML is available at http://www.bioconductor.org

F-25

FACT: Explorative Data Analysis of High-throughput Experiments

Felix Kokocinski[1], Nicolas Delhomme[2], Gunnar Wrobel, Peter Lichter
[1]f.kokocinski@dkfz.de, DKFZ; [2]n.delhomme@dkfz.de, DKFZ

FACT (Flexible Annotation and Correlation Tool) was developed as a flexible framework for the explorative meta-analysis of genomic, proteomic or other experiments. Heterogeneous experimental data and annotational sources can be integrated and diverse analytical algorithms can be applied with the goal of finding distinct patterns and illuminating inherent characteristics.

F-26

GEDA-C: Gene Expression Data Analyzer for Clustering

Miyoung Shin[1], Eunmi Kang[2], Seon-Hee Park
[1]shinmy@etri.re.kr, ETRI; [2]emkang@etri.re.kr, ETRI

GEDA-C was developed as a clustering tool for gene expression data, especially incorporating various types of prior-knowledge regarding genes as seed information, which is called seed clustering. Currently, three types of prior-knowledge are allowed to use as seed information, which include previously obtained clustering results, functional annotations, and gene profiles of user's interest.

F-27

Development, Management and Analysis of Human MitoChip

Lenka Piherova[1], Robert Ivanek[2], Alena Cizkova, Petr Divina, Jan Paces, Stan Kmoch
[1]lenka.piherova@lf1.cuni.cz, Institute of Inherited Metabolic Disorders, 1st Faculty of Medicine, Charles University; [2]ivanek@biomed.cas.cz, Dept. Mammalian Molecular Genetic, Institute of Molecular Genetics AS CR

The Human MitoChip is an oligonucleotide microarray, which interrogate ultimately expression of around 800 human genes involved in mitochondria biogenesis, maintenance, and metabolism. We will present the management and the analysis of data obtained from our MitoChip as well as analysis of public data related to our topic.

F-28

Finding Splicing Variants with Gene Expression Data

Tadashi Kadowaki[1], Gozoh Tsujimoto[2], Satoshi Shiojima, Yasushi Okuno
[1]kadowaki@pharm.kyoto-u.ac.jp, Kyoto University; [2]gtsuji@pharm.kyoto-u.ac.jp, Kyoto University

The method of finding splicing variants was developed. We found some disease related splicing variants and non-characterized variants by analyzing our data obtained from disease samples and normal samples. The results shows that our approach works and the large scale experiments are needed to reveal the roles and mechanisms of the spliceome system.

F-29

cMAMS: cDNA Microarray Data Analysis and Management System

Sang-Bae Kim[1], Young-Jin Kim[2], Hyo-Mi Kim, Ho-Youl Jung, Eun-Jung Lee, Jung-Sun Park, Yun-Ju Park, Ku-Chan Kim, In-Song Koh
[1]ksb003@korea.com, National Genome Research Institute; [2]inthistime@lycos.co.kr, National Genome Research Institute

cMAMS is a web-based pipeline process system. It provides the integrated multi-functions for magnifying the efficiency of analysis program capability from the data deposit and handling to the data analysis including gene annotation. It was implemented on Linux platform by JSP, Perl, and R languages. cMAMS is being improved, and the draft version is available at http://www.nih.go.kr/

F-30

ExpressMiner: A Tool for Mining the Gene Expression Data from Microarray Experiments

Sunitha Jonnakuty[1], Coral del Val[2], Karl-Heinz Glatting, Sándor Suhai
[1]S.jonnakuty@dkfz.de, Department of Molecular Biophysics, Deutsches Krebsforschungszentrum(DKFZ),69120 Heidelberg, Germany;
[2]C.delval@dkfz.de, Department of Molecular Biophysics, Deutsches Krebsforschungszentrum(DKFZ),69120 Heidelberg, Germany

ExpressMiner is a tool designed to identify, annotate and compare large collections of up- or down-regulated genes/clones from microarray experiments. ExpressMiner groups information and simultaneously visualize all genes by constructing functional profiles based on tissue-specificity, biochemical-pathway information , GO-slims and user-specified sorting strategies after reducing inconsistency by cross-checking database information.

F-31

Using TFBS Information to Bias Genetic Network Estimation via Bayesian Network Learning

Mathaeus Dejori[1], Andreas Naegele[2], Martin Stetter
[1]mathaeus.dejori.external@mchp.siemens.de, Siemens AG CT IC4; [2]naegele@in.tum.de, Dep. of Computer Science, Technical University of Munich

We combine sequence and microarray data analysis to infer gene regulatory networks. For this, TFBS information is used for gene-preselection and as a Bayesian Prior to guide Bayesian structure learning from microarray data. By use of this ability the resulting network better describes the underlying regulatory system.

F-32

An Integrated Universal Binning and Classification Feature Selection Algorithm to Extract Minimal Number of Variables

Liang Goh[1], Nikola Kasabov[2]
[1]liang.goh@aut.ac.nz, KEDRI; [2]nkasabov@aut.ac.nz, KEDRI

We propose a new integrated universal feature selection algorithm that not only selects significant variables, but also selects only those variables that contribute to the classification or the prediction rate of the model. This results in a minimal set of features that gives the maximum classification rate for the model.

F-33

Integration of the MAGE-OM and CDA

Yu Rang Park[1], Ju Han Kim[2], Ji Yeon Park, Hwa Jeung Seo
[1]kirarang@snu.ac.kr, Seoul Nat'l University School of Medicine; [2]juhan@snu.ac.kr, Seoul Nat'l University School of Medicine

To assess the relevance of microarray analysis it is required to combine clinical and genetic information. However current standard model for microarray (MAGE-OM) cannot include modeling for the clinical data of source. So we propose the method that integration of MAGE-OM and CDA.

F-34

arrayMagic: Two-colour DNA Array Quality Control and Preprocessing

Andreas Buneß[1], Wolfgang Huber[2], Klaus Steiner, Holger Sültmann, Annemarie Poustka
[1]a.buness@dkfz.de, German Cancer Research Institute; [2]w.huber@dkfz.de, German Cancer Research Institute

arrayMagic is a software package which facilitates the analysis of two colour DNA microarray data. The package is written in R (http://www.r-project.org) and integrates into Bioconductor (http://www.bioconductor.org). The automated analysis pipeline comprises data loading, normalisation and quality diagnostics. The pipeline is flexbile and can be adjusted for specific needs.

F-35

Functional Food Ingredients Against Colorectal Cancer: the Bioinformatics Challenge

Marjan van Erk[1], Cyrille Krul[2], Eric Caldenhoven, Rob Stierum, Ruud Woutersen, Ben van Ommen
[1]vanerk@voeding.tno.nl, TNO Nutrition and Food Research; [2]krul@voeding.tno.nl, TNO Nutrition and Food Research

Transcriptomics technologies were used to study effects and mechanisms of food compounds with possible cancer preventive action in cultured colon cancer cells. Since colon cancer prevention is multifactorial, multivariate statistical analysis tools like principal component analysis were combined with other bioinformatics tools like pathway analysis.

F-36

Molecular Profiling of Clinical Features in Breast Cancer Using Principal Component Analysis

Miryung Han[1], Mihyeon Kim[2], Wonshik Han, Hyejin Kim, Hye Won Lee, Dong-Young Noh, Ju Han Kim
[1]gene0309@snu.ac.kr, Seoul National University Biomedical Informatics (SNUBI);
[2]aroo12@snu.ac.kr, Seoul National University Biomedical Informatics (SNUBI)

To identify important biological aspects of malignancy, we analyzed cDNA microarray data from breast cancers using Principal Component Analysis (PCA). We correlated resulting PCs with clinical data and revealed the difference between ER(+) patient with recurrence and ER(+) patient without recurrence by comparing genes with extreme (negative or positive) loading values.

F-37

A Novel Statistical Framework for the Design of Microarray Experiments and Effective Detection of Differential Gene Expression

Shu-Dong Zhang[1], Timothy W. Gant[2]
[1]sdz1@le.ac.uk, MRC Toxicology Unit, University of Leicester; [2]twg1@le.ac.uk, MRC Toxicology Unit, University of Leicester

A novel statistical framework is presented for the detection of differential gene expression (DGE) in microarray experiments. We have derived a formula to determine the success rate of DGE detection, which can be routinely used in the design of experiments or in post-experiment assessment.

F-38

Characterize Each Type of Relations in Biological Pathways with Microarray Data

Mingoo Kim[1], Ji Yeon Park[2], Hyejin Kim, Ju Han Kim
[1]satyrs1@snu.ac.kr, SNUBI; [2]ys802@snu.ac.kr, SNUBI

Gene expression data from the microarray show weaknesses in deriving specific relations among genes. By surveying expression patterns of well-known relations in biological pathways, we try to associate expression patterns with the specific biological relations.

F-39

Mixed Models for Probe-level Oligonucleotide Array Data Analysis in a Lymphocyte Development Study

Dick de Ridder[1], Karin Pike-Overzet[2], Floor Weerkamp, Menno C. van Zelm, Marcel J.T. Reinders and Frank J.T. Staal
[1]D.deRidder@ewi.tudelft.nl, Information & Communication Theory Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology;
[2]k.pike-overzet@erasmusmc.nl, Department of Immunology, Erasmus University Medical Center

We study stages in human lymphocyte development using Affymetrix Microarrays. Due to the large number of stages and difficulties in obtaining sufficient material for some stages, only a small number of Microarrays are used per stage. We discuss algorithms employed for (differential) expression calculation, based on RMA and two-way ANOVA.

F-40

Microarray Experiments: New Statistical Tools Facilitate Biological Interpretation

Rainer Breitling[1], Anna Amtmann[2], Pawel Herzyk
[1]R.Breitling@bio.gla.ac.uk, University of Glasgow;
[2]A.Amtmann@bio.gla.ac.uk, University of Glasgow

We present a number of statistical tools for the automated interpretation of microarray results that provide a rapid biological summary of the most relevant expression changes. The flexible combination of expression data and previous knowledge allows a more sensitive detection and helps to guide further exploration of the results.

F-41

ArrayExpress: A Standards-based Public Repository for Microarray Data

Tim Rayner[1], Alvis Brazma[2], Niran Abeygunawardena, Sergio Contrino, Anna Farne, Gonzalo Garcia Lara, Ele Holloway, Misha Kapushesky, Gaurab Mukherjee, Ahmet Oezcimen, Helen Parkinson, Philippe Rocca-Serra, Susanna-Assunta Sansone, Ugis Sarkans, Anjan Sharma, Mohammadreza Shojatalab
[1]rayner@ebi.ac.uk, European Bioinformatics Institute;
[2]brazma@cam.ac.uk, European Bioinformatics Institute

ArrayExpress is a public repository of well-annotated microarray data representing over 5000 hybridizations from 177 experiments performed in 15 different organisms. The database provides submission and query interfaces, expert curation, and analysis tools, and adheres to community annotation standards. ArrayExpress is available at http://www.ebi.ac.uk/arrayexpress/

F-42

Predicting Transcript Isforms from EST Data for cDNA Microarray Design and Quality Control

Beaudoing[1], Gautheret[2], Hingamp
[1]beaudoing@tagc.univ-mrs.fr, Inserm erm206; [2]hingamp@tagc.univ-mrs.fr, Inserm erm206

We propose a method to locate known cDNA clones in order to assist a priori cDNA microarray design as well as a posteriori quality control of existing cDNA designs flagging reporters liable to above pitfalls

F-43

Detection of State Transitions Based on Gene Expression in a Whole Cell

Ryoko Morioka[1], Shigehiko Kanaya[2], Keiko Matsuda, Tetsuo Sato, Kazuo Kobayashi, Naotake Ogasawara, Kotaro Minato
[1]ryouko-m@is.naist.jp, Nara Institute of Science and Technology; [2]skanaya@gtc.naist.jp, Nara Institute of Science and Technology

Living organisms change their own states in order to adapt to the environments. We estimate the transition time points extracted from expression profiles of Bacillus subtilis based on a linear dynamical model and SOM. Significant transition points can be obtained, and the results are also interpreted by metabolome of Bacillus subtilis.

F-44

Network Analysis and Data Processing System for Gene Expression Time-series Data

Tominaga Daisuke[1], Takahashi Katsutoshi[2], Kadota Koji, Yoshikawa Tomohiro, Miyake Masato
[1]tominaga@cbrc.jp, Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology; [2]takahashi-k@aist.go.jp, Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology

We develop a system for quantitative time-series data of gene expression levels which obtained by recent technology, such as transfection arrays. Elimination of outlier data and polynomial fitting with statistical criterion are done as data processing, then estimation of gene clusters relations among clusters, and changes of clustering are performed.

F-45

Search for Regulatory Motifs in Drosophila Melanogaster Genome

Anastasia Samsonova[1], Christoph Dieterich[2], Martin Vingron and Alvis Brazma
[1]nastja@ebi.ac.uk, European Bioinformatics Institute; [2]dieteric@molgen.mpg.de, Max-Planck-Institute for Molecular Genetics

With published genomic sequences of D.melanogaster and D.pseudoobscura, we identified conserved regions on the genome that contain putative transcription factor binding sites. For the clusters of co-expressed genes found via microarray data analysis, we ran searches for statistically over-represented motifs in the corresponding groups of non-coding conserved sequences.

F-46

Insight into Pax-5 Gene Expression Program from Targeted DT40 B Line

Pekka Kohonen[1], Kalle-Pekka Nera[2], Elli Veistinen, Anne Peippo, Perttu Terho, Jean-Marie Buerstedde and Olli Lassila
[1]pekka.kohonen@utu.fi, University of Turku; [2]kalle-pekka.nera@utu.fi, University of Turku

The Pax-5 transcription factor is central to B cell commitment and development. We have used knockouts in the avian DT40 cells to study Pax-5 targets with array containing 14000 probes derived from B cells. Target genes yield insights into B cell development, signaling and somatic hypermutation or gene conversion.

F-47

A Quality Measure Model and System for Microarray Images

Pan-Gyu Kim[1], Kiejung Park[2], Hwan-Gue Cho
[1]pgkim@smallsoft.co.kr, Information Technology Institute, SmallSoft Co., Ltd., Junmin-Dong 461-71, Yusung-Gu, Daejeon 305-811, Korea, Department of Computer Science and Engineering, Pusan National University, Pusan 609-735, Korea; [2]kjpark@smallsoft.co.kr, Information Technology Institute, SmallSoft Co., Ltd., Junmin-Dong 461-71, Yusung-Gu, Daejeon 305-811, Korea

We developed a microarray image analysis system which provides five quality measure functions of signal noise, background noise, scale invariant, spot regularity, and spot alignment, to check the quality of microarray images and validate the correctness of experiments. Each quality measure expresses and quantifies the property of good or bad microarray quality.

F-48

ArrayPipe: a Powerful Processing Pipeline for Microarray Data with User-friendly Automation and Parallelization

Karsten Hokamp[1], Fiona M. Roche[2], Michael Acab, Marc-Etienne Rousseau, Byron Kuo, David Goode, Dana Aeschliman, Jenny Bryan, Lorne A. Babiuk, Robert E.W. Hancock, Fiona S.L. Brinkman
[1]karsten_hokamp@sfu.ca, Simon Fraser University; [2]fiona_roche@sfu.ca, Simon Fraser University

ArrayPipe is a multi-functional tool for the analysis of dual-labeled microarray data. It can be run as a web server or through the command line and allows very flexible construction of complex analysis processes. Open source ArrayPipe is freely available at www.pathogenomics.ca/arraypipe

F-49

Automatic Pipeline for Standardized Analysis of Gene Expression Profiling Experiments at RZPD

Karsten R. Heidtke[1], Rico Basekow[2], Robert Wagner, Bernd Drescher
[1]heidtke@rzpd.de, German Resource Center for Genome Research GmbH, Berlin, Germany; [2]basekow@rzpd.de, German Resource Center for Genome Research GmbH, Berlin, Germany

Standardizing the analysis of gene expression profiling experiments is the only way to make them comparable to each other and to control the quality of the results. The Resource Center for Genome Research (RZPD) is establishing an automatic pipeline for standardized analysis including statistical methods, functional annotation and available clones.

F-50

Graph Algorithms on Signal Transduction Network to Decipher Gene Expression: Introducing Indirect Edges and Chains

Nico Voss[1], Alexander Kel[2], Anatolij P. Potapov, Claudia Choi, Susanne Pistor, Mathias Krull, Olga Kel-Margoulis and Edgar Wingender
[1]nvo@biobase.de, BIOBASE GmbH; [2]ake@biobase.de, BIOBASE GmbH

TRANSPATH is a manually curated signal-transduction database with associated analysis and visualization tools. It enables efficient analysis of gene expression mass data, prediction of possible common upstream regulators and visualization of a detailed, dynamically predicted pathway. These appoaches could be further improved with respect to false positive predictions by introducing indirect edges and chains into the signaling network.

F-51

RNA Expression as Quantitative Trait in Animal Models of Multiple Sclerosis

Steffen Möller[1], Patrik Wernhoff[2], Pablo Serrano, Uwe K. Zettl, Hans-Jürgen Thiesen, Dirk Koczan, Saleh M. Ibrahim
[1]moeller@pzr.uni-rostock.de, University of Rostock, Institute of Immunology; [2]patrik.wernhoff@inflam.lu.se, University of Rostock, Institute of Immnunology

Applied to murine experimental autoimmune encephalomyelitis, this work extends the clinical phenotypes of genotyped mice to the DNA chip expression data taken from the same individual. Taken as quantitative trait, an analysis yields chromosomal loci (and their pairwise interaction) associated with the control of the expression of each gene.

F-52

Weighted Analysis of Microarry Gene Expression Using Maximum Likelihood

David Bakewell[1], Ernst Wit[2]
[1]d.bakewell@beatson.gla.ac.uk, Glasgow University; [2]ernst@stats.gla.ac.uk, Glasgow University

This poster describes the development of Maximum Likelihood Estimation based algorithms that assign weights to microarray spot mean values based on pixel variance. Using data from experiments and simulations, comparisons with current methods that exclude pixel variance shows gene expression estimates using weighted-spot means leads to improvements in differential detection.

F-53

Construction of Predictive Models for Patient's Survival Time Based on a Combination of Gene Expression Data and Text Mining Networks

Christian Gieger[1], Hartwig Deneke[2], Juliane Fluck
[1]christian.gieger@scai.fhg.de, Fraunhofer Institute SCAI; [2]hartwig.deneke@scai.fhg.de, Fraunhofer Institute SCAI

We present an approach which combines gene expression data and text mining networks for predicting a patient's survival time using a Cox proportional hazard model. By incorporating a priori biological knowledge through protein interaction networks extracted by textmining of Medline abstracts, such a model includes pathway information, which facilitates the biological interpretation.

F-54

Combining Databases, Microarrays, and Genetics to Identify Novel Imprinted Genes

Jonathan D. Choi[1], Andrew Wood[2], Lara Underkoffler, Joelle Collins, Rebecca Oakey
[1]jonathan.choi@kcl.ac.uk, Dept. of Med. and Mol. Genetics, Kings College London; [2]andrew.wood@genetics.kcl.ac.uk, Dept. of Med. and Mol. Genetics, Kings College London

Genomic Imprinting is an epigenetic phenomena in which genes are preferentially expressed from one allele in a parent-of-origin specific fashion. We are combining Microarrays, Databases, and genomic feature analysis with traditional genetics to identify novel imprinted genes in the mouse. Several novel imprinted genes have been found that define two imprinted clusters.

F-55

Microarray Data Integration

Kjell Petersen[1]
[1]kjellp@cbu.uib.no, Bergen Center for Computational Science, Computational Biology Unit, University of Bergen

New genome scale experiments make grounds for new important scientific discoveries, but also raise the need for standard representation and documentation. Discussion of the MAGE standard for microarray data, and examples of implementation using the MAGE software toolkit for integration of resources such as J-Express (analysis tool), ArrayExpress (public repository@EBI), and BASE (lab database) is presented.

F-56

Comparison of Gene Selection Methods Used in Microarrays Analysis

Ian Jeffery[1], Aedin Culhane[2], Linda McArdle, Des Higgin, Ray Stallings
[1]ian.jeffery@ucd.ie; [2]aedin.culhane@ucd.ie

We preformed a statistical analysis on microarray data from 22 Neuroblastoma samples. This poster will illustrate the selectivity and sensitivity of MaxT, Significance Analysis of Microarray data (SAM), and Between Group Analysis (BGA) using the results obtained in this study, as well as results from other methods of gene selection, using this data.

F-57

GenomeCubeExpress - a Public Repository for Microarray Data

Robert Wagner[1], Bernd Drescher[2]
[1]wagner_r@rzpd.de, Resource Center for Genome Research GmbH; [2]drescher@rzpd.de, Resource Center for Genome Research GmbH

GenomeCubeExpress is a public repository for microarray experiments, that stores data in accordance with the MGED recommendations. Experiments can be queried by certain criteria and examined using an automated analysis pipeline. Data submitted will also be pipelined to ArrayExpress, the international microarray database, where it will get a stable accession number.

F-58

Alteration of the Gene Coexpression Network in Cancer

Jung Kyoon Choi[1], Sangsoo Kim[2], Ook Joon Yoo
[1]jkchoi@kaist.edu, KAIST; [2], NGIC

We investigated the gene coexpression network created from multiple gene expression profiles on tumors of diverse tissue types. Tissue-wide coexpression changes in human cancer were analyzed and discussed via the comparison of a normal network and a tumor network.

F-59

A Generalised Method for Motif Identification from ChIP Array Data

David Huen[1]
[1]mh1008@cus.cam.ac.uk, Dept. of Genetics, University of Cambridge

I have developed a novel method for the identification of binding sites from ChIP array data. I will present results from a pilot study applying this approach to the analysis of results obtained within the FlyChip facility.

F-60

DBRF-MEGN Method: an Algorithm for Deducing Minimum Equivalent Gene Networks from Large-scale Gene Expression Profiles of Gene Deletion Mutants

Shuichi Onami[1], Koji Kyoda[2]
[1]sonami@bio.keio.ac.jp, Graduate School of Science and Technology, Keio University, and Institute for Bioinformatics Research and Development, Japan Science and Technology Agency, Yokohama, Japan; [2]kyoda@so.bio.keio.ac.jp, Graduate School of Science and Technology, Keio University, Yokohama, Japan

Signed directed graph is the most common representation of gene network in genetics and cell biology. We developed an algorithm for deducing the most parsimonious signed directed graph consistent with the expression profiles of gene deletion mutants. The algorithm was applied to large-scale gene expression data of S. cerevisiae.

F-61

Subspace Clustering for Microarray Data Analysis: Multiple Criteria and Significance Assessment

Hui Fang[1], Chengxiang Zhai[2]
[1]hfang@cs.uiuc.edu, University of Illinois at Urbana-Champaign; [2]czhai@cs.uiuc.edu, University of Illinois at Urbana-Champaign

We report a new clustering method for analyzing microarray data that improves existing approaches in three aspects — capturing gene similarities under a subset of gene expression conditions, combining multiple criteria to capture trend similarity, and assigning statistical significance to detected clusters.

F-62

Analysis and Visualization of Microarray Data for Comparative Genomi Hybridization Experiments

Wendy A. Findlay[1], Eduardo N. Taboada[2], Simon J. Foote, John H.E. Nash
[1]Wendy.Findlay@nrc-cnrc.gc.ca, NRC; [2]Ed.Taboada@nrc-cnrc.gc.ca, NRC

A full genome microarray of Campylobacter jejuni (cause of acute bacterial enteritis) was used to characterize the genetic diversity of 52 strains. We developed java tools to statistically analyse and visualize genome-wide gene conservation patterns, and identified 120 highly variable genes, a potential basis for future "genotyping" of new strains.

F-63

A Computational Linguistics Approach to Representation and Reasoning of DNA Microarray Experiments - With Applications in Medicinal Plant and Formulation Authentication

Siu-wai Leung[1], Yanbo Zhang[2], Dave Robertson
[1]siu@inf.ed.ac.uk, School of Informatics, The University of Edinburgh; [2]ybzhang@cuhk.edu.hk, Department of Biochemistry, The Chinese University of Hong Kong

A DNA linguistic formalism Basic Gene Grammars (BGGs) was developed to represent DNA microarray experiments, particularly the relationship of target sequences and their specific probes. Possible conceptual experiment designs can be generated based on a grammar and the normal practice (preferences and constraints) of the laboratory for medicinal plant and formulation authentication.

F-64

Splice Variant-based Expression Profiling Using High-density Oligonucleotide Microarrays

Hui Wang[1], Alan Williams[2], Jing-shan Hu, Gang Lu, Tyson Clark, Melissa Cline, Ray Wheeler, Gangwu Mei, John Burke2 and John Blume1
[1]hui_wang@affymetrix.com, Affymetrix; [2]alan_williams@affymetrix.com, affymetrix

Alternative splicing is an essential, biological process, involving an estimated ~60% of all human genes. We have designed arrays to access splice variation. Here, we have described our array design concepts, analysis algorithms and their biological applications.

F-65

A Knowledge-Based Clustering Algorithm Driven by Gene Ontology

Jill Cheng[1], Melissa Cline[2], John Martin, David Finkelstein, Tarif Awad, David Kulp, and Michael A. Siani-Rose
[1]jill_cheng@affymetrix.com, Affymetrix, Inc.; [2]melissa_cline@affymetrix.com, Affymetrix, Inc.

We have developed an algorithm for inferring the degree of similarity between genes using the graph structure of Gene Ontology. We apply this knowledge-based similarity metric to identify related genes; we also combine it with an expression-based distance metric in a co-cluster analysis to identify genes that are similar in both expression profiles and biological characteristics.

F-66

The ChipYard Framework For Array Based Data Analysis

Toedt, Grischa[1], Engel, Felix[2], Del Homme, Nicolas; Kokocinski, Felix
[1]g.toedt@dkfz.de, Deutsches Krebsforschungszentrum;
[2]f.engel@dkfz.de, Deutsches Krebsforschungszentrum

One of the major tasks in DNA-chip based experiments is to handle the large amount of data generated. The ChipYard Framework addresses this by providing a solution for tracking, analysis and visualization of microarray data (i.e. expression profiling data, matrix CGH data).

F-67

Microarray Data Analysis and Pathway Activity Inference in PATIKA

Ozgun Babur[1], Emek Demir[2], Asli Ayaz, Ugur Dogrusoz, Onur Sakarya
[1]ozgun@cs.bilkent.edu.tr, Bilkent University;
[2]emek@cs.bilkent.edu.tr, BIlkent University

Pathway activity inference attempts to infer differentially active pathways of cellular networks, given a qualitative state - transition model of the cellular network and an expression profile of RNA molecules. We present a new efficient algorithm, implemented as part of microarray data analysis component of PATIKA.

F-68

Variability Sources in Gene Expression Data

Johanna Hardin[1]
[1]jo.hardin@pomona.edu, Pomona College

The distribution of microarray data is often unknown. Transformed microarray data can have heavier tails than a normal distribution. We have provided a method of fitting a t-distribution to a data set using a robust estimate of scale. We investigate the source of the variability for genes.

F-69

Patterns of Housekeeping Genes in Cancer

Wooi Keng Tan[1], David K. Smith[2]
[1]wooikeng@hkusua.hku.hk, Department of Biochemistry, University of Hong Kong; [2]dsmith@hku.hk, Department of Biochemistry, University of Hong Kong

Housekeeping genes are always expressed to maintain the basic functions of a normal cell. We compare the expression of housekeeping genes in cancer cells and normal tissues using oligonucleotide microarray data and suggest the use of housekeeping genes as a reference standard to validate the quality of microarray experiments.

F-70

MPP: A Microarray-to-phylogeny Pipeline for Gene Content Analysis

Robert Davey[1], Ian Roberts[2], Jo Dicks, V.J. Rayward Smith
[1]robert.davey@bbsrc.ac.uk, John Innes Centre;
[2]ian.roberts@bbsrc.ac.uk, Institute of Food Research

MPP is a microarray analysis tool that constructs a "microarray-to-phylogeny pipeline" - a pathway to generate phylogenetic trees directly from DNA:DNA microarray data. We focus on the modeling of microarray data to infer gene content and we use the resultant data sets to discover key features of genome evolution.

F-71

ProbeLynx: A Tool for Updating the Association of Microarray Probes to Genes

Fiona Roche[1], Karsten Hokamp[2], Michael Acab, Lorne A. Babiuk, Robert E.W. Hancock, Fiona S.L. Brinkman
[1]froche@sfu.ca, Simon Fraser University;
[2]karsten_hokamp@sfu.ca, Simon Fraser University

ProbeLynx is a web-based tool which uses the latest genomic sequence and gene prediction data to re-examine microarray probe specificity and provide updated annotations on identified targets. ProbeLynx is freely available (including source code) at http://www.pathogenomics.ca/probelynx.

F-72

Large-scale Analysis of Coexpression of Human Genes

Paul Pavlidis[1], Homin K. Lee[2], Amy Hsu, Jon Sajdak, Jie Qin
[1]pp175@columbia.edu, Columbia University;
[2]hkl7@columbia.edu, Columbia University

We describe the analysis of coexpression patterns in a database of 60 human studies (3924 Microarrays). Using a simple meta-analysis approach, we show that confirmation of coexpression in multiple data sets is correlated with functional relatedness, and provide a basis for further analysis of high-confidence coexpression patterns.

F-73

Microarray-based Study of Craniofacial Development in Transgenic Mice

Ying JIANG[1], D.K.Smith[2], Simmonds, R.E., Zhou, Z., Wong, E.Y.M., Cheah, K.S.E., Sham, M.H.
[1]yyjiang@hkusua.hku.hk, Laboratory of Bioinformatics, Department of Biochemistry, Faculty of Medicine, the University of Hong Kong; [2]dsmith@hku.hk, Laboratory of Bioinformatics,Department of Biochemistry, Faculty of Medicine, the University of Hong Kong

A transgenic mouse ectopically expressing HOXB3 was found to have craniofacial abnormalities. We have conducted microarray studies of gene expression in the branchial arches of transgenic and wild-type embryonic mice to study craniofacial development. The data have been analyzed using several techniques to detect genes involved in craniofacial development.

F-74

Biologically Inspired Neural Networks and Genetic Algorithms for Microarray Data classification and Identification of Informative Genes

Anastasios Oulas[1], Panayiota Poirazi[2]
*[1]oulas@imbb.forth.gr, Foundation of Research and Technology, Hellas (FO.R.T.H) - Institute of Molecular Biology and Biotechnology (IMBB);
[2]poirazi@imbb.forth.gr, Foundation of Research and Technology, Hellas (FO.R.T.H) - Institute of Molecular Biology and Biotechnology (IMBB)*

A biologically inspired artificial neural network has been developed and used for microarray data classification. The method outperforms algorithms used in reference publications. Sophisticated clustering and genetic algorithms are currently used to identify informative groups of genes towards a better understanding of the molecular basis of the diseases under study.

F-75

Genome Diversity in Lactobacillus Plantarum

Douwe Molenaar[1], Roland Siezen[2], Frank Schuren, Francoise Bringel, Willem de Vos and Michiel Kleerebezem
[1]douwe.molenaar@nizo.nl, Wageningen Centre for Food Sciences; [2]siezen@cmbi.kun.nl, Wageningen Centre for Food Sciences

A comparative DNA microarray-based genotyping analysis of 20 strains of Lactobacillus plantarum showed considerable variation in the presence/ absence of different DNA regions. Several highly variable regions correlated with a high base-deviation index, suggesting horizontal transfer.
A 200-kb region encoding sugar metabolism displays genome plasticity and represents a life-style adaptation island.

F-76

MAGIC Microarray: a Complement to MAGIC Gene Discovery

Marie-Michele Cordonnier-Pratt[1], Lee Pratt[2], Dmitri Kolychev, Feng Sun, Chun Liang, Robert M. Freeman, Jr.
*[1]mmpratt@uga.edu, University of Georgia;
[2]leepratt@uga.edu, University of Georgia*

MAGIC Microarray is the gene-expression half of a modular approach to a genomic, integrated and comprehensive resource for gene discovery and expression. It includes a MIAME-compliant Oracle 9i database tightly integrated with the DNA sequence, clustering, and annotation information in MAGIC Gene Discovery. Additional information is at http://fungen.org.

F-77

Testing Non-IID-hypotheses in Microarray Experiments

Claus D. Mayer[1]
[1]claus@bioss.ac.uk, Biomathematics & Statistics Scotland

Many common tests for differential expression in microarray experiments are based on a permutation principle that requires the data to be independently identically distributed (IID) under the null hypothesis. We study methods to check microarray data for variance heterogeneity and discuss which tests to use if the IID-assumption is violated.

F-78

A Software Platform for the Storage, Visualization and Analysis of Microarray Gene Expression Data

Thomas M. Karopka[1], Sven Bansemer[2], Thomas Scheel, Änne Glass, Lothar Gierl
[1]thomas.karopka@medizin.uni-rostock.de, University of Rostock; [2]sven.bansemer@medizin.uni-rostock.de, University of Rostock

gEn0m is a comprehensive client server system for oligonucleotide microarray analysis. Besides the storage and management capability, the software features a rich set of algorithms for analysis as well as the ability to link the data to major genomic Databases. gEn0m is available upon request from the authors.

Posters

F-79

Comparison of Ratio Statistics in cDNA Microarray Experiments

Kiwoong Kim[1], Taesung Park[2], Sung-Gon Yi, SeungYeoun Lee, Jin Hyuk Kim and Yong-Sung Lee
[1]kiwoong@biostats.snu.ac.kr, Department of Statistics, Seoul National University; [2]tspark@stats.snu.ac.kr, Department of Statistics, Seoul National University

In this paper, we consider two types of log-transformed ratio statistics. We compare the new ratio statistics with the conventional ratio statistic commonly used in cDNA microarray experiment.

F-80

Interferon Inducible Expression of Human Tyrosyl-tRNA Synthetase in Cancer Cell Lines

Sergiy Ivakhno[1], Alexander Kornelyuk[2]
[1]ivahser@i.com.ua, Taras Shevchenko Kyiv National University; [2]kornelyuk@imbg.org.ua, Institute of Molecular Biology and Genetics of NAS of Ukraine

Bioinformatic analysis of microarray data revealed interferon-gamma inducible transcription of human tyrosyl-tRNA synthetase(TyrRS) in cancer cell lines. Upstream sequences of TyrRS were also found to contain interferon response elements. Our data suggest the presence of antiangiogenic activities for C-module of TyrRS and implicate regulation of these activities at the transcriptional level.

F-81

From Microarray Data to Pathways: a Bioinformatics Pipeline for Rational Drug Discovery

Pierre Stanislawski[1], Simon de Bernard, John Brozek, Gautier Koscielny, Laurent Buffat
[1]pierre.stanislawski@it-omics.com, IT.OMICS

We present a bioanalysis platform for deciphering biological events from microarray experiments. It includes probe level analysis of affymetrix chi for enhanced gene identification, gene/protein annotation, and extraction of gene/protein functional relationships described in MEDLINE. Exploration of the neighborhood of regulated genes allows to identify phenotype specific pathway and potential

F-82

Gene Expression Signatures in Blood Cells Classify and Predict Posttraumatic Stress Disorder Among Trauma Survivors

Ronen Segman[1], Noa Shefi[2], Prof. Nir Friedman, Dr. Naftali Kaminski, Tanya Goltser-Dubner
[1]sronen@md2.huji.ac.il, Hadassah medical center, Jerusalem; [2]shefi@cs.huji.ac.il, Hebrew Uni. Jerusalem

Gene expression profiles of peripheral blood that was taken from trauma survivors few hours after the trauma, and 4 months later, can identify the progression of post-traumatic stress disorder. This information may provide clues to the pathogenesis of a psychological disorder, and be applied to guide early treatment.

F-83

maxdLoad2: Managing Microarrays and MetaData - From Spreadsheet to Relational Database to Structured Document

David Hancock[1], Andy Brass[2]
[1]d.hancock@cs.man.ac.uk, The University of Manchester; [2]a.brass@man.ac.uk, The University of Manchester

This poster begins with an overview of the various methodologies currently in use for storing meta-data about transcriptomic experiments. It then introduces ‘maxdLoad2’ an open-source microarray data storage solution built around a relational database. The novel aspects of this work are the tools for data importing and flexible template-driven, structured document generation.

F-84

New Challenges in Gene Expression Data Analysis and the Extended Suite of Web Tools GEPAS and FatiGO/Fatiwise

Javier Herrero[1], Juan M. Vaquerizas[2], Fatima Al-Shahrour, Lucia Conde, Alvaro Mateos, Javier Santoyo Ramon Diaz-Uriarte and Joaquin Dopazo
[1]jherrero@cnio.es, CNIO; [2]jvaquerizas@cnio.es, CNIO

GEPAS (http://gepas.bioinfo.cnio.es., Herrero et al., 2003 NAR), the web-based suite of tools for microarray data analysis, has evolved to cope with the new challenges that have recently arisen in the this field field. GEPAS incorporates several interconnected methods, for normalization, clustering, gene selection, predictor building, datamining, and other.

F-85

What Your Microarrays are Really Trying to Tell You

L.A.Gilhuijs-Pederson[1], AHC van Kampen[2]
[1]l.a.pederson@amc.uva.nl, AMC/UvA;
[2]a.h.vankampen@amc.uva.nl, AMC/UvA

Novel ANOVA-style algorithms are presented which lower the threshold of discovery while reducing the number of false positives due to noise/biases. Applied to the Gene Set Enrichment dataset of Mootha et. al., hundreds of genes consistent with what is already understood about diabetes are individually detected.

F-86

Metabolite Fingerprinting: an ICA Approach

Matthias Scholz[1], Stephan Gatzek[2], A.Sterling, O.Fiehn, J.Selbig
[1]scholz@mpimp-golm.mpg.de, MPI of Molecular Plant Physiology, Potsdam, Germany;
[2]gatzek@mpimp-golm.mpg.de, MPI of Molecular Plant Physiology, Potsdam, Germany

Independent component analysis (ICA) is applied to high dimensional samples, given by metabolite mass spectra. It is shown that ICA has a higher informative power when it is combined with suitable pre-processing and evaluation criteria. The resulting components are interpreted. Two of the resulting variables discriminate between sample types, while another variable is correlated to an experimental artefact.

F-87

Meta-analysis of Chemosensitivity Profiles and Multiple Gene Expression Datasets of NCI60 Cell Lines

Aedin C Culhane[1], Guy Perriére[2], Sharon McKenna, Desmond G. Higgins
[1]Aedin.Culhane@ucd.ie, Conway Institute, UCD;
[2]perriere@biomserv.univ-lyon1.fr , Laboratoire de Biometrie et Biologie Evolutive, Universite Claude Bernard

We have applied a powerful multivariate technique called co-inertia analysis to explore trends in drug sensitivity data with strong correlations to trends in gene expression data. We have analysed three independent gene expression datasets of the NCI60 panel of cancer cell lines. Significant classes of drug-gene trends were identified.

F-88

Missing Value Estimation for DNA Microarrays Based on k Nearest Individuals

Anja Bråthen Kristoffersen[1], Ole Christian Lingjærde[2]
[1]anjab@ifi.uio.no, Department of Informatics, University of Oslo; [2]ole@ifi.uio.no, Department of Informatics, University of Oslo

A common method for estimation of missing values in Microarrays is to use a weighted mean of the k nearest genes. Here, a novel method is proposed based on a weighted mean over the k nearest individuals. A comparison of different missing value estimation methods is also included

F-89

Web-based Application for Bayesian Modelling of Regulatory Networks from Continuous Microarray Data

Carsten Friis[1], Ole Winther[2], Hanne Jarmer, Anders Gorm Petersen, Steen Knudsen
[1]carsten@cbs.dtu.dk, CBS; [2]owi@imm.dtu.dk, IMM

We have developed a Bayesian framework to quantify how well proposed regulatory network graphs fit available data, thus identifying the most eligiable of several candidate regulatory hypotheses. To exemplify, we investigate the TnrA/GlnR system in Bacillus subtilis. The framework can be accessed at: www.cbs.dtu.dk/services

F-90

Gene Expression Profiling Distinguish Subgroups in T-Cell Lymphoma

Benoît Ballester[1], Rémi Houlgatte[2], Philippe Gaulard and Luc Xerri
[1]ballester@tagc.univ-mrs.fr, TAGC INSERM ERM206;
[2]houlgatte@tagc.univ-mrs.fr, TAGC INSERM ERM206

Peripheral T-cell lymphomas (PTCLs) represent a rare and heterogeneous group of cancers which to date remained to be described using microarray. Our results support the existence of distinct gene expression subtypes but also distinguish new type of PTCLs with different outcome providing a basis for improved disease diagnosis and management.

F-91

Cyber-T: Identifying Gene Changes in DNA Microarray Data

Suman Sundaresh[1], She-pin Hung[2], G. Wesley Hatfield, Pierre Baldi
[1]suman@uci.edu, University of California, Irvine;
[2]shung@uci.edu, University of California, Irvine

Cyber-T is a statistics program with a web interface that can be conveniently used on high-dimensional array data for the identification of statistically significant differentially expressed genes and for the estimation of experiment-wide false positive and negative levels. It can be freely accessed at http://www.igb.uci.edu/servers/dmss.htm

F-92

Search Engine Technology for Microarray Expression Data

Julie Morrison[1], Rainer Breitling [2], Desmond J Higham, David R. Gilbert
[1]rs.jmor@maths.strath.ac.uk, University of Strathclyde / University of Glasgow; [2]r.breitling@bio.gla.ac.uk, University of Glasgow

Where differentially expressed genes tend to be well-connected by other evidence (annotations, interactions ...), re-ordering genes by combining expression and connectivity information will be as effective means to highlight most relevant genes. We propose a technique using search engine technology to combine expression data with connectivity information from, e.g. GO annotations.

F-93

Calculating the Statistical Significance of Changes in Pathway Activity from Gene Expression Data

Jörg Rahnenführer[1], Francisco S. Domingues[2], Jochen Maydt, Thomas Lengauer
[1]rahnenfj@mpi-sb.mpg.de, Max-Planck-Institute for Informatics ; [2]doming@mpi-sb.mpg.de, Max-Planck-Institute for Informatics

We present a statistical approach to score changes in activity of metabolic pathways from gene expression data. The method identifies the biologically relevant pathways with corresponding statistical significance. Suitable measures for co-regulation of genes are adaptively determined, and the topology of the metabolic pathway is integrated into the significance score.

F-94

Cyanobacterial Gene Expression: Analysis with a Novel Desktop Microarray Facility

Heike Eckes[1], Röbbe Wünschiers[2]
[1]he@heikeslair.net, University of Cologne;
[2]rw@biowasserstoff.de, University of Cologne

A novel microarray technique has been used to investigate gene expression in Anabaena sp. PCC 7120. For data processing a specific database was setup. It allows to correlate the gene-expression data obtained from the experiments with metabolic pathways and other data. Project partners can access this database at HyDaBA.

F-95

A Bayesian Method for Clustering Gene Expression Temporal Profiles from Microarray Data

Fulvia Ferrazzi[1], Paolo Magni[2], Riccardo Bellazzi
[1]fulvia@aim.unipv.it, Università di Pavia;
[2]magni@aimed11.unipv.it, Università di Pavia

We propose a new Bayesian method for clustering gene expression temporal profiles. The method, based on Random walk models, is designed to offer a principled way to identify the number of clusters and to explicitly model inter-gene variability within each cluster. A validation was performed on real and simulated data.

F-96

Extracting Networks of Influences from Microarray Data

Liviu Badea[1]
[1]badea@ici.ro, National Institute for Research in Informatics

We apply a conditional independence-based network structure inference algorithm to the lung cancer dataset of Garber et al. Although network inference algorithms cannot help in reconstructing the complete causal network, we generate a sample-specific network that succinctly describes the dependencies among the variables in the given dataset.

F-97

ProbeHunter: Computer Aided Probes Design for Microarray

Raoul J.P. Bonnal[1], Claudio Ferretti[2], Gianluca De Bellis, Giancarlo Mauri
[1]raoul.bonnal@itb.cnr.it, Consiglio Nazionale delle Ricerche, Istituto di Tecnologie Biomediche; [2]ferretti@disco.unimib.it, Università di Milano - Bicocca, Dipartimento di Informatica, Sistemistica e Comunicazione

Several biological contexts require molecular probes which recognize a single sequence or a group of sequences within a set of strongly homologous sequences, microorganisms 16S rRNA or human HLA gene. ProbeHunter is a software tool which aids researchers in identifying such probes to be used in microarray technology.

F-98

Mapping of Probes and Reconstruction of Transcripts from Genome-wide High Density Oligonucleotide Microarrays

Javier De Las Rivas[1], Alberto De Luis[2], Carlos Prieto
[1]jrivas@usal.es, CIC; [2], CIC

Bioinformatic tool DAGA (Dynamic Annotation of GeneChip probesets from Affymetrix) to map probes from high-density oligo nucleotide expression Microarrays. The method uses BLAST against current curated genome Databases to locate each individual oligonucleotide probe either in cDNA Databases (RefSeq, UniGene, ENSEMBL) or in full genome Databases (ENSEMBL golden path).

F-99

CDNA Probe Selection from Expressed Sequence Tags (ESTS) for Microarray Design

Yian A. Chen[1], David Mckillen[2], Shuyuan Wu, Javier Robalino, Matthew J. Jenny, Paul S. Gross, Gregory W. Warr, Jonas S. Almeida
[1]chenya@musc.edu, Department of Biostatistics, Bioinformatics, and Epidemiology, Medical University of South Carolina, USA; [2]mckilldj@musc.edu, Department of Biochemistry and Molecular Biology, Medical University of South Carolina, USA

We propose a microarray design procedure to select subsets of ESTs for spotting on cDNA microarray by forming non-redundant clusters, which retain functional diversity, assessed through the Gene Ontology annotation. Different clustering methods are variably advantageous depending on the purposes of the experiments. The proposed quantitative indices of tracking amalgamation schemes enable the optimal microarray design.

F-100

GenMAPP and MAPPFinder 2.0: Tools for Viewing and Analyzing Genomic Data Using Gene Ontology and Biological Pathways

Kam D. Dahlquist[1], Scott W. Doniger[2], Nathan Salomonis, Kristina Hanspers, Karen Vranizan, Lynn Ferrante, Alexander C. Zambon, Jeff C. Lawlor, Steven C. Lawlor, and Bruce R. Conklin
[1]kadahlquist@vassar.edu, Vassar College; [2]GenMAPP@gladstone.ucsf.edu, Gladstone Institute of Cardiovascular Disease

GenMAPP is designed for viewing expression data on biological pathways. GenMAPP automatically color-codes the genes according to criteria supplied by the user. MAPPFinder matches expression data to the Gene Ontology and indicates whether there is a statistically significant over-representation of genes meeting the user's criterion for each GO term. http://www.GenMAPP.org.

F-101

SNP Selection for a Microarray Based High-throughput Genotyping Assay

Guoying Liu[1], Jing Huang[2], Xiaojun Di, Raymond Wheeler, Giulia C Kennedy, Rui Mei, Alan J. Williams and Keith W Jones
[1]guying_liu@affymetrix.com, Affymetrix; [2]jing_huang@affymetrix.com, Affymetrix

We developed a microarray based high-throughput SNP genotyping method for simultaneously genotyping over 100,000 SNPs. The Microarrays were designed to interrogate SNPs predicted to be located on 500-2000-base-pair fragments of XbaI or HindIII digestion. The pool of SNPs was collected from TSC (the SNP Consortium), dbSNP and Perlegen.

F-102

Effect of Amplification on Gene Expression Profiles

Rachel I.M. van Haaften[1], Blanche Schroen[2], Ben J.A. Janssen, Jos F.M. Smits, Yigal M.Pinto and Chris T.A. Evelo
[1]Rachel.vanhaaften@bigcat.unimaas.nl, BiGCaT Bioinformatics BMT-TU/e & UM ; [2]b.schroen@cardio.unimaas.nl, Experimental and Molecular Cardiology/ CARIM, Maastricht University

Gene expression Microarrays (Rat expression Set 230 Array from Affymetrix) are used to compare gene expression profiles between T7 linear amplified biopsy samples and non-amplified full tissue samples from rat. For this comparison several types of analyses were used. The (in our view very interesting) results will be presented.

F-103

Expression-based Detection of Statistically-significant Alternative Splicing Events

Melissa Cline[1], John Blume[2], Simon Cawley, Tyson Clark, Jing-Shan Hu, Hui Wang, and Alan Williams
[1]*melissa_cline@affymetrix.com, Affymetrix, Inc;*
[2]*john_blume@affymetrix.com, Affymetrix, Inc.*

Novel microarray platforms can elucidate alternative splicing by measuring the expression levels of millions of transcript features simultaneously. This work describes a method to detect alternative splicing events by applying ANOVA statistical testing in conjunction with a Li and Wong-style estimation.

F-104

XPS, a Novel Framework for Distributed Storage and Analysis of Microarray Data in the Terabyte Range: An Alternative to BioConductor

Christian Stratowa[1]
[1]*christian.stratowa@vie.boehringer-ingelheim.com, Boehringer Ingelheim Austria*

XPS is a Data Mining tool for large-scale storage and analysis of microarray data. It is based on ROOT, a framework developed for high energy physics data in the petabyte range. The specialized storage method as compressed trees could be adopted as novel standard for storage of microarray data.

F-105

Co-regulated Gene Clusters in Breast Cancer Determined by Fuzzy Clustering and Promoter Analysis

Miroslava Cuperlovic-Culf[1], Nabil Belacel[2], Mark Laflamme, Rodney J. Ouellette
[1]*miroslavac@health.nb.ca, BRMI;* [2]*Nabil.Belacel@nrc-cnrc.gc.ca, NRC*

A fuzzy clustering method is used for the determination of coregulated genes derived from the subsets of tumors in the breast cancer data set. Following the cluster analysis, we looked for more detailed information about the coregulation, and possible function of the most significant gene clusters using promoter analysis and gene ontology.

F-106

Normal Probability Plots for Quality Assessment of Microarray Experiments

Jeroen Pennings[1], Siem Heisterkamp[2]
[1]*Jeroen.Pennings@rivm.nl, RIVM;*
[2]*SH.Heisterkamp@rivm.nl, RIVM*

We present the use of normal probability plots for comparing data distributions among microarray slides. This can be used as a means for slide quality assessment, especially to identify experimental outliers. Knowledge of the data distribution will also help to optimize the power of statistical analyses for microarray experiments.

G-1

Automatic Service Composition of Data and Applications in Drug Discovery

Dr.Tanveer Syeda-Mahmood[1], Dr. Bhooshan Kelkar[2]
[1]stf@almaden.ibm.com, IBM Almaden Research Center; [2]bkelkar@us.ibm.com, IBM Healthcare and Life Sciences

Rapid advances in sequencing and HTS coupled with emergence of numerous commercial and public domain analytic tools necessitate seamless composition of data (numerical, text), their analysis and visualization components. The current paper presents mechanisms and results of services composition with implications in all aspects of drug discovery, development and delivery.

G-2

An Experimental Grid Interface for Protein Sequence Analysis

Blanchet C.[1], Lecluse A.[2], Combet C., Deleage G.
[1]Christophe.Blanchet@ibcp.fr, CNRS IBCP; [2], CNRS IBCP

Bioinformatics has to derive valuable information from the large amounts of data provided by all the complete genome projects (1087, march 2004). Grid computing, as EGEE infrastructure, would be a viable solution to distribute data, computing and storage resources for Genomics. GPS@ web portal, "Grid Protein Sequence Analysis" (http://gpsa.ibcp.fr), could be a user-friendly interface for these grid genomic resources.

G-3

Mapping Alternate Splice Variants onto Protein Structures Using Biocomputing Methods

Shoba Ranganathan[1], Vivek Gopalan[2], Chitra L. Madhwacharyula
[1]shoba@els.mq.edu.au, Macquarie University & National University of Singapore; [2]vivek@bic.nus.edu.sg, National University of Singapore

Alternative splicing is a major mechanism responsible for the complexity of the eukaryotic genome, resulting in several proteins arising from a single gene. To understand how these alternatively spliced products affect functionality, we present a schema for mapping alternate splice variants onto 3D protein structures

G-4

A Study on Aligning High-dimensional MS Proteomics Data

Amol Prakash[1], Benno Schwikowski[2]
[1]aprakash@systemsbiology.org, Institute for Systems Biology, Seattle, WA; [2]benno@systemsbiology.org, Institut Pasteur, Paris, France

Analysis of complex protein mixtures on the basis of liquid chromatography (LC), followed by mass spectrometry (LC-MS or LC-MS/MS), is a key technology for the systematic large-scale exploration of cellular processes. We present a novel and general integrative approach for the interpretation of this data, and our study of the results on different large datasets.

G-5

Language Independent Web Service Interoperability: A UML-Centric Approach for Creating Bioinformatics Web Services

Chad Matsalla[1]
[1]matsallac@agr.gc.ca, Agriculture & Agri-Food Canada

A UML model-centered approach to developing interoperable bioinformatics software provides numerous benefits including design, implementation, and interoperability. This is important because of the vast diversity of applications used to accomplish analytical tasks. A system that generates interoperable software components from UML models allows applications to exchange complex objects regardless of the choice of implementation language.

G-6

Proteochemometric Modelling of Cyclic Peptides Interaction with Wild-type and Chimeric Melanocortin Receptors

Aleksejs Kontijevskis[1], Peteris Prusis[2], Ramona Petrovska, Ilze Mutule, Staffan Uhlén, Jan Komorowski, Jarl E.S. Wikberg
[1]Aleksejs.Kontijevskis@lcb.uu.se, The Linnaeus Centre for Bioinformatics, Uppsala University; [2]Peteris.Prusis@farmbio.uu.se, The Linnaeus Centre for Bioinformatics, Uppsala University

The interaction of peptides with wild-type and chimeric melanocortin receptors was analysed using proteochemometrics technique. The descriptors of peptides and receptors were correlated to the affinities using Partial Least Squares projection to latent structures modelling. Application of proteochemometrics approach allowed us to explain the selectivity of peptides towards MC4 receptor.

G-7

Efficient Drug Screening Using Active Learning

Asogawa[1], Osoda[2]
[1]m-asogawa@bq.jp.nec.com, NEC Corporation; [2]osoda@aj.jp.nec.com, NEC Corporation

We applied the active learning method as an effective chemical screening method and shown its effectiveness by both computer simulations using known chemical data and actual wet experiments. It is shown that one fifth screening is enough for finding ninety percent of all hit compounds.

G-8

Automatic Annotation of Enzyme Classes with Disease Information from the Biomedical Literature

Hofmann, Oliver[1], Schomburg, Dietmar[2]
[1]o.hofmann@smail.uni-koeln.de, Department of Biochemistry, University of Cologne; [2]d.schomburg@uni-koeln.de, Department of Biochemistry, University of Cologne

We present a system to automatically annotate enzyme classes with disease related information extracted from the biomedical literature. Utilizing the co-occurence of concepts allowed for a precision of 92% at 50% recall, sufficient for the inclusion in a high-quality database.

G-9

Cloning Computing Workstations for Scientific Research

Daniel Swan[1], Bela Tiwari[2], Milo Thurston, Nic Betrand, Dawn Field
[1]dswan@ceh.ac.uk, EGTDC; [2]btiwari@ceh.ac.uk, EGTDC

We describe a strategy that allows the creation of scientific workstations, customised for researchers. The approach is based on the customisation of a master workstation that can be "cloned". In this poster we outline the strategy involved in developing "Bio-Linux" workstations for researchers in environmental genomics.

G-10

Association of Variations in NFKBIE with Graves' Disease Using Classical and myGrid Methodologies

Peter Li[1], Anil Wipat[2], Keith Hayward, Claire Jennings, Kate Owen, Tom Oinn, Robert Stevens, Simon Pearce
[1]peter.li@ncl.ac.uk, Computing Science, University of Newcastle; [2]anil.wipat@ncl.ac.uk, Computing Science, University of Newcastle

We have compared the use of myGrid and classical approaches for performing bioinformatics experiments in the genetic analysis of Graves' disease. Both the classical and myGrid methodologies identified NFKBIE as a candidate gene involved in Graves' disease. The myGrid software is available at http://www.mygrid.org.uk.

G-11

Glycan Alignment and Scoring

Kiyoko F. Aoki[1], Hiroshi Mamitsuka[2], Minoru Kanehisa
[1]kiyoko@kuicr.kyoto-u.ac.jp, Bioinformatics Center, Kyoto University; [2]mami@kuicr.kyoto-u.ac.jp, Bioinformatics Center, Kyoto University

The alignment of glycan structures has been previously presented, but elementary scoring methods were used. We have performed a statistical analysis of glycan data and constructed a glycan score matrix. We present our calculations to generate our score matrix and illustrate the improvement in performance of the glycan alignments.

G-12

African Society of Bioinformatics and Computational Biology

Patrick O Erah[1], Nicola Mulder[2], Jaco de Ridder, Daniel Masiga, Raphael Isokpehi
[1]erah@uniben.edu, University of Benin, Benin City, Nigeria; [2]nicky@sanbi.ac.za, SANBI, University of Western Cape, South Africa

African Society of Bioinformatics and Computational Biology (ASBCB) is a non-profit professional association dedicated to the advancement of bioinformatics and computational biology in Africa. Transforming from the African Bioinformatics Network (ABioNET), ASBCB was established in February 2004 at a meeting in Cape Town, South Africa.

G-13

Duplicons Harbor Complex Forms of Sequence Variation that may Mislead SNP Genotyping Studies

David Fredman[1], Anthony Brookes[2], Stefan White, Johan den Dunnen
[1]david.fredman@cgb.ki.se, Karolinska Institute; [2]anthony.brookes@cgb.ki.se, Karolinska Institute

We genotyped 157 SNPs across 17 duplicated segments in diploid and fully homozygous genomes. This revealed a new type of polymorphism representing the sum of the signals from many individual duplicon elements that can masquerade as normal SNPs when genotyped.

G-14

BioJava 2: The Next-generation Bioinformatics Library for Java

Matthew Pocock[1], Thomas Down[2]
[1]matthew.pocock@ncl.ac.uk, University of Newcastle Upon Tyne; [2]td2@sanger.ac.uk, The Sanger Center

BioJava is a mature open-source project, affiliated with the Open Bioinformatics Foundation. BioJava 2 is a ground-up redesign of these APIs, providing greatly enhanced ease-of-use for the casual user, while still allowing the power user total control. The pervasive use of ontologies and a custom data-integration engine makes BioJava 2 the natural choice for developing post-genomic applications.

G-15

Accessing Databases Provided by the European Bioinformatics Institute Using Web Services

Sharmila Pillai[1], Rodrigo Lopez[2], John Tate, Sameer Velanker, Adel Golovin, Martin Senger, Tom Oinn, Kim Henrick, Peter Rice
[1]sharmila@ebi.ac.uk, EBI; [2]rls@ebi.ac.uk, EBI

The EBI provides various biological Databases and analysis tools. Integrating these heterogeneous resources to obtain useful data for further analysis poses a challenge to bioinformaticians. To overcome this, the EBI uses Web Services technology to provide programmatic access to these resources. These services are accessible at http://www.ebi.ac.uk/Tools/webservices.

G-16

The Most Important Protein Superfamilies as Revealed by Graph Theoretic Analysis of the Protein Structural Interactome

Dan Bolser[1], Jong Park[2], Panos Dafas, Michael Schroder, Richard Harrington
[1]dmb@mrc-dunn.cam.ac.uk, MRC-Dunn; [2]j@bio.cc, Korean Advanced Institute of Science and Technology

PSIMAP is a broad protein family interaction map, which highlights the evolutionary history of protein-protein interactions. Here we analyse this network of family-family interactions for its topological features. We conclude that the most important and ubiquitous protein superfamilies are highlighted in the network topology.

G-17

KEGG API: New Features of KEGG Web Service

Shuichi Kawashima[1], Toshiaki Katayama[2], Minoru Kanehisa
[1]shuichi@kuicr.kyoto-u.ac.jp, Bioinformatics Center, Institute for Chemical Research, Kyoto University; [2]ktym@hgc.jp, Human Genome Center, Institute of Medical Science, University of Tokyo

The KEGG API provides valuable means for accessing the KEGG system such as searching and computing biochemical pathways in cellular processes or analyzing the universe of genes in the completely sequenced genomes. The users access the KEGG API server using SOAP over HTTP.

G-18

Providing Text Mining Services for the Bio-community

National Centre for Text Mining[1], National Centre for Text Mining[2]
[1]jak@co.umist.ac.uk, National Centre for Text Mining; [2]S.Ananiadou@salford.ac.uk, National Centre for Text Mining

The main aims and objectives of a newly established UK Centre for Text Mining, funded by JISC, BBSRC and EPSRC are outlined. The main objective of the Centre (which has a special focus on biology/biomedicine) is to provide text mining services initially to the academic community, followed by extension of services to other sectors, including industry.

G-19

Data Modeling and Universal Data Exchange

Rasmus[1], Wim[2], Wayne Boucher, John Ionides, Anne Pajon, Tim Stevens, Ernest Laue
[1]Fogh, H; [2]Vranken,

CCPN presents a framework for automatically generating and maintaining data exchange standards and associated subroutine libraries solely from a description of the structure of the data. We also present a data model for Systems Biology, including macromolecular NMR spectroscopy, and biomolecular LIMS systems, generated within the framework.

G-20

eScience and Pfam

Robert Finn[1], Alex Bateman[2]
[1]rdf@sanger.ac.uk, Wellcome Trust Sanger Institute;
[2]agb@sanger.ac.uk, Wellcome Trust Sanger Institute

Pfam is part of eFamily, a data intensive eScience project aimed to bridge protein structure and sequence resources. Deployed web-services enable access to data and compute resources by the scientific community. Integration of web-services into work flow engines allow eScience to be performed. Research using the web-services is presented.

G-21

Exploring Williams-Beuren Syndrome Using myGrid

Robert Stevens[1], Hannah Tipney[2], Chris Wroe, Tom Oinn, Martin Senger, Phillip Lord, Carole Goble, Andy Brass, May Tassabehji
[1], University of Manchester; [2]h.j.tipney@man.ac.uk, University of Manchester

We describe the use of myGrid middleware to run in silico experiments in a semantic Grid aware environment. Workflows have been used to map the complex genomic region associated with Williams-Beuren Syndrome and scientists have produced biologically interesting and valid results, while improving their productivity compared to previous manual undertakings

G-22

Literature Mining of Medline/MeSH for Biomarkers in Systemic Lupus Erythematosus

Douglas M. Blair[1], Hong-wei Sun[2], Gabor G. Illei, John J. Grefenstette, Peter E. Lipsky
[1]blaird1@mail.nih.gov, NIH/NIAMS;
[2]sunh1@mail.nih.gov, NIH/NIAMS

Systemic Lupus Erythematosus (SLE) is an autoimmune disease of unknown etiology, variable clinical manifestation, and imprecise characterization. We utilize a custom database of the medical literature and its metadata in an attempt to infer novel, testable candidates for biomarkers in lupus risk prediction, diagnosis, treatment, and prognosis

G-23

Investigation of Nuclear Factor-kappa Beta (NF-kb) and DNA Interactions Using Macromolecular Docking and Simulation Experiments

Mahmud Tareq Hassan Khan[1], Arjumand Ather[2], Robarto Gambari
[1]mahmud.khan@hej.edu, Dr. Panjwani Center for Molecular Medcine and Drug Research, International Center for Chemical Sciences;
[2]arjumand.ather@hej.edu, Dr. Panjwani Center for Molecular Medcine and Drug Research, International Center for Chemical Sciences

Macromolecular docking experiments of the NF-kb with ssDNA has been done using the docking algorithm HEX 4.2 and molecular dynamic simulation experiments also been employed. It was found that there is an interaction between T5′ and SER335 and distances between these two were calculated to be 7.19 and 8.75 Å.

G-24

The Notebook Project: A Personal Data Store and Smart Client Environment for the Researcher

Gregory B. Quinn[1], Blair Jennings[2], Mark A. Miller
[1]quinn@sdsc.edu, San Diego Supercomputer Center;
[2]blair@sdsc.edu, San Diego Supercomputer Center

The Notebook is a new breed of personal computer application which addresses limitations of the web paradigm by providing the researcher with rich interfaces to the next generation of online SOAP-based data and analytical services, coupled with the convenience of a personal local database, automated background data searching and a collaboration environment. Information on the web at: http://www.notebookproject.org

G-25

Constructing an Interactive Transcriptome Atlas of the Mouse Brain at Cellular Resolution

James P. Carson[1], Gregor Eichele[2], Tao Ju, Christina Thaller, Joe Warren, Wah Chiu
[1]james.carson@bcm.tmc.edu, National Center for Macromolecular Imaging, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX, USA; [2]gregor.eichele@mpihan.mpg.de, Max-Planck Institute of Experimental Endocrinology, Hannover, Germany

Geneatlas.org is a molecular atlas of the postnatal mouse brain consisting of gene expression patterns generated via high-throughput in situ hybridization. These patterns are mapped into the common context of a deformable geometric brain, thus enabling spatial queries that can assist in identifying genes involved in biological pathways.

G-26

The Immunology Grid: a Computational Pipeline for Immunoinformatics

Barry R Smith[1], David S Moss[2], Clare E. Sansom, Shikta Das, Mark Halling-Brown, Raheel Shaban, Matthew Davies.
[1]b.smith@mail.cryst.bbk.ac.uk, Birkbeck College; [2]d.moss@mail.cryst.bbk.ac.uk, Birkbeck College

The 'Immunology Grid' project aims to integrate a series of methods which together will form an easy-to-use Grid-enabled tool for prediction of binding affinities between MHC molecules and peptides. Improved knowledge of peptide binding has far-reaching benefits in areas of molecular medicine, such as histocompatibility, vaccine design, and autoimmunity

G-27

Integrating Bioinformatics Resources Using Shims

Duncan Hull[1], Robert Stevens[2], Phillip Lord, Carole Goble
[1]duncan.hull@cs.man.ac.uk, Computer Science Dept, University of Manchester; [2]duncan.hull@cs.man.ac.uk, Computer Science Department, University of Manchester

Workflows allow integration of autonomous and distributed resources during in silico experiments. Ideally workflows should hide the details of orchestrating resources. In practice, many intermediate or ``shim'' services have to be constructed in order to enable heterogeneous services to interoperate. This poster presents an initial classification of bioinformatics shims

G-28

2D-PAGE

Jonathan Epstein[1], Jonathan Gordin[2], Donita Garland, Gary Giulian
[1]Jonathan_Epstein@nih.gov, NICHD; [2]jsg53@cornell.edu, NICHD

We present a data model for two-dimensional gel electrophoresis (2D-PAGE) in the Protege-2000 system. The goal is to produce an expert system to permit a researcher who is not a 2D-PAGE expert to select an appropriate protocol for running their gel.

G-29

Community Structure in Epidemiological Networks

Shweta Bansal[1]
[1]sbansal@ices.utexas.edu, University of Texas, Austin

Mathematical models of human contact patterns allow us to quantify important information about the process of disease spread. We are developing methods for capturing community structure in contact networks. We show that the presence of community structure in a network accelerates local disease outbreaks but slows large-scale epidemic transmission.

H-1

Intron Evolution in the Vertebrates

Michael M. Hoffman[1], Ewan Birney[2]
[1]hoffman@ebi.ac.uk, EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, England; and Graduate School of Biological, Medical and Veterinary Sciences, Cambridge, England; [2]birney@ebi.ac.uk, EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, England

We developed a model of gene evolution that incorporates varying selective pressures on introns and coding sequence, and a method for aligning intronic regions among homologous genes. This method finds many gene pairs with significant intronic conservation. We also propose that intron conservation provides a useful neutral evolutionary distance measure.

H-2

The False Positive Selection Induced by Saturated Sequence Divergences

Feng Wang[1], Yasuhiko Wada[2]
*[1]wangfeng@cc.saga-u.ac.jp, Saga University;
[2]ywada@cc.saga-u.ac.jp, Saga University*

We demonstrate that when the sequence divergences are saturated, both distance and likelihood method can induce the false detection of positive selection, such computation results should be treated cautiously.

H-3

Genome-scale Phylogenetic Tree Construction Based on Cellular Metabolic Pathway Content

Soon Ho Hong[1], Tae Yong Kim[2], Dong-Yup Lee, Sang Yup Lee
*[1]totenkof@webmail.kaist.ac.kr, Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical & Biomolecular Engineering and BioProcess Engineering Research Center, Korea Advanced Institute of Science and Technology;
[2]kimty@kaist.ac.kr, Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical & Biomolecular Engineering and BioProcess Engineering Research Center, Korea Advanced Institute of Science and Technology*

A new phylogenetic tree can be constructed on the basis of metabolic network profiles for completely sequenced 42 microorganisms. The resultant phylogenetic tree represents both the evolutionary time scale (change in sequence) and the evolutionary process (change in metabolism).

H-4

A Phylogenetic Analysis of the Y-family of DNA Polymerases

Shwetal Patel[1], Errol C Friedberg[2]
*[1]shwetal.patel@utsouthwestern.edu, The University of Texas Southwestern Medical Center;
[2]errol.friedberg@utsouthwestern.edu, The University of Texas Southwestern Medical Center*

Phylogenetic analysis of the Y-family of DNA polymerases reveals a large number of conserved regions in eukaryotic members. Based on experimental evidence of known interaction partners, correlation of mutations is computed for the conserved regions to assess the applicability of the in silico two-hybrid approach to predict protein-protein interaction partners.

H-5

Gamma Chain Receptor Interleukins: Comparison of the Primate and Rodent Immune Systems Yields Evidence for Positive Selection

Mary J. O'Connell[1], James O.McInerney[2]
*[1]mary.oconnell@dcu.ie, Dublin City University, Glasnevin, Dublin 9, Ireland;
[2]james.o.mcinerney@may.ie, National University of Ireland Maynooth, Maynooth, Co.Kildare*

The interleukin-2 receptor g chain (gc) is an indispensable subunit of the functional receptor complexes for interleukin-2, -4, -7, -9, -15 and -21. We show that competition for the recruitment of the gc component of these receptor complexes has caused rapid evolution in this group of interleukins.

H-6

WeightLESS: a Program for Branch Testing in Phylogenies Reconstructed from Distance Measures Using Weighted Least-squares Likelihood Ratio Test

Borys Wróbel[1], Rafael Sanjuán[2]
*[1]borys.wrobel@uv.es, Institut Cavanilles de Biodiversitat i Biología Evolutiva, Universitat de València;
[2]rafael.sanjuan@uv.es, Institut Cavanilles de Biodiversitat i Biología Evolutiva, Universitat de València*

WeightLESS is a program to perform weighted least-squares likelihood ratio test (WLS-LRT) for testing phylogenies reconstructed from distance measures. WLS-LRT uses only the distances and some of their associated variances; the program implements also the Felsenstein F-test meant to be used when no variances are known, and is available for download at http://www.iopan.gda.pl/~wrobel.

H-7

Evolution of Paralogous Gene Regulation in Saccharomyces Cerevisiae

Ilan Wapinski[1], Nir Friedman, Aviv Regev, Avi Pfeffer
[1]ilan@eecs.harvard.edu, Division of Engineering and Applied Sciences and Bauer Center for Genome Research, Harvard University

Gene duplication is a major source of evolutionary innovation. We investigate the relative changes in the regulatory elements and coding sequences in pairs of close paralogs in the Saccharomyces cerevisiae genome to uncover how novel genes are incorporated into an existing regulatory network.

H-8

Nucleotide Composition of Ribosomal RNA Genes and Thermal Adaptation

Huaichun Wang[1], Donal Hickey[2]
[1]hcwang@uottawa.ca, University of Ottawa;
[2]dhickey@uottawa.ca, University of Ottawa

We used the method of phylogenetic independent contrasts to reexamined the relationship between optimal growth temperature and the nucleotide content and length of 16S rRNA genes. We conclude that the positive relationship between the nucleotide content of structural RNAs and bacterial growth temperature is not due entirely to phylogenetic history, but reflects a repeated selective response to elevated environmental temperature.

H-9

Adaptive Evolutionary Analysis of Genes in Parasitic Nematode Trapping Fungi

Balaji Rajashekar[1], Anders Tunlid[2]
[1]balaji.rajashekar@mbioekol.lu.se, Lund University;
[2]anders.tunlid@mbioekol.lu.se, Lund University

Comparative genomic analysis of EST data in Monacrosporium haptotylum (nematode-trapping fungi) was performed to identify genes showing positive selection. These genes are specific to Ascomycetes and closely related nematode-trapping fungi. Potential positively selected genes were further analysed using microarray experiments (infection and time-course).

H-10

Correlating Evolutionary Properties of Amino Acid Positions with Structural Data

Roure Béatrice[1], Lartillot Nicolas[2], Philippe Hervé
[1]beatrice.roure@umontreal.ca, Université de Montréal;
[2]nicolas.lartillot@umontreal.ca, Université de Montréal

Most models of sequence evolution assume that sites are identically distributed. With a large data set of bacterial proteins, we study the correlation between structural data and the classes inferred by a Bayesian mixture model that assumes the existence of distinct classes differing by their equilibrium frequencies over the 20 residues.

H-11

Chordate Innovations in Evolution via EST Analysis of Saccoglossus Kowalevskii

Robert M Freeman Jr[1], Marie-Michele Cordonier-Pratt[2], Lee Pratt, Christopher J. Lowe, John Gerhart, and Marc Kirschner
[1]bob_freeman@hms.harvard.edu, Harvard Medical School; [2]mmpratt@uga.edu, University of Georgia

To study the developmental innovations introduced into the chordate phylum after the divergence of the deuterostomes and proteostomes, we have analyzed 70,736 high-quality EST reads from three libraries of the hemichordate Saccoglossus kowalevskii using various computational tools and generated 44,487 unique genes/gene fragments

H-12

Inferring Property Selection Pressure from Psitional Residue Conservation

Rose Hoberman[1], Yong Lu[2], Judith Klein-Seetharaman and Roni Rosenfeld
[1]roseh@cs.cmu.edu, CMU; [2], CMU

We present a statistical procedure for systematically identifying significant conservation of specific physical-chemical properties at each site in a multiple sequence alignment. We use a Monte-Carlo method to ascertain the significance of the findings and to control the False Discovery Rate, then present results for the diverse GPCRA family.

H-13

Finding Phylogenetic Signal in Genomic Kmer Distributions

Elinor K. Karlsson[1], Ted Sharpe[2], Michael C. Zody, Michael Sorenson, Jill P. Mesirov, Eric S. Lander
[1]elinor@bu.edu, Program in Bioinformatics, Boston University; [2]tsharpe@broad.mit.edu, Broad Institute/MIT

Most genomes have a unique and distinguishable kmer distribution (the frequency of each possible oligonucleotide of length k across the genome). We investigate the source of this phenomenon and the feasibility of using the kmer distributions of both mitochondrial and nuclear whole genome sequences to construct accurate phylogenetic trees.

H-14

Classification of BNIP2-Cdc42GAP Homology Domain Family

Keng Hwa Tan[1], Ranganathan Shoba[2], Tin Wee Tan, Boon Chuan Tan
[1]dbstkh@nus.edu.sg, National University of Singapore; [2]shoba@bic.nus.edu.sg, Macquarie University

A novel protein-protein interaction BCH domain is classified as a lipid-binding domain in current protein domain Databases. To understand the presence of a protein-binding domain within an established lipid-binding domain family, we employed bioinformatics analysis methods and classified the BCH domain as a distinct subgroup of the BCH/Sec14-like domain family.

H-15

Molecular Evolution of Biotin-dependent Carboxylases

Wieslawa Mentzen[1], Eve Wurtele[2], Basil Nikolau
[1]wiesia@iastate.edu, Iowa State University; [2]mash@iastate.edu, Iowa State University

Biotin-dependent carboxylases and decarboxylases form a family of metabolically diverse enzymes that transfer CO2 group to or from the substrate via a covalently-bound biotin. The systematic phylogenetic analysis of constituent domains of this family reveals the complex evolutionary history, with multiple domain duplications, fusions, rearrangements, possible fissions and replacements.

H-16

Human Genome Duplications and Reversed Duplications

Heli Hiisila[1]
[1]heli.hiisila@hut.fi, Laboratory of Computer and Information Science, Helsinki University of Technology

Events of duplications or reversed duplications were looked for by going through the human chromosomes piece by piece. Both duplications and reversed duplications found are in average about 30 bp long, but also regions with large-scale rearrangements were found. Duplications are more frequent than reversef duplications.

H-17

Phylogeny of Dinoflagellates - Methodological Considerations

Jarno Tuimala[1]
[1]jarno.tuimala@csc.fi, CSC

Phylogeny of dinoflagellates was inferred from all publicly available 18S rRNA sequences. The effect of sequence alignment parameters and different tree recovery methods on the inferred phylogeny was studied. Tree search strategy combining sectorial searches with tree drifting and fusing proved fastest, but results depended much on the sequence alignment.

H-18

Vestige: Maximum Likelihood Based Phylogenetic Footprinting

Matthew Wakefield[1], Gavin Huttley[2], Peter Maxwell
[1]matthew.wakefield@kangaroo.genome.org.au, CBiS, The Australian National University; [2]gavin.huttley@anu.edu.au, CBiS, The Australian National University

Vestige is a phylogenetic footprinting framework built on the PyEvolve evolutionary modelling toolkit. Vestige uses phylogenetic tree based maximum-likelihood methods to estimate parameters in a model ofmolecular evolution including the conventional evolutionary distance We present analysis of the SCL locus and the BRCA1 Ka/Ks ratio. GPL. Download: http://cbis.anu.edu.au/software/

H-19

The Parkinson's Disease Gene Parkin and its Relatives: Comparative Genomics of the RBR Family

J. Ignasi Lucas[1], Ignacio Marín[2]
[1]Departamento de Genética, Universidad de Valencia. Spain; [2]Ignacio.Marin@uv.es, Departamento de Genética, Universidad de Valencia. Spain

Parkin has an RBR signature, that consists in two RING fingers and an IBR/DRIL domain. A comprehensive analysis of all available sequences of proteins with the RBR signature supports their classification in 12 subfamilies, and offers hints about the functions and evolutionary history of this eukaryotic family of ubiquitin ligases.

H-20

Phylogenetic Reconstruction of Ancestral Gene Expression Profiles by use of Minimum Evolution Methods and Promoter Regions

Roald Rossnes[1], Ingvar Eidhammer[2], David Liberles
[1]roaldr@ii.uib.no, Department of Informatics, University of Bergen; [2]ingvar.eidhammer@ii.uib.no, Department of Informatics, University of Bergen

As genomes evolove under selective pressure, gene regulatory sites effecting expression profiles and splice site usage also evolve. We describe software that produces ancestral states of such continous values and correlate them with an analysis of the actual regulatory sequence to highlight changes that may come from selective pressure.

H-21

Phylogenetic Footprinting by Combining Tree-Gibbs Sampling with Treeclustering

Stefan Van Yper [1], Olivier Thas[2], Jean-Pierre Ottoy and Wim Van Criekinge
[1]Stefan@biomath.UGent.be, Department of Applied Mathematics, Biometrics and Process Control, Ghent University; [2]Olivier.Thas@UGent.be, Department of Applied Mathematics, Biometrics and Process Control, Ghent University

Tree-Gibbs sampling is an extension of the classic Gibbs sampler resulting in an improved detection and alignment of locally conserved regions (motifs) in datasets of promoter sequences from both co-expressed and orthologous genes. Sometimes the Tree-Gibbs algorithm doesn't converge to a valuable consensus root motif model, but this problem is solved by combining the Tree-Gibbs algorithm with treeclustering.

H-22

Mutation Rates in Human and Rodents Are Correlated

Jeffrey Chuang[1], Hao Li[2]
[1]jchuang@genome.ucsf.edu, UCSF; [2]haoli@genome.ucsf.edu, UCSF

The conservation of mutation rates across lineages is a signature of common mutational pressures. We analyze such conservation in human, mouse, and rat. We find that substitution rates in gene silent-sites correlate overall between the species. Certain regions have extremely similar patterns of substitution, covarying over megabase length scales

H-23

Metazoan Deep Phylogenies - can the Cambrian Explosion be Resolved with Molecular Markers?

G. Fritzsch[1], M. Schlegel[2], P. Stadler
[1]Fritzsch@izbi.uni-leipzig.de, IZBI, Leipzig; [2]schlegel@rz.uni-leipzig.de, University of Leipzig

Fossils are our primary window to the history of life. The appearance and radiation within ten million years is referred to as the Cambrian explosion. We started a comprehensive multi-gene sequence analysis in order to contribute to the solution of this outstanding evolutionary phenomenon in metazoan evolution.

H-24

Comparative Genomics of Glutathione Transferases

Antonio Marco[1], Ignacio Marín[2]
[1]Departamento de Genética, Universidad de Valencia. Spain; [2]Ignacio.Marin@uv.es, Departamento de Genética, Universidad de Valencia. Spain

Glutathione transferases (GSTs) are a highly heterogeneous group of detoxifying enzymes. In a phylogenetic analysis that involved 700 sequences, at least fourteen main groups of GSTs were found. A novel group includes the human gene GDAP1, involved in Charcot-Marie-Tooth disease. GDAP1 group proteins have unique structural features, including transmembrane domains.

H-25

Comparative Genomics and Proteomics of the Septin Gene Family

Sergio Mars[1], Vicente Arnau[2], Ignacio Marín[3]
[1]Departamento de Genética, Universidad de Valencia. Spain; [2]Departamento de Informática, Universidad de Valencia. Spain; [3]Ignacio.Marin@uv.es, Departamento de Genética, Universidad de Valencia. Spain

We describe the phylogenetic relationships among septins, a group of proteins involved in cytokinesis and other processes. A comparative analysis of septin-interacting proteins in Saccharomyces cerevisiae and Drosophila melanogaster that suggests that septins are functioning differently in animals and fungi, is also presented.

H-26

Reconstructing Longitudinal Evolutional Phylogenetic Tree of HIV-1 Protease Gene under HAART by Linking Within-patient Orthologous Viral Population

Naoki Hasegawa[1], Wataru Sugiura[2], F. Ren, M. Matsuda, H. Tanaka
[1]hasegawa.bio@tmd.ac.jp, Department of Bioinformatics, Medical Research Institute, Tokyo Medical and Dental University; [2]wsugiura@nih.go.jp, AIDS Research Center, National Institute of Infectious Diseases

Understanding of the mechanism of drug resistant mutation acquisition is important for anti-HIV treatment. We reconstructed a longitudinal phylogenetic tree of HIV-1 protease gene by linking within-patient orthologous viral population. It clarified not only the evolutionary relationship, but also the dynamic change of the viral population.

H-27

Phylogeny of the Archaea Using Combined Protein Datasets

Chris Cooper[1], Sandra L. Baldauf[2], Dominic P. Tolle, Katherine Tindall, Jill Cockrill
[1]christopher.cooper@ndcls.ox.ac.uk, University of Oxford; [2]slb14@york.ac.uk , University of York

The phylogeny of nineteen archaeal species was examined using a concatenated orthologous protein data set. Parsimony and distance analyses were performed on individual proteins and also random combinations of protein and character subsets. The significance of clade arrangement and resolution are discussed in the context of a concatenated dataset approach.

H-28

Evidence of Positive Darwinian Selection in Putative Meningococcal Vaccine Antigens

David Fitzpatrick[1], Chris Creevey[2], James McInerney
[1]david.a.fitzpatrick@may.ie, NUI Maynooth; [2], NUI Maynooth

Serogroup B Meningococcal meningitides is a life threatening disease with no effective vaccine. Highly conserved membrane proteins are potential vaccine antigens and presently a number are in clinical trials. This analysis has found evidence of positive Darwinian selection in a number of these putative vaccine antigens, this may have important implications for the effectiveness of these vaccines.

H-29

Protein Evolution, the Younger the Faster?

Hannes Luz[1], Eike Staub[2], Antje Krause, Martin Vingron
[1]luz@molgen.mpg.de, MPI for Molecular Genetics Berlin; [2]staub@molgen.mpg.de, MPI for Molecular Genetics Berlin

It is observed that secreted proteins evolove relatively fast. Given a set of 3640 protein families each containing ortholog of man, fugu, fly and worm, we investigate the hypothesis that rates of protein evolution depend on the time having passed since the protein's evolutionary emergence.

H-30

Selective Forces Driving Guanine + Citosine (GC%) Content in Prokaryotes

Héctor Romero[1], Hugo Naya[2], Alejandro Zavala, Fernando Álvarez, Giorgio Bernardi and Héctor Musto
[1]eletor@fcien.edu.uy, Lab. Org. Evol. del Genoma - Facultad de Ciencias - Universidad de la República; [2]hnaya@fcien.edu.uy, Lab. Org. Evol. del Genoma - Facultad de Ciencias - Universidad de la República

In prokaryotes, the range of GC% levels (from 25% to 75%) have puzzled biologists for more than 4 decades. This variability can be seen under two standpoints: selectionist or neutralist. Here we analyze from a novel point of view two selective forces that probably shape GC%: aerobiosis and Optimal Growth Temperature. The general implications of these findings are discussed.

H-31

Comparative Analysis of Ten Animal Genomes Reveals Recent Expansions of an Ancestral Set of Ion Channel Genes

Christian M Zmasek[1], Tim Jegla[2], Serge Batalov, Brice Campo, Ardem Patapoutian
[1]czmasek@gnf.org, Genomics Institute of the Novartis Research Foundation; [2]tjegla@gnf.org, Genomics Institute of the Novartis Research Foundation

We present a detailed phylogenomic analysis of 33 ion channel gene families across ten completely sequenced animal genomes. In particular, we analyzed where and when gene family producing gene duplications are likely to have occurred and what inferences we can make regarding ancestral channel types

H-32

Inferring and Visualising Intraspecific Phylogenies Using a Minimum Geodetic Set Cover Approximation

S. C. Dyer[1], B. J. Stapley[2]
[1]sarah.dyer@postgrad.umist.ac.uk, UMIST; [2]b.stapley@umist.ac.uk, UMIST

Undirected graphs are constructed to visualise the evolutionary history of sequences which have evolved with frequent recombination. This method generates a median closure (Bandelt et al. 1995) - incorporating heteroplasies - for a set of sequences. The network is reduced by retaining most probable paths from geodetic sets between sequence pairs

H-33

Variable Inter-Subtype Evolutionary Dynamics within HIV-1 Group M

Simon A.A. Travers[1], Mary J. O'Connell[2], James O. McInerney, Grace P. McCormack
[1]simon.a.travers@may.ie, Bioinformatics Laboratory, NUI Maynooth; [2]mary.oconnell@dcu.ie, School of Biotechnology, Dublin City University

HIV-1 group M comprises of a number of phylogenetically distinct subtypes all of which are believed to have arisen from a founder virus. We have analysed the evolutionary processes occurring across the subtype lineages and have observed subtype specific evolution occurring within HIV-1 group M.

H-34

Maternally and Paternally Silenced Imprinted Genes Differ in their Intron Content

Marie E Fahey[1], Desmond G Higgins[2], Walter Mills, Tom F Moore
[1]marie.fahey@ucd.ie, University College Dublin; [2]des.higgins@ucd.ie, University College Dublin

Imprinted genes exhibit silencing of one of the parental alleles during embryonic development. We investigate intron content of maternally and paternally silenced imprinted genes, in both mouse and human, relative to a non-imprinted control set. We find that there are significant differences with respect to a variety of intron-related parameters.

H-36

Taxonomy, Biology's First Ontology, and the Tree of Life, Biology's Grandest Endeavour

Nadia Anwar[1]
[1]n.anwar@udcf.gla.ac.uk, Univeristy of Glasgow

This project will develop methodologies and tools to support phylogenetic Databases. Using a web-based portal to taxonomy data a user can search for scientific names. The results will expand to include the terms entered by the user, synonyms and equivalent terms. Exemplified using phylogenetic tree database, users will be able to retrieve all trees required for a super tree analysis.

H-37

A Simple Algorithm to Automatically Detect Recombinants in DNA/RNA Sequence Alignments

Iain Milne[1], Frank Wright[2]
[1]iainm@bioss.ac.uk, BioSS; [2]frank@bioss.ac.uk, BioSS

The TOPALi software provides an intuitive GUI for working with statistical methods designed to detect recombination within DNA/RNA sequence alignments. Here we describe the algorithm used to automatically predict recombinant sequences using a parametric bootstrapping approach, resulting in a graphical display showing the similarity of putative recombinant sequences to non-recombinants

H-38

Puzzling: Supertree Construction Software

Melissa Pentony[1], James O McInerney[2]
[1]Melissa.M.Pentony@may.ie, NUI Maynooth; [2]James.O.McInerney@may.ie, NUI Maynooth

Supertree methods are becoming an increasingly popular way of constructing phylogenies. We present a new method, Puzzling, for constructing robust supertrees using either sequence alignments or nested parenthesis trees as input. Puzzling is freely available at http://bioinf.may.ie/software/puzzling

H-39

everEST - A Database Package for Phylogenomic Analysis of EST Sequences

Steinke, Dirk[1], Salzburger, Walter[2], Axel Meyer
[1]Dirk.Steinke@uni-konstanz.de, Uni Konstanz; [2]Walter.Salzburger@uni-konstanz.de, Uni Konstanz

everEST is a EST database software which includes software features for genomic comparisons by BLAST searches against common available Databases and doing phylogenomic analysis.

H-40

Comparative Analysis of Protein Coding Sequences from Human, Mouse and the Domesticated Pig

Frank Grønlund Jørgensen[1], Asger Hobolth[2], Henrik Hornshøj, Christian Bendixen, Merete Fredholm and Mikkel H. Schierup
[1]frank@birc.dk, Dept. Genetics and Ecology, Univ. Aarhus, Denmark; [2]asger@birc.dk, BiRC, Univ. Aarhus, Denmark

Evolutionary analysis of 1120 full lengths cDNA sequences from pig, human and mouse was performed to estimate evolutionary rates and identiofy genes under positive selection.

H-41

Eukaryotic Phylogenetic Model Based on Mitochondrial Genome Evolution

Mikyung Jang[1]
[1]jmk@mrc-dunn.cam.ac.uk, MRC Dunn Human Nutrition Unit

Mitochondria are thought to originate from a common alpha-proteobacterial ancestor. The proto-mitochondrial genome is thought to have been mostly deleted and/or transferred to the nuclear genome during evolution. This study constructed phylogenic trees of 26 eukaryotic phyla based on the presence, localisation and sequence of genes in the mitochondrial genomes.

H-42

Gene Subfunctionalization and Neofunctionalization in Arabidopsis Gene Families Studied Using Gene Expression and Sequence Data

Kiana Toufighi[1], Nicholas Provart[2], David Guttman
[1]kiana.toufighi@utoronto.ca, Dept. of Botany, Uni. Toronto; [2]provart@botany.utoronto.ca, Dept. of Botany, Uni. Toronto

Gene expression data for 3377 gene families in Arabidopsis thaliana were used along with sequence data to explore the fate of duplicated genes. Almost all duplicated genes exhibited low expression correlation, and Kn/Ks ratios were << 1, suggesting that subfunctionalization, and not neofunctionalization, is the fate of gene duplicates.

H-43

Investigation Into an Ant-Colony Optimisation (ACO) Approach to Phylogenetic Tree Optimisation

Alexander Moore[1], Dr. Andrew Dalby[2]
[1]a.d.moore@ex.ac.uk, Exeter University Bioinformatics Centre; [2]a.r.dalby@ex.ac.uk, Exeter University Bioinformatics Centre

We present a novel application of the Ant-Colony Optimisation algorithm to Maximum-Likelihood phylogenetic tree construction. This poster details current results as well as future research directions.

H-44

Phylogenetic Super-Networks from Partial Trees

Tobias Dezulian[1], Daniel H. Huson[2], Tobias Kloepper, Mike A. Steel
[1]dezulian@informatik.uni-tuebingen.de, Center for Bioinformatics (ZBIT), Tuebingen University; [2]huson@informatik.uni-tuebingen.de, Center for Bioinformatics (ZBIT), Tuebingen University

Incomplete, possibly contradictory phylogenetic data, such as a set of partial trees derived from different markers challenges a phylogenetic analysis. We present an efficient approach to infer a phylogenetic super-network from such data. A biological application example illustrates the usefulness of the method and an experimental study confirms its potential.

H-45

Protein Evolution: Universal Sharing Patterns

Gustavo Caetano-Anolles[1]
[1]gca@uiuc.edu, University of Illinois at Urbana-Champaign

The evolutionary tracing of how protein architecture crosses organismal domain boundaries shows that many ancestral folds were once common to all life and that defined episodes of architectural diversification associated with Archaea, Bacteria and Eucarya occcurred relatively late in protein evolution and in defined order.

H-46

Detecting Coevolving Protein Residues Using Bayesian Phylogenetic Techniques

Matt Dimmic[1], Melissa Todd[2], Carlos D. Bustamante, Rasmus Nielsen
[1]mdimmic@umich.edu, Cornell University; [2]melissa@mail.bscb.cornell.edu, Cornell University

We describe a Bayesian approach to detect correlated substitutions in protein families. This method minimizes errors due to phylogenetic correlations and rate variation, and Bayesian posterior predictive distributions are used to determine statistical significance. Sensitivity and specificity are assessed on simulated datasets, and we examine the coevolutionary signal between interacting protein domains.

H-47

Divergent Transcriptional and Translational Signals in Archaea

Elfar Torarinsson[1], Hans-Peter Klenk[2], Roger A. Garrett
[1]elfar@binf.ku.dk, University of Copenhagen; [2]Hans-Peter.Klenk@egene-biotech.de, e.gene Biotechnologie GmbH

Many archaea, in contrast to bacteria, produce a high proportion of leaderless transcripts, show a wide variation in their consensus Shine-Dalgarno sequences, and frequently use GUG and UUG start codons. In order to understand the basis for these differences, eighteen archaeal genomes were examined for sequence signals that are positionally conserved upstream from genes.

H-48

AriadneDB: an Application for Automatically Storing and Constructing Sequence Similarity Sets (S3) used in the Construction of Phylogenetic Trees

Derek Huntley[1], Nadia Anwar[2], Kosmas Theodorides
[1]d.huntley@imperial.ac.uk, Imperial College London; [2]n.anwar@udcf.gla.ac.uk, University of Glasgow

AriadneDB is an automated system for generating sequence alignments from ESTs to provide a data source for phylogeny reconstruction. The system has a GUI for user interaction to customise parameters, build and edit sequence alignments and view with CLUSTALX.

H-49

Ectopic Gene Conversions in Pathogenic and Non-pathogenic E. Coli Genomes

Robert Morris[1], Guy Drouin[2]
[1]rmorr329@science.uottawa.ca, University of Ottawa; [2]gdrouin@science.uottawa.ca, University of Ottawa

We characterized the ectopic gene conversions in the genomes of the K-12 MG1655, O157:H7 Sakai, O157:H7 EDL933, and CFT073 strains of E. coli. We found that recombination in pathogenic strains is more frequent and has less stringent requirements than in non-pathogenic strains, perhaps due to mutations in mismatch repair genes.

H-50

Microsatellite Markers for Genetic Typing of Standard Rat Strains

B.Kiran Kumar[1], Dr.N.Vijaya Bhanu[2], Dr.N.V.Giridharan, National Institute of Nutrition, Email:nappanveettil@yahoo.co.in
[1]sobian_007@rediffmail.com, National Institute of Nutrition; [2]nbhanuin@yahoo.com, National Institute of Nutrition

In an effort to genetically type three commonly used rat strains, WNIN, WKY, and Fischer-344, we are using the sequence tagged microsatellite markers(STMS). Out of the 60STMSs tested, 7 are polymorphic and are useful for routine genetic monitoring of inbred rat strains and provide a valuable tool to study phylogenetic relationship among them.

H-51

Shuffling of Sulfolobus Genomes by Autonomous and Non-autonomous Mobile Elements

Elfar Torarinsson[1], Kim Brügger[2]
[1]elfar@mermaid.molbio.ku.dk, University of Copenhagen; [2]brugger@mermaid.molbio.ku.dk, University of Copenhagen

Each of the sequenced Sulfolobus genomes contains large numbers of mobile elements. Moreover, the gene order between the two organisms differs greatly. Recently, a third Sulfolobus genome was completed which contains no mobile elements. Comparison of the gene orders of the three genomes provide evidence for mobile element-induced rearrangements.

H-53

Approximative Estimation in Codon-based Models with Context Dependent Substitution Rates

Ole F. Christensen[1], Asger Hobolth[2], Jens Ledet Jensen
[1]olefc@birc.dk, Bioinformatics Research Center Aarhus University, Denmark; [2]asger@birc.dk, Bioinformatics Research Center Aarhus University, Denmark

We introduce pseudo-likelihood type of approximative estimation methods for a codon model allowing CpG effects over codon boundaries. This makes the model much more attractive to use for data analysis as a supplementary to the commonly used codon models implemented in software package PAML

H-54

Pruning the Probabilistic Divergence Measure for Improve Detection of Interspecific Recombination

Dirk Husmeier[1], Frank Wright[2], Iain Milne
[1]dirk@bioss.ac.uk, Biomathematics & Statistics Scotland; [2], Biomathematics & Statistics Scotland

The present study investigates how the shortcomings of the probabilistic divergence measure (PDM) methods for detecting interspecific recombination can be redeemed with a topology-based pruning scheme.

H-55

The Real Rates of RNA Evolution under Selecgtion

Janet Siefert[1], George Fox[2], Chad Shaw, Marek Kimmel
[1]siefert@rice.edu, Rice University; [2]fox@uh.edu, University of Houston

We report the population structure and evolutionary history of laboratory evolved ribozymes. This analysis provides insight into the possible diversity and amount of time necessary to populate an RNA world with the requisite number of molecules at the boundry of a chemical/pre-biological system.

H-56

The Effects of Evolutionary Distance on Comparative Annotation of Transctiption Factor Binding Sites in S. Cerevisiae

Alan Moses[1], Derek Chiang[2], Dan Pollard, Michael Eisen
[1]amoses@lbl.gov, Graduate Group in Biophysics, Center for Integrative Genomics, UC Berkeley, Berkeley, CA, USA; [2]dchiang@ocf.berkeley.edu, Department of Molecular and Cell Biology, UC Berkeley, Berkeley, CA, USA

Comparative genomics stands to improve computational annotation of functional non-coding sequence elements. Using a new software tool ("monkey"), we search alignments of non-coding sequence for instances of known transcription factor binding motifs. By comparing the results to genome-wide functional data, we show that as evolutionary distance increases, our predictive power increases.

H-57

Regional and Time-resolved Base Substitution Patterns of the Human Genome

Peter F Arndt[1]
[1]arndt@molgen.mpg.de, Max Planck Institute for Molecular Genetics

Utilizing various repetitive elements we map regional differences in nucleotide substitution patterns along the human chromosomes on the 1Mbp scale. Comparing differently old repeats it is also possible to reconstruct the history of those substitution patterns. Special attention is given to the CpG methylation deamination process, which is the predominant substitution process for vertebrates.

H-58

An SVD-based Comparison of Nine Whole Eukaryotic Genomes Supports a Coelomate rather than Ecdysozoan Lineage

Gary W. Stuart[1], Michael W. Berry[2]
[1]G-Stuart@indstate.edu, Indiana State University; [2]berry@cs.utk.edu, Univerisity of Tennessee Knoxville

A non-alignment method based on Singular Value Decomposition (SVD) was used to compare the predicted proteins from nine whole eukaryotic genomes. This analysis resulted in the identification of over 400 well conserved motifs and gene families, and produced a species tree supporting one of two conflicting hypotheses of metazoan relationships.

H-59

Large-scale Evaluation and Prediction of Function Shift in Protein Families

Saraswathi Abhiman[1], Erik L.L. Sonnhammer[2]
[1]abhiman.saraswathi@cgb.ki.se, Center for Genomics and Bioinformatics, Karolinska Institute, Stockholm, Sweden; [2]Erik.Sonnhammer@cgb.ki.se, Center for Genomics and Bioinformatics, Karolinska Institute, Stockholm, Sweden

Evaluation and prediction of function changes was performed on a large dataset of enzyme protein families derived from the Pfam database. We present an approach for analyzing subfamilies of a protein family using two kinds of sites namely Conservation shifting and Rate shifting sites. Using this approach we predicted cases of functional shifts in the entire Pfam database.

H-60

First Comparative Genomics Analyses Supporting the Ecdysozoa (Arthropods-Nematodes) Monophyly

Hernán J. Dopazo[1], Joaquín Dopazo[2]
[1]hdopazo@cnio.es, Bioinformatics Unit. Centro Nacional de Investigaciones Oncologicas. CNIO; [2]jdopazo@cnio.es, Bioinformatics Unit. Centro Nacional de Investigaciones Oncologicas. CNIO

There are many reasons to consider the coelomata-ecdysozoa problem, the most astonishing issue in animal systematics and one of the major open ended subjects in evolutionary biology. Previous single gene and genome scale analyses contradicted each other. We present first genome scale evidence supporting the ecdysozoa hypothesis.

H-61

MetaPIGA2.0, a Software Implementing the Metapopulation Genetic Algorithm for very Large Phylogeny Inference under Maximu Likelihood

Michel C. Milinkovitch[1], Alan Lemmon[2], Pascal Dal Farra
[1]mcmilink@ulb.ac.be, Free University of Brussels; [2]alemmon@evotutor.org, Biology, University of Texas, Austin

This new, cross-platform, version of MetaPIGA incorporates multiple ML models of nucleotide and amino-acid evolution. It allows rapid exploration of search space, identifies optimal trees and produces branch probability indices. Single-processor and parallelized versions of METAPIGA2.0 (distributed at http://dbm.ulb.ac.be/ueg) are available.

H-62

Do Alpha Proteobacteria have Composite Genomes?

Zara Ghazoui[1], Peter Young[2]
[1]zg105@york.ac.uk, University of York; [2]jpy1@york.ac.uk, University of York

Orthologous proteins found in all species include both housekeeping (basic) and accessory genes. Basic genes are strongly overrepresented on the main chromosome, but accessory genes are mostly on secondary chromosomes or plasmids. Most orthologs, especially on the chromosome, support the ribosomal RNA phylogeny, but an interesting minority support other trees.

H-63

When and Where Have Protein Folds Emerged: An Evolutionary View

Song Yang[1], Phil Bourne[2]
[1]soyang@sdsc.edu, Department of Chemistry and Biochemistry, University of California, San Diego; [2]bourne@sdsc.edu, San Diego Supercomputer Center, University of California, San Diego

We analyze the protein domain distribution across 123 Archaea, Bacteria and Eukaryota with complete genomes using SUPERFAMILY database. Using the venn diagram, we investigated questions such as the evolutionary origin of the three kingdoms, the disulfide-bond dependent domains and the two unique domains in Archaea.

H-64

Variation in Base Composition across Sites and its Effect in Phylogenetic Inference

Vivek Gowri-Shankar[1], Magnus Rattray[2]
[1]vivek.gowri-shankar@cs.man.ac.uk, University of Manchester; [2]magnus.rattray@cs.man.ac.uk, University of Manchester

Nucleotide composition in RNA coding-genes varies across sites and fast evolving sites in RNA helices are relatively G+C poor. The usual assumption of compositional homogeneity across sites in phylogenetics is therefore violated. We present some effects of this model misspecification in phylogenetic inference and question the ancestral G+C estimate found previously through use of a nonhomogeneous model.

H-65

Benchmarking the Detection and Analysis of Recombination

Cheong Xin Chan[1], Robert G. Beiko[2], Mark A. Ragan
[1]*c.chan@imb.uq.edu.au, Institute for Molecular Bioscience and ARC Centre in Bioinformatics;* [2]*r.beiko@imb.uq.edu.au, Institute for Molecular Bioscience and ARC Centre in Bioinformatics*

Simulating genealogy with recombination under a coalescent model backward in time, followed by simulating sequence along the model forward in time, is a realistic approach for simulating sequence evolution. These simulations are used to benchmark algorithms designed to detect and analyse recombination.

H-66

A hidden Markov Model of Evolution and Structure for Multiple Sequence Alignment, with an Application to Phylogenetic Gene Finding

Ari Loytynoja[1], Nick Goldman[2]
[1]*ari@ebi.ac.uk, EMBL-European Bioinformatics Institute;* [2]*goldman@ebi.ac.uk, EMBL-European Bioinformatics Institute*

A new method that simultaneously aligns sequences and infers their internal structure is presented, and its application to phylogenetic gene finding is described. The method combines a HMM defining the match/gap-process and sequence structure, progressive algorithms for reconstructing ancestral sequences, and time-dependent insertion-deletion and substitution processes in different structure states.

H-67

Using Fish Phylogenetic Footprinting to Identify Vertebrate Specific Regulatory Regions

Adam Woolfe[1], Greg Elgar[2], Martin Goodson, Debbie Goode, Phil Snell, Sarah F. Smith, Tanya Vavouri, Gayle McEwen, Krys Kelly
[1]*awoolfe@rfcgr.mrc.ac.uk, MRC RFCGR;* [2]*gelgar@rfcgr.mrc.ac.uk, MRC RFCGR*

Conserved non-coding sequences identified in alignments between the teleost fish Fugu rubripes and Mammals are good indicators of regulatory elements and cluster almost exclusively around developmental genes. Here we have identified a number of these in multiple alignments. These sequences are likely to comprise the fundamental set of regulatory elements for vertebrate development.

H-68

Losses of Universally Conserved Genes in our Genome

Etienne Danchin[1], Pierre Pontarotti[2]
[1]*edanchin@up.univ-mrs.fr, Université de Provence;* [2]*Pierre.Pontarotti@up.univ-mrs.fr, CNRS / Université de Provence*

We report here 24 gene families that are evolutionary conserved throughout evolution (from yeast as Saccharomyces to animals such as Drosophila) and are absent from the human genome. Most of these ancestrally conserved genes whose function we describe in details, were lost in the whole vertebrates' lineage.

H-69

A Sequence Sub-sampling Algorithm (Ali-shuffle) to Increase the Power to Detect Distant Homologues

Kate Johnston[1], Denis Shields[2]
[1]*kjohnston@rcsi.ie, RCSI;* [2]*dshields@rcsi.ie, RCSI*

Investigation of protein alignments (PFAM) revealed there is often a weak relationship between the inferred tree and the pair-wise distances among proteins. Our algorithm (ali-shuffle) allows for similarities among sequence subsets that do not adhere to phylogeny. Ali-shuffle has greater sensitivity than HMMER and PSI-BLAST higher permitted false positive rates.

H-70

Detection of Horizontal Gene Transfer in Rumen Ciliates

G.Ricard[1], M.Huynen[2]
[1]*G.Ricard@cmbi.kun.nl, CMBI / NCMLS;* [2]*M.Huynen@cmbi.kun.nl, CMBI / NCMLS*

To study horizontal gene transfer (HGT) from bacteria to eukaryotes, we analyzed ESTs sequences from ciliates. Rumen ciliates were chosen as they live in close contact with bacteria in the gut of ruminants. Using phylogenetic approaches we found 88 ciliates genes (representing ~20 different functions) acquired by HGT from bacteria.

H-71

The Phylogenetic Relationship between Human, Fruitfly and C. Elegans

Avril Coghlan[1], Des Higgins[2]
[1]avril.coghlan@ucd.ie, University College Dublin;
[2]des.higgins@ucd.ie, University College Dublin

It is very controversial whether flies are more closely related to humans or to nematodes. To avoid the long-branch attraction artefacts to which previous studies have been prone, we analysed 64 genes using sponge/jellyfish outgroups (instead of yeast/plant) and sequence from the slow-evolving nematode Brugia malayi (instead of C. elegans).

H-72

Birth, Sex and Death of Genes in the Actinobacteria: a study of how Gene Duplication, Gene Loss and Horizontal Gene Transfer have Shaped the Evolution of these Organisms

Rhoda J. Kinsella[1], James O. McInerney[2], none
[1]Rhoda.J.Kinsella@may.ie, National University of Ireland Maynooth; [2]James.O.McInerney@may.ie, National University of Ireland Maynooth

The phylum Actinobacteria contains bacteria whose genomes vary considerably in size and range from species that are pathogenic to humans and animals to species that are used by the biotechnology industry for human benefit. This research aims to identify the evolutionary events that have shaped the extant Actinobacteria.

H-73

Phylogenetic Analysis of the Yeast Peroxisomal Proteome

Frank van Zimmeren[1], Toni Gabaldon [2], Martijn A. Huynen
[1]fvzimmer@cmbi.kun.nl, CMBI;
[2]T.GAbaldon@cmbi.kun.nl, CMBI-NCMLS

Phylogenetic analyses of 58 yeast peroxisomal proteins revealed that half are of eukaryotic origin, most of which are involved in peroxisome organization and biogenesis. In contrast, metabolic pathway proteins tend to have bacterial origin. Proteins with similar phylogenetic profiles and genes with a significant level of co-expression also showed functional correlation.

H-74

RAMBLE: The Adaptive Evolution Simulator

Jennifer M. Commins[1], Dr. James O. McInerney[2]
[1]jennifer.commins@may.ie, NUI Maynooth;
[2]james.o.mcinerney@may.ie, NUI Maynooth

Software for simulating sequence evolution has been in existence for quite some time. Current simulation programs do not attempt to simulate selection and instead concentrate on the simulation of mutation. Our aim is to address this issue and accurately simulate the processes of both adaptive and neutral evolution over time.

H-75

GASP: Gapped Ancestral Sequence Prediction of Proteins

Richard J Edwards[1], Denis C Shields[2]
[1]redwards@rcsi.ie, RCSI; [2]dshields@rcsi.ie, RCSI

GASP (Gapped Ancestral Sequence Prediction) is a new algorithm to predict ancestral sequences from multiple protein alignments of any size. These alignments may include gaps. GASP was tested on simulated datasets, based on real phylogenies. Performance was slightly inferior to CODEML for ungapped data but considerably better for gapped data.

H-76

Evolution of the Mitochondrial Metabolism

Toni Gabaldon [1], Martijn Huynen[2]
[1]T.Gabaldon@cmbi.kun.nl, CMBI-NCMLS;
[2]huynen@cmbi.kun.nl, CMBI-NCMLS

Using large-scale phylogenetic analyses we reconstruct the proto-mitochondrial proteome and compare its metabolic diversity with that of two experimentally determined sets of human and yeast mitochondrial proteomes. The analysis reveals that the transition from endosymbiont to modern organelle was accompanied by major changes in the proteome and its metabolic capability.

H-77

PDA: a Pipeline to Explore and Estimate Polymorphism in Large DNA Databases

Sònia Casillas[1], Antonio Barbadilla[2]
[1]Sonia.Casillas@uab.es, Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Catalonia, Spain.; [2]Antonio.Barbadilla@uab.es, Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Catalonia, Spain

PDA is a Web server that automatically can search for polymorphic sequences in large Databases and estimate their genetic diversity on different functional regions. The output includes a database with available sequences and estimations, the performed alignments and a histogram maker tool. PDA is publicly available at http://pda.uab.es/

H-78

Identification of Chromosomal Evolutionary Breakpoints in Vertebrate Genomes

Megy[1], Rocha[2]
[1]km369@cam.ac.uk, University of Cambridge; [2]dr219@mole.bio.cam.ac.uk, University of Cambridge

We propose a method to identify chromosomal evolutionary breakpoints given an annotate reference genome and a genome of interest The latter can be represented by a full sequence or a set of clones allowing the analysis of partially sequenced genomes

H-79

A Common Evolutionary Origin for Nuclear Pore Complexes and Coated Vesicles

Damien Devos[1], Michael P. Rout[2], Svetlana Dokudovskaya, Frank Alber, Marc A. Marti-Renom, Rosemary Williams, Brian T. Chait, Andrej Sali
[1]damien@salilab.org, UCSF; [2]rout@mail.rockefeller.edu, Rockefeller U., NY

We report a structural analysis of the seven proteins that comprise the yNup84/vNup107 subcomplex. Our results suggest a common evolutionary origin for nuclear pore complexes and coated vesicles in an early membrane-curving module that led to the formation of the internal membrane systems in modern eukaryotes.

H-80

Signatures of Selection at Molecular Level in Two Genes Implicated in Human Familial Cancers

Krzysztof A. Cyran[1], Joanna Polanska[2], Ranajit Chakraborty, David Nelson, Marek Kimmel
[1]chrisc1@rice.edu, Statistics Department, Rice University, Houston TX, USA; [2]jkp@stat.rice.edu, Institute of Automation, Silesian University of Technology, Gliwice, Poland

We tested for natural selection SNPs haplotypes from four genes implicated in human familial cancers: ATM, RECQL, BLM and WRN. Using standard tests and original graphical method we detected selection in ATM and RECQL. To evaluate the influence of different demographic scenarios we employed various null hypotheses, assuming: constancy, growth and substructure of population.

H-81

Deterministic Model of Evolution from the Point of View of Control Theory

Alexey V. Melkikh[1]
[1]mav@dpt.ustu.ru, Ural State Technical University, Ekaterinburg, Russia

The deterministic model of evolution allowed to remove the contradiction on characteristic times of species origin is constructed. It is shown, that in the case of uncertainty of an environment movement of an organism to the next ecological niches will be slowed down. Such approach pulls together deterministic theory of evolution with morphogenesis.

H-82

On Protein Sequence, Structure and Evolution: What can integration of the Databases reveal?

David John Howorth[1], Antonina Andreeva[2]
[1]dhoworth@mrc-lmb.cam.ac.uk, MRC Centre for Protein Engineering Hills Road, Cambridge; [2]MRC Centre for Protein Engineering Hills Road, Cambridge

Recent structural and sequence information has revealed protein relationships that are not obvious in major Databases. As part of the eFamily integration project, we addressed this problem by comparing PROSITE, Pfam, CATH and SCOP. We present a detailed statistical analysis and discuss some examples of non-trivial protein relationships.

I-1

Rapid Identification of Ligand Binding Sites

Randy J. Zauhar[1], Michael F. Bruist[2]
[1]r.zauhar@usip.edu, University of the Sciences in Philadelphia; [2]m.bruist@usip.edu, University of the Sciences in Philadelphia

We present new approaches to the computational identification of potential ligand binding sites in protein structures. Our methods involve modification and extension of the Shape Signatures technique, which has shown success in finding matches between known receptor sites and compounds in chemical Databases

I-2

Computational Studies of the Thioredoxin Superfamily

Efrosini Moutevelis[1], Jim Warwicker[2]
[1]E.Moutevelis@postgrad.umist.ac.uk, UMIST; [2]j.warwicker@umist.ac.uk, UMIST

We present a method for the prediction of pKa and redox potentials in the thioredoxin superfamily. This method combines electrostatic calculations using the Finite Difference Poisson-Boltzmann method with sidechain rotamer variation for the CXXC motif of this superfamily and a clustering method that uses matrices of distances for speed.

I-3

DAROGAN: Enzyme Function Prediction from Multiple Sequence Alignments

Russell S. Hamilton[1], Dietlind L. Gerloff[2]
[1]Russell.Hamilton@ed.ac.uk, University of Edinburgh; [2]D.Gerloff@ed.ac.uk, University of Edinburgh

DAROGAN is a novel, proof of principle, enzyme function prediction method. Unordered sets of conserved functional residues (Treads) are encoded as vectors and describe enzymes of known function. A database of these reference Treads is used to infer functional relatedness from similarity to a query Tread for a putative enzyme.

I-4

Computational Design of Antimicrobial Peptides

Kyle Jensen[1], Mark Sty[2], Gregory Stephanopoulos
[1]kljensen@mit.edu, MIT; [2]marksty@mit.edu, MIT

Antimicrobial peptides (AmPs) are effectors of the innate immune system that combat bacterial infection in multicellular eukaryotes. Here we describe the computation design of novel AmPs that have no homology to natural AmPs but show strong bacteriostatic activity against several species of bacteria

I-5

Symmetric Scores for Discriminating Regulatory Elements

James Taylor[1], Webb Miller[2], Francesca Chiaromonte
*[1]james@bx.psu.edu, Center for Comparative Genomics and Bioinformatics, Penn State University;
[2]Center for Comparative Genomics and Bioinformatics, Penn State University*

We improve Regulatory Potential scores by constraining model selection to be symmetric with respect to the species aligned. These scores discriminate regulatory elements in multiple species alignments, using a mixture of conservation, composition, and short patterns, based on training with known regulatory elements and a neutral model.

I-6

Docking Ligands to Flexible Protein Receptors Using a New Scoring Function Based on Statistical Potentials

Ernesto Moreno[1], Luis A. Diago[2]
[1]emoreno@ict.cim.sld.cu, Center of Molecular Immunology; [2]diago@electrica.cujae.edu.cu, Instituto Superior Politecnico

The flexibility of a protein binding site, as explored by molecular dynamics, is taken into account in docking simulations by using a new scoring function based on statistical potentials. Tests performed for hundreds of complexes show that this approach is suitable for virtual screening of small-molecule Databases.

I-7

Using Structural Sequence Information to Predict Helix-turn-helix DNA Binding Motifs in Proteins of Unknown Function

Marialuisa Pellegrini-Calace[1], Janet M. Thornton
[1]marial@ebi.ac.uk, EBI

A study of the potentiality of transferring structural knowledge back to the sequence level in the prediction of DNA-binding helix-turn-helix (HTH) motifs is shown. The study was carried out by means of HMMs derived from both PFAM multiple alignments and structure-feature-based sequence alignments.

I-8

Prediction of G-Protein Coupling Specificity of GPCR

Betty Yee Man Cheng[1], Jaime G. Carbonell[2], Judith Klein-Seetharaman
[1]ymcheng@cs.cmu.edu, Carnegie Mellon University; [2]jgc@cs.cmu.edu, Carnegie Mellon University

Understanding the signalling mechanism of G-protein coupled receptors requires knowledge of the receptors' G-protein coupling specificity. Using the k-NN classifier on alignment-based and ngram-based features from the whole sequence, we have developed a coupling specificity prediction method that outperforms the current state-of-the-art in precision, recall and F1.

I-9

Exploring the Aggregation Mechanisms of Amyloidic Tetramers

Geneviève Boucher[1], Adrien Melquiond[2], Normand Mousseau, Philippe Derreumaux
[1]genevieve.boucher@umontreal.ca, Université de Montréal; [2], Institut de Biologie Physico-Chimique

Recent results suggest that the toxicity associated with _-amyloid fibrils is caused by the soluble oligomers observed at the onset of fibrillogenesis, raising the interest in determining the initial events in the aggregation process. We present a numerical study of the aggregation mechanisms of KFFE tetramers using the activation-relaxation technique with an approximate energy model (OPEP).

I-10

Combining Biological Networks to Predict Genetic Interactions

Sharyl L. Wong[1], Lan V. Zhang[2], Amy H. Y. Tong, Zhijian Li, Debra S. Goldberg, Oliver D. King, Guillaume Lesage, Marc Vidal, Brenda Andrews, Howard Bussey, Charles Boone, Frederick P. Roth
[1]sharyl_wong@student.hms.harvard.edu, Harvard Medical School; [2]lan_zhang@student.hms.harvard.edu, Harvard Medical School

We predicted genetic interactions using probabilistic decision trees to integrate multiple types of data including mRNA expression, physical interaction, protein function, and characteristics of network topology. Experimental evidence demonstrated the reliability of this strategy, which may prove valuable in discovering drug targets for cancer therapy and in identifying genes responsible for multigenic diseases.

I-11

Active Learning Strategies for Drug Screening

Megon Walker[1], Simon Kasif[2]
[1]megonw@bu.edu, Bioinformatics Program, Boston University, Boston, MA 02215, USA; [2]kasif@bu.edu, Bioinformatics Program, Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA

From a large chemical library, distinguish compounds that bind a target molecule in as few biochemical testing iterations as possible. The proposed drug screening approach combining query by committee, bagging, and boosting with various sample selection strategies is demonstrated on two pharmaceuticals datasets. Perl source code is available at http://genomics10.bu.edu/megonw.

I-12

Is Aromatic Feature Good Criterion to Predict Protein-Protein Interactions?

Yoojin Chung[1], Illyoung Choi[2]
[1]chungyj@hufs.ac.kr, Hankuk University of Foreign Studies; [2]bugsoda@hanmail.net, Hankuk University of Foreign Studies

We try to predict protein-protein interactions directly from its amino acid sequence and associated features using a Support Vector Machine. As a result of experiments using all possible combinations of 9 features (hydrophobic, positive, negative, polar, charged, small, tiny, aromatic, aliphatic), aromatic shows best results (98% average accuracy with high precision and recall).

I-13

Secondary Structure Prediction of RNA Pairs

Mirela Andronescu[1], Anne Condon[2]
[1]andrones@cs.ubc.ca, University of British Columbia; [2]condon@cs.ubc.ca, University of British Columbia

PairFold is an algorithm for secondary structure prediction of two interacting RNA molecules, a simple extension of Zuker and Siegler algorihm. When testing our algorithm on pseudoknot-free RNA duplexes shorter than 120 nucleotides, we correctly predict approximately 90% of the base pairs on average. PairFold is available online at http://www.RNAsoft.ca

I-14

A Computational Approach for Identification of Pathogenicity Islands in Prokaryotic Genomes

Sung Ho Yoon[1], Cheol-Goo Hur[2], Ho-Young Kang, Jihyun F. Kim
[1]moncher@kribb.re.kr, Genome Research Center, Korea Research Institute of Bioscience and Biotechnology;
[2]hurlee@kribb.re.kr, Genome Research Center, Korea Research Institute of Bioscience and Biotechnology

We developed a computational method for identifying pathogenicity islands (PAIs) in prokaryotic genomes by combining sequence similarities and abnormalities in genomic composition. Of the 141 prokaryotic chromosomes, 23 pathogens and one Escherichia coli K-12 contained 1,286 PAI-homologous genes in 201 regions. 642 PAI-homologous genes in 79 regions showed anomalies in G+C composition and codon usage.

I-15

Loss-Based Estimation Using the Deletion/ Substitution/Addition Algorithm with Applications in Genomics

Sandra E. Sinisi[1], Mark J. van der Laan[2]
[1]ssinisi@stat.berkeley.edu, Division of Biostatistics;
[2]laan@stat.berkeley.edu, Division of Biostatistics

The Deletion/Substitution/Addition (D/S/A) algorithm is a novel regression methodology. Important steps are to define the parameter of interest in terms of a loss function; parameterize the parameter space in terms of basis functions; construct candidate estimators; and use cross-validation to select among candidate estimators. It is applied to an HIV-1 dataset

I-16

Specificity of Asparaginyl Endopeptidase: How are the Peptides used for MHC Class II Presentation Generated? How are the Peptides used for MHC Class II Presentation Generated?

Can Kesmir[1], Colin Watts [2]
[1]c.kesmir@bio.uu.nl, Utrecht University, NL; [2], University of Dundee, UK

Asparaginyl endopeptidase (AEP) is an important enzyme that generates peptides from microbial proteins to induce an immune response. We here develop a method that can predict the specificity of AEP using artificial neural networks and available experiemental data. The method has high 90% sensitivity and 72% specificity.

I-17

Hybrid Computational Method for Pol II Promoter Detection

Ki-Bong Kim[1], Mi-Kyung Lee[2], Jeong Seop Sim
[1]kbkim@smu.ac.kr, Sangmyung University;
[2]mklee@smu.ac.kr, Sangmyung University

We deals with the development of a hybrid computational method which will be employed to detect the core promoter region and the transcription start site (TSS) in vertebrate genomic DNA sequences, an issue of obvious importance for Genome Annotation efforts. The hybrid method internally consists of several SVMs (support vector machines) and several probabilistic models.

I-18

Analysis and Prediction of Leucine-rich Nuclear Export Signals

Tanja la Cour[1], Lars Kiemer[2], Anne Mølgaard, Ramneek Gupta, Karen Skriver, Flemming Poulsen, Søren Brunak
[1]tanja@cbs.dtu.dk, Center for Biological Sequence Analysis, Technical University of Denmark;
[2]lars@cbs.dtu.dk, Center for Biological Sequence Analysis, Technical University of Denmark

We analyse and compare leucine-rich nuclear export signals (NES). Furthermore, we have constructed a predictor based on both neural networks and hidden Markov models. Using this approach we obtain a prediction Matthews correlation coefficient above 0.5 with cross-validation.

I-19

Predicting the Protein Coding Region in cDNA with Frame Shift Errors by Using a Classifier Based on Boosting

Kana Shimizu[1], Yoichi Muraoka[2]
[1]kana@muraoka.info.waseda.ac.jp, Waseda University;
[2]muraoka@waseda.jp, Waseda University

We introduce a new method for predicting protein coding regions in cDNA. Our method detects frame-shift errors. Our approach is based on a boosting algorithm that can efficiently combine a lot of biological information In our experiments, we reduced our method's incidence of errors to a level comparable to that of conventional tools

I-20

MarSel: LD-based Marker Selection System for Large-scale Haplotype Data

Sang-Jun Kim[1], Sung Kwon Kim[2], Kyung-Rak Na, Sang-Soo Yeo
[1]jjuns@alg.cse.cau.ac.kr, Chung-Ang University; [2]skkim@cau.ac.kr, Chung-Ang University

MarSel is a software system to divide haplotypes into blocks, by a mixture of computational methods of dynamic algorithms and biological LD-based methods to select tag SNPs. Through comparisons with other existing systems we show that MarSel performs well in real data as well as artificial data

I-21

Regulogs: Conserved Regulatory Networks in Bacteria

Wynand Alkema[1], Boris Lenhard[2], Wyeth Wasserman
[1]Wynand.Alkema@cgb.ki.se, Karolinska Institutet; [2]Boris.Lenhard@cgb.ki.se, Karolinska Institutet

We present Regulogger (regulogs.cgb.ki.se), a method that predicts regulatory networks in bacteria based on co-conservation of regulatory signals and protein sequences. Regulogger performed well on a test set of known regulatory networks in Escherichia coli. Application of Regulogger to Staphylococcus aureus yielded predictions for new proteins involved in iron uptake

I-22

Computational Prediction of Signals Regulating the Herpesvirus Infection Gene Network

Nicolás Bellora[1], M.Mar Albà[2]
[1]nicolas.bellora@upf.edu, Universitat Pompeu Fabra; [2]malba@imim.es, Universitat Pompeu Fabra

Recognition of sequence signals by transcription factors is a fundamental step in the regulation of gene expression networks. We have performed a computational analysis of the regulatory signals in herpesvirus genomes, considering different virus gene expression classes, to identify candidate interactions for the regulation of the herpesvirus infection network.

I-23

Protein Sequence-structure Alignments for Prediction, Classification and Visualization

Jim Procter[1], Andrew Torda[2]
[1]procter@zbh.uni-hamburg.de, ZBH - University of Hamburg; [2]torda@zbh.uni-hamburg.de, ZBH - University of Hamburg

This poster examines the use of sequence-structure alignments for protein classification, and evaluates the effect that fold conformation and architecture have on the performance of a threading based fold recognition method. The dataset used is also presented as VRML based visualization available for browsing at http://www.zbh.uni-hamburg.de/wurst/protspace/

I-24

Protein Secondary Structure Prediction Using Hidden Markov Models

Martin[1], Gibrat[2], Rodolphe
[1]Juliette.Martin@jouy.inra.fr, INRA; [2]Jean-Francois.Gibrat@jouy.inra.fr, INRA

We present an attempt to predict the secondary structure of proteins with Hidden Markov Models. Several models are evaluated : from the simplest three-states model without memory, to more sophisticated model architecture Our best model at this time uses single-sequence information and has a Q3 score of 66.3 %.

I-25

Computational Assignation of Protein Subcellular Location in a Reverse Vaccinology Context

Paul D. Taylor[1], Darren R. Flower[2]
[1]paul.taylor@jenner.ac.uk, Edward Jenner Institute for Vaccine Research; [2]darrren.flower@jenner.ac.uk, Edward Jenner Institute for Vaccine Research

In recent years reverse vaccinology has utilised the wealth of information provided by genome sequencing to identify and characterise a whole host of new vaccine targets. Here we present methods for the separation of bacterial proteins into subcellular compartments, from sequence, as a tool to provide new possible vaccine candidates.

I-26

A Novel Algorithm for Automated Prediction of Protein-Protein Interactions

Ali Selim Aytuna[1], Ozlem Keskin[2], Attila Gursoy
[1]aaytuna@ku.edu.tr, Koc University - Center for Computational Biology and Bioinformatics; [2]okeskin@ku.edu.tr, Koc University - Center for Computational Biology and Bioinformatics

We present a novel, fully automated parallel algorithm to address the problem of protein-protein interaction prediction. The algorithm discovers possible binary interactions between every polypeptide chain in the PDB by detecting the structural and evolutionary similarities of their non-homologous subset to members of a clustered, non-redundant, representative dataset of existing interactions in PDB.

I-27

Learning an Integrated Probabilistic Protein-Protein Interaction Map

Ariel Jaimovich[1], Gal Elidan[2], Hanah Margalit, Nir Friedman
[1]arielj@cs.huji.ac.il, Hebrew University; [2]galel@cs.huji.ac.il, Hebrew University

We build a joint probabilistic model of protein interactions, interaction assays and the attributes of each protein. As we show, by using this model for inference, different sources of information can be integrated, generating a joint interaction map. We evaluate our approach on protein interactions in Baker's yeast

I-28

Parsimonious Markov Models

Bourguignon[1], Robelin[2]
[1]bourguignon@genopole.cnrs.fr, Laboratoire Statistique et Genome; [2]robelin@genopole.cnrs.fr, Laboratoire Statistique et Genome

Markov models have limitations concerning the order of the models used in bioinformatics, because of the exponential growth of the parameter's dimension with the order of the model. We propose parsimonious Markov models as an alternative for Markov models when high order is required, and show that they exhibit interesting performances.

I-29

Enhanced Statistics for Local Alignment of Multiple Alignments Improves Protein Function and Structure Prediction

Milana Morgenstern[1], Shmuel Pietrokovski[2]
[1]milana@wicc.weizmann.ac.il, Weizmann Institute of Science; [2]shmuel.pietrokovski@weizmann.ac.il, Weizmann Institute of Science

We present an enhanced score significance estimation procedure for LAMA, a method for ungapped profile-to-profile comparison. It is based on column-to-column score probabilities combined using Fisher's Chi-Square method. The improved LAMA was found most sensitive in a large scale comparison with other profile-to-profile comparison methods using various parameters.

I-30

GISMO: Gene Identification Using Homology Information and Support Vector Machines for ORF Classification

Lutz Krause[1], Alice McHardy[2], Jens Stoye, Folker Meyer
[1]Lutz.Krause@CeBiTec.Uni-Bielefeld.DE, Center for Biotechnology (CeBiTec), Bielefeld University, 33594 Bielefeld Germany; [2]Alice.McHardy@CeBiTec.Uni-Bielefeld.DE, Center for Biotechnology (CeBiTec), Bielefeld University, 33594 Bielefeld Germany

For finding protein coding genes in prokaryotic genomes we present the program GISMO, which combines a homology-based approach with OR classification using Support Vector Machines. The algorithm has an average sensitivity of 94% and a specificity of 95%. GISMO outperforms GLIMMER and ZCURVE and is similar in overall performance to CRITICA while being higher in sensitivity.

I-31

New Approaches in Gene Regulation Bioinformatics Using the Constrained Binding Site Diversity within Families of Transcription Factors

Albin[1], Boris[2], Wyeth Wasserman
[1]albin.sandelin@cgb.ki.se, Sandelin; [2]Boris.lenhard@cgb.ki.se, Lenhard

We introduce a new direction in gene regulation bioinformatics based on the constraints imposed by the structures of DNA-binding proteins. Incorporation of such constraints into pattern discovery procedures confers dramatic sensitivity improvements. We present initial results of using generalized models for classes of transcription factors to classify sets of genes.

I-32

Predicting Protein Function From Structure

Paul D. Dobson[1], Andrew J. Doig[2]
[1]p.dobson@postgrad.umist.ac.uk, UMIST;
[2]andrew.doig@umist.ac.uk, UMIST

By charactering the general structural properties of proteins with the same top-level enzyme class, we develop a method to predict function that is independent of structural similarity. The model performs significantly better than random, with the top-ranked prediction accurate to approximately 40%, rising to around 60% with the top two ranks.

I-33

Modeling of Receptor-Ligand Interaction Using Rule Based Methods

Helena Strömbergsson[1], Peteris Prusis[2], Herman Midelfart & Jan Komorowski
[1]helena.strombergsson@lcb.uu.se, Linnaeus Centre for Bionformatics, Uppsala University;
[2]Peteris.Prusis@farmbio.uu.se , Department of Pharmaceutical Pharmacology, Uppsala University, Linnaeus Centre for Bionformatics, Uppsala University

Models of receptor-ligand interaction between melanocortin receptors and peptide ligands have been induced. Validations of the models resulted in an AUC mean of 0.94 (SE 0.12) and 0.91 (SE 0.19) for the two datasets used in this study. This indicates that the rough set approach may be used for modeling and predicting receptor-ligand interaction.

I-34

DTW and Stagewise Regression for the Molecular Profiling of Microarray Time Series

Cesare Furlanello[1], Giuseppe Jurman[2], Maria Serafini, Samantha Riccadonna, Diego Giuliani, Stefano Merler
[1]furlan@itc.it, ITC-irst; [2]jurman@itc.it, ITC-irst

Given temporal microarray data, a panel of time series is selected by integrating a stagewise boosting algorithm with the Dynamic Time Warping distance considered as similarity measure. The approach is applied on the Cardiogenomics PGA mouse model of Myocardial Infarction to identify a molecular profile of the ventricular remodeling process.

I-35

Structural Stability Prediction of Short Beta-hairpin Peptides by Molecular Dynamics and Knowledge Based Potentials

Karin Noy[1], Nir Kalisman[2], Dr. Chen Keasar
[1]noyk@bgumail.bgu.ac.il, Ben Gurion University, Israel;
[2]nirka@cs.bgu.ac.il, Ben Gurion University, Israel

We present a computational approach for structural stability prediction of short peptides by using two orthogonal computational techniques: Molecular Dynamics and knowledge based potentials. The major observation is a clear correlation between the experimental and computed both stabilities. The prediction scheme implied by this correlation can help the design of efficient combinatorial peptide libraries.

I-36

The Role of Electrostatic Coupling and Strain in Protein and Enzyme Function

Richard Greaves[1], Jim Warwicker[2]
[1]R.Greaves@umist.ac.uk, UMIST;
[2]j.warwicker@umist.ac.uk, UMIST

Electrostatic strain should be reflected in particularly strong coupling between ionisable groups and/or dehydration of key groups in computational studies of protein ionisation behaviour. We are investigating the association between electrostatic coupling or strain and protein function in proton pumping proteins and for the prediction of enzyme active site location

I-37

Unsupervised Discovery from Gene Tracking with RFE Classification Systems

Cesare Furlanello[1], Stefano Merler[2], Maria Serafini, Giuseppe Jurman
[1]furlan@itc.it, ITC-irst; [2]merler@itc.it, ITC-irst

We present a discovery method for predictive classification in DNA microarray experiments coming from recursive feature elimination systems and implemented on a high-throughput computing (HTC) facility. We developed analyzers for single-sample response to the gene selection process and gene-importance tracking, allowing the discovery of subclasses of different molecular profiles.

I-38

Towards a Probabilistic Model of Peptide Fragmentation in Mass Spectrometry

Behshad Behzadi[1], Benno Schwikowski[2]
[1]behzadi@lix.polytechnique.fr, LIX, Ecole Polytechnique; [2]benno@schwikowski.de, 1)Institut Pasteur, Paris, France 2)Institute for Systems Biology, Seattle, WA

Building on a dataset that contains hunreds of thousands of ms/ms spectra, we model the following components of a probabilistic model for CID peak intensity: distribution of peak intensities of single peaks in repeat experiments and the effects of whitin-sequence positions, charge, and isotopes on peak intensity.

I-39

Analysis of Variation of Gene Expression Levels

Steinfath[1], Wierling[2], Marie-Laure Yaspo, Marc Sultan, Hans Lehrach, Ralf Herwig
[1]steinfat@molgen.mpg.de, Max-Planck-Institut for molecular genetics; [2], Max-Planck-Institut for molecular genetics

The change in the variation of gene expression between different groups of individuals is investigated. The measured gene expression is the combination of biological and technical variance. Methods to estimate and overcome the statistical and systematical errors due to these influences are tested with artificial and experimental data.

I-40

In Silico Prediction of Beta-barrel Transmembrane Proteins

Andrew Garrow[1]
[1]andy@bioinformatics.leeds.ac.uk, University of Leeds

Beta-barrel transmembrane proteins can be found spanning the outer membranes of gram -ve bacteria, mitochondria and chloroplasts. To date, prediction remains troublesome, with transmembrane domains hidden by a cryptic inside-outside dyad repeat motif. We are experimenting with a number of prediction approaches including HMMs and machine learning methods.

I-41

Prediction of -1 and +1 Frameshift Signals in Genomic Sequences

Kyungsook Han [1], Yanga Byun[2], Sanghoon Moon
[1]khan@inha.ac.kr, Inha University; [2]quaah@hanmail.net, Inha University

We have developed an algorithm that predicts both −1 and +1 frameshift signals in the entire genomic sequences, and implemented the algorithm in a program called FSFinder (Frameshift Signal Finder). FSFinder was tested on 146 genomic sequences, and frameshift signals were predicted with greater sensitivity and specificity than existing approaches.

I-42

Identification of Genes that are Directly Regulated by Transcription Factor GCN4p: Utilization of Gene Expression Data, Regulatory Motif Finding and Feature Selection

Sonika Tyagi[1], Alok Bhattacharya[2], Probal Choudhury, S. N. Maheswari
[1]st0645@students.jnu.ac.in, School of Information Technology; [2]alok0200@mail.jnu.ac.in, School of Information Technology

Identification of coregulated genes is a major problem in biology. Here we focus on a data set of GCN4p regulated genes and utilize different motif finding methods along with some of the features based on biophysical and statistical properties to identify sequence motifs associated with genes that are directly activated by GCN4p.

I-43

A Two Level Computational Method to Detect Non Coding RNA

Gard Thomassen[1], Josef Thingnes[2]
[1]gardt@ifi.uio.no, Centre for Molecular biology and Neuroscience; [2]joseft@ifi.uio.no, Centre for Molecular biology and Neuroscience

We are working on a two level computational method to detect novel non-coding RNA (ncRNA) where the first level searches for transcription signals, while the second level employs a Covariance Model (CM) to search for a consensus in secondary as well as in primary structure.

I-44

Using TDNNs to Predict Protein Interactions by Locating Relevant Sequence Features

Wolfgang Lehrach[1], Dirk Husmeier[2], Chris Williams, David Barber
[1]s0239540@sms.ed.ac.uk, University of Edinburgh and Biomathematics & Statistics Scotland;
[2]dirk@bioss.sari.ac.uk, Biomathematics & Statistics Scotland

A novel Time Delay Neural Network based model of protein interactions allows prediction of interactions, the identification of known motifs and the discovery of new sequence features that are important for interactions to take place. The model is evaluated on generated data and the Database of Interacting Proteins.

I-45

Investigating the 3D Conformation of N-terminally Constrained Oligopeptides

Itizar Frades[1], Richard J Edwards[2], Denis C Shields
[1]fitziar@rcsi.ie, Royal College of Surgeons in Ireland;
[2]redwards@rcsi.ie, Royal College of Surgeons in Ireland

Some N-terminally constrained oligopeptides show activity in biological systems while others do not but the characteristics responsible for activity are yet to be determined. N-terminal regions of resolved 3D structures protein can be used as a surrogate for such peptides to investigate conformational stability and its role in bioactivity.

I-46

Improvements to an Analysis of Disease Susceptibility Loci to Identify Novel Disease Genes

Frances Turner[1], Colin Semple[2]
[1]fturner@hgu.mrc.ac.uk, MRC Human Genetics Unit;
[2]colins@hgu.mrc.ac.uk, MRC Human Genetics Unit

We have previously published a computational method for the prioritisation of positional candidate genes for diseases where multiple susceptibility loci exist. Here we present improvements to the method based on the inclusion of publicly available SAGE, microarray, and EST data to identify co-expressed disease genes.

I-47

The Use of Operators as Predictors of Operons in Streptomyces Coelicolor

E Laing[1], S Hubbard[2]
[1]e.laing@postgrad.umist.ac.uk, UMIST;
[2]simon.hubbard@umist.ac.uk, UMIST

Methods that make use of regulatory units (RUs) to predict operons concentrate on finding a gene or set of genes bracketed by an upstream promoter and downstream terminator. Here, we present a strategy that attempts to predict operons through the identification of a different RU, the operator.

I-48

Why do Internal Promoters Play Second Fiddle?

Aditi Kanhere[1], Manju Bansal[2]
[1]aditi@mbu.iisc.erner.in, Indian Institute of Science;
[2]mb@mbu.iisc.ernet.in, Indian Institute of Science

We calculated various structural and physical properties for 279 Escherichia coli promoter sequences. Interestingly, most of the promoters, which do not show any distinct structural and physical characteristics, are either internal promoters or weak promoters. The secondary role of internal promoters may be due to the absence of any such special properties.

I-49

Contact Map Prediction Using Neural Networks

Pelin Akan[1], Berrin Yanikoglu[2], Osman Ugur Sezerman
[1]pa2@sanger.ac.uk, Wellcome Trust Sanger Institute;
[2]berrin@sabanciuniv.edu, Sabanci University

Neural networks are used to predict contacting residues in tertiary structure of proteins. Networks are fed by several physical and chemical properties of the residues and their neighbours. The predictor is 7.2 times better than a random predictor and 11% of the contacting residues predicted with 2% false positive ratio.

I-50

Accurate Plant Promoter Prediction with Confidence Estimation

I.A.Shahmuradov[1], V.V.Solovyev1[2], A.J.Gammerman
[1]ilham@cs.rhul.ac.uk , Royal Holloway, University of London; [2]victor@softberry.com, Softberry Inc., USA

Using SVM/TCM discriminative approaches, we developed a new program TSSP-TCM for prediction of plant TATA and TATA-less promoters with 95% confidence level. The program predicted TSS for 88% (35/40) and 84% (21/25) of known TATA and TATA-less promoters, respectively, with a median deviation of several nucleotides from the annotated TSS.

I-51

Novel Computational Method for Human Cis Regulatory Elements Prediction

Alberto Ambesi-Impiombato[1], Diego di Bernardo[2]
[1]ambesi@tigem.it, Telethon Institute of Genetics and Medicine; [2]dibernardo@tigem.it, Telethon Institute of Genetics and Medicine

BID is a novel algorithm for CRE prediction of human genes based on a new way to model background noise and to integrate ortholog information. BID was tested on a set of 208 human genes for which binding sites were known, and its performance was compared to alternative algorithm MatchTM

I-52

Using Biologically Relevant Structural Data to Predict Protein-protein Interactions in Proteins of Unknown Sequence

Emily Jefferson[1], Geoff Barton[2]
[1]emily@compbio.dundee.ac.uk, Dundee University; [2]geoff@compbio.dundee.ac.uk, Dundee University

We investigated the number of different unique domain-domain interactions seen in PDB structures in comparison to those seen in PQS from the MSD. It was found that there are more domain interactions observed in PQS than if only PDB structures were used. These extra domains interactions will increase the reliabilty of interaction prediction tools which use structural data.

I-53

Prediction of Protein Kinase Substrate Specificity

Diego Miranda-Saavedra[1], Geoffrey J Barton[2], n/a
[1]diego@compbio.dundee.ac.uk, University of Dundee; [2]geoff@compbio.dundee.ac.uk, University of Dundee

The prediction of protein kinase substrate specificity is an active area of research. We are taking a structural approach to predicting the specificity of protein kinases.

I-54

Naive Bayesian Prediction of Interaction of MHC and MHC Superfamily with B2M and Characterization of IMGT Positions

Duprat E[1], Gascuel O[2], Lefranc M-P
[1]duprat@ligm.igh.cnrs.fr, Laboratoire d'Immuno Génétique Moleculaire, Université Montpellier II, UPR CNRS 1142 IGH, 141 rue de la Cardonille, F-34396 Montpellier, France; [2]gascuel@lirmm.fr, Département d'Informatique Fondamentale et Applications, LIRMM, 161 rue Ada, 34392, Montpellier, France

We present results from naive Bayesian prediction of MHC and MHC superfamily interaction with B2M (90% of accuracy), based on IMGT domain standardization. This method, cross-validated by leave-one-out procedures, provides identification of important positions implicated in this interaction and relation with their respective amino acid physico-chemical properties.

I-55

Simultaneous Prediction of MicroRNAs and their Targets in Arabidopsis Thaliana

Morten Lindow[1], Anders Krogh[2]
[1]morten@binf.ku.dk, University of Copenhagen; [2]krogh@binf.ku.dk, University of Copenhagen

We demanding near-perfect complementarity to mRNAs, ability of precursors to form hairpins and phylogenetic conservation, we have developed a method to predict MicroRNAs in plants with 75% specificity and around 50% sensitivity

I-56

Phobius: a Combined Transmembrane Topology and Signal Peptide Predictor

Lukas Käll[1], Anders Krogh[2], Erik L.L. Sonnhammer
[1]Lukas.Kall@cgb.ki.se, Karolinska Institutet; [2]krogh@binf.ku.dk, University of Copenhagen

We have constructed and trained a combined transmembrane topology and signal peptide predictor. The predictor makes far fewer errors due to cross predictions of trans membrane segments and signal preptides compared to conventional predictors. The method isavailable at http://phobius.cgb.ki.se/ as well as at http://phobius.binf.ku.dk/

I-57

Sequence Shuffling as a Model for Assessing Oligomer Frequencies

E. A. Rodland[1]
[1]e.a.rodland@labmed.uio.no, Rikshospitalet

We present results on oligomer frequencies in shuffled sequences (constrained Markov model), including exact formulas for the oligome counts. Short-comings of this model are discussed.

I-58

Developing Genome-Scale Prediction System for Transcription Factors and Their Targets

Akinori Sarai[1], Shandar Ahmad[2], Michael M. Gromiha, Hidetoshi Kono
[1]sarai@bse.kyutech.ac.jp, Kyushu Institute of Technology; [2]shandar@bse.kyutech.ac.jp, Kyushu Institute of Technology

We have developed a strategy for predicting transcription factors and their targets based on various kinds of information such as sequence information, experimental binding data, Genome Annotation information, and computer simulations. We are integrating these methods to create a genome-scale prediction system for transcription factors and their targets.

I-59

Detection of New Protein-peptide Mediated Interactions

Neduva, V.[1], Linding, R.[2], Stark A., Gibson T. and Russell R.
[1]neduva@embl.de, EMBL; [2]linding@embl.de, EMBL

Many biological processes are governed by interactions between globular domains and short linear peptides (linear motifs). We present an approach to find the best motifs from a set of sequences sharing an interaction partner or some other common function. We also present the results of the systematic search of the new motifs using genome-scale interaction discovery studies.

I-60

Making Optimal use of Empirical Energy Functions - the YAMBER Force Field

Elmar Krieger[1], Tom Darden, Sander B. Nabuurs, Alexei Finkelstein, Gert Vriend
[1]elmar@cmbi.kun.nl, CMBI

Increasingly accurate energy functions are required to improve protein models built by homology, for example during molecular dynamics refinement. By simulating protein crystals and iteratively adjusting the AMBER force field parameters to minimize the deviations from experiment, we arrive at the YAMBER force field, which is shown to move homology models in the right direction.

I-61

Disulfide Connectivity Prediction Using Secondary Structure Information

Fabrizio Ferre[1], Peter Clote[2]
[1]ferref@bc.edu, Boston College; [2]clote@bc.edu, Boston Colege

Starting from the observation that there is a bias in the secondary structure preferences of free cysteines and half-cystines, we develop a neural network to learn disulfide bond preferences of both amino acid residues and secondary structure assignment of the symmetric flanking regions centered at partner half-cystines.

I-62

Comparison of Protein Structural Comparison Methods, VAST and SHEBA, with the SCOP Classification, Using Statistical Methods

Vichetra SAM[1], Chin-Hsien (Emily) Tai[2], Jean Garnier, Jean-Francois Gibrat, Byungkook Lee, Peter J Munson
[1]vsam@helix.nih.gov, MSCL/DCB/CIT/NOH/DHHS; [2]taic@pop.nci.nih.gov, MMBS/LMB/NCI/NIH/DHHS

We conducted a comparison of the structural comparison methods VAST and SHEBA with the structural classification SCOP. Higher agreement with SCOP was obtained when combinations of scores from VAST and SHEBA were used. Fold-classification confusions were quantified. In depth visual analysis of the structures of confused folds will be presented.

I-63

HLA and HIV Infection Progression: Application of the Minimum Description Length Principle to Statistical Genetics

Peter Hraber[1], Bette Korber[2], Steven Wolinsky, Henry Erlich, Elizabeth Trachtenberg, Thomas B. Kepler
[1]phraber@exagendiagnostics.com, Exagen Diagnostics; [2]btk@lanl.gov, Los Alamos National Laboratory

In HIV-infected cells, HLA molecules present antigen fragments to T-cells with varying specificity. The minimum description length principle enabled us to classify allele associations with viral load. Variation in progression is strongly associated with HLA-B. Individuals without B58s alleles have 3.6 times greater viral loads than individuals with them.

I-64

Developing a Confidence Measure for Protein Interactions

Ruth Isserlin[1], Christopher W.V. Hogue[2]
[1]isserlin@mshri.on.ca, Samuel Lunenfeld Research Institute, University of Toronto; [2]hogue@mshri.on.ca, Samuel Lunenfeld Research Institute, University of Toronto

Advances in high throughput techniques for the generation of protein-protein interactions have produced a wealth of data. Inherent to these techniques is a large amount of false positive results. We are developing a probabilistic framework for assigning confidence to binary protein interactions to enable the efficient use of interaction data.

I-65

Classification of Escherichia Coli Microarray Genes into Co-expressed Modules Based on Genes Regulation

Haleh Yasrebi[1], Maia Angelova[2]
[1]haleh.yasrebi@unn.ac.uk, Department of informatics, Northumbria University; [2]maia.angelova@unn.ac.uk, Department of informatics, Northumbria University

Identifying co-expressed genes allows to infer the function of unknown genes by comparing the co-regulated genes to the genes with characterized function. Here, we study the functional roles and regulation of new genome-wide microarray expression data for Escherichia coli by using Principal Component Analysis and clustering, individually and in combination with each other.

I-66

Domain Linker Prediction By Amino Acid Composition

Michel Dumontier[1], Christopher W.V. Hogue[2], Rong Yao
[1]mjdumont@blueprint.org, Blueprint Initiative; [2]chogue@blueprin.torg, Blueprint Initiative

The identification and annotation of protein domains is a crucial step in the accurate determination of molecular function. An amino acid index was derived from the amino acid composition of linker regions and compact domains. Armadillo (http://armadillo.blueprint.org) provides a simple method for the prediction of protein linker domains by sequence.

I-67

Protein Flexibility Prediction from Sequence

Avner Schlessinger[1], Burkhard[2]
[1]as2067@columbia.edu, Columbia University; [2]rost@columbia.edu, Rost

We performed a large scale analysis of B-values of residues. Then, we created a neural network that identifies flexible residues in proteins. It was shown that prediction from sequence of flexible and rigid segments in proteins helps us identify conformational switches and regions with rigidity that is crucial for function.

I-68

Predicting Function From Sequence

Andrew Kernytsky[1], Burkhard Rost[2]
[1]kernytsky@maple.bioc.columbia.edu, Columbia University; [2]rost@bioc.columbia.edu, Columbia University

Using neural networks we have created an algorithm that predicts the first Enzyme Classification (EC) number of a protein sequence. The method attempts to improve upon previous methods by focusing on local sequence information, proper partitioning of the protein at domain boundaries, and using an improved learning method.

I-69

Discovery of Novel Conserved Protein Domains: Sequence Partitioning Using the Libraries of Already Discovered Domains

Deendayal Dinakarpandian[1], Stijn van Dongen, Arcady Mushegian
[1]dinakard@umkc.edu, University of Missouri Kansas City

A semi-automated approach to the identification of conserved domains is presented that offers an efficient alternative to an exhaustive approach. The validity of the method is demonstrated on the proteome of Baker's yeast, with the ultimate aim of finding all conserved domains in the nr database.

I-70

SOKOS/CAN: A Novel Method to Find Non-coding RNAs in a Genomic Sequence

Taishin KIN[1], Koji Tsuda[2], Kiyoshi Asai
[1]taishin@cbrc.jp, Computational Biology Research Center/AIST; [2]koji.tsuda@tuebingen.mpg.de, Max Planck Institute for Biological Cybernetics

Finding non-coding RNAs in a genomic sequence is one of the most important yet difficult tasks in bioinformatics. We tackle this problem by introducing a novel approach that incorporates stochastic context free grammar and support vector machine. Our approach is realized as a software package called SOKOS/CAN.

I-71

New Coding Measures Based on Singular Spectrum Analysis (SSA)

Rolando Hong[1], Miguel Sautie[2], Carlos Martinez, Jose Luis Hernandez Caceres
[1]hong@cecam.sld.cu, CECAM; [2]hong@cecam.sld.cu, CECAM

We propose 8 closely related parameters based on singular spectrum analysis (SSA) to separate DNA protein coding sequences from non-coding sequences. We evaluate comparatively the proposed coding measures on DNA sequences from several genomes and asses the size dependency of the SSA-parameters.

I-72

PCB: A Predictive System for Classifying Multimodal Brain Tumor Images in an Image-guided Medical Diagnosis Model

Lau Phooi Yee[1], Ozawa Shinji[2]
[1]laupy@ozawa.ics.keio.ac.jp, Keio University; [2]ozawa@ozawa.ics.keio.ac.jp, Keio University

PCB is a diagnosis research-oriented system having the capability to classify and make connections between a disease and its tumor images properties, to its respective diagnosis and treatment prescription. PCB focuses on three different perspectives for analyzing input images: content-based, shape-based and texture-based techniques, irrespective of age and gender.

I-73

A Knowledge Based Model for Protein-DNA Interactions: a Structural Approach

Richard Gamblin[1], Richard Jackson[2]
[1]bms8rjg@bmb.leeds.ac.uk, University of Leeds; [2]jackson@bmb.leeds.ac.uk, University of Leeds

We have developed a statistical model for protein-DNA interactions based on an analysis of hydrogen bonding and contact interaction patterns from a non-redundant set of protein-DNA complex structures We describe the application of this model to specific cases of DNA binding proteins, in addition to its assessment as compared with equivalent experimentally derived PSSM models.

I-74

Avoiding Erroneous Prediction in Transcriptomics, Proteomics and Metabolomics

Amir A. Handzel[1]
[1]ahandzel@beyondgenomics.com, Beyond Genomics

Cross validation enables good generalization in statistical prediction from finite sets. It may fail, however, when the number of samples is much smaller than the number of variables. Numerical experiments show that large validation sets should be used, whereas the popular leave-one-out gives poor results and should be abandoned in molecular biology practice

I-75

Prediction of Migration Time of Cations in Capillary Electrophoresis - Mass Spectrometry Using Artificial Neural Networks

Masahiro Sugimoto[1], Shinichi Kikuchi[2], Masanori Arita, Tomoyoshi Soga, Masaru Tomita
[1]msugi@sfc.keio.ac.jp, Keio Univ.;
[2]kikuchi@sfc.keio.ac.jp, Keio Univ.

The computational method to predict migration time of cationic metabolites measured by Capillary Electrophoresis – Mass Spectrometry using Artificial Neural Networks ensemble method was developed. This method is adaptable for metabolomics analysis to identify many kinds of cationic metabolites exhibiting unknown peaks.

I-76

Generation of 3D Templates for Enzymes in the Catalytic Site Atlas: Analysis of Catalytic Residue Geometry and Utility of Automatically-generated Templates for Function Prediction from 3D Structure

Gail J. Bartlett[1], James W. Torrance[2], Craig T. Porter, Jonathan A. Barker, Alex Gutteridge, Malcolm W. MacArthur, Janet M. Thornton
[1]g.bartlett@imperial.ac.uk, Imperial College London;
[2]torrance@ebi.ac.uk, European Bioinformatics Institute

3D active site templates were automatically generated from enzyme structures in a database of catalytic sites. Templates were used to analyse catalytic residue geometry within homologous enzyme families, and also to test a template-based active site recognition method. Template effectiveness was measured using RMSD and statistical significance scores

I-77

Prediction of Transcription Factor Binding Sites by Modelin Multiple Features

Rainer Pudimat[1], Rolf Backofen[2]
[1]rpudimat@informatik.uni-jena.de, FSU Jena;
[2]backofen@inf.uni-Jena.de, FSU Jena

The objective of the presented work is to face the weak predictiveness of current approaches for detecting transcription factor binding sites. First, this is done by applying a wide range of sequence-related properties for modeling sites and second by considering statistical dependencies among these properties.

I-78

Support Vector Machine Approach to Active Sites Prediction Using Local Sequence Information

Dariusz Plewczynski[1], Leszek Rychlewski[2], Adrian Tkacz, Lucjan Wyrwicz
[1]darman@bioinfo.pl, BioInfoBank Inst.;
[2]leszek@bioinfo.pl, BioInfoBank Inst.

The AutoMotif Server (AMS) predicts functional patterns in proteins using the support vector machine SVM approach. A list of possible functional motifs for a given query protein is predicted using only query protein sequence and the database of proteins annotated for certain types of biological processes by Swiss-Prot database. The method is available as an internet server at http://automotif.bioinfo.pl/.

I-79

WAPP - Prediction of MHC Class I Antigen Processing

Pierre Dönnes[1], Oliver Kohlbacher[2]
[1]doennes@informatik.uni-tuebingen.de, Department for Simulation of Biological Systems, Eberhard Karls University Tübingen; [2]oliver.kohlbacher@uni-tuebingen.de, Department for Simulation of Biological Systems, Eberhard Karls University Tübingen

WAPP offers integrated prediction of the whole antigen processing pathway of MHC class I peptides. Predicted peptides are likely to be correctly cleaved at the C-terminal, transported by TAP and bound by MHC class I molecules.
WAPP is available at http://www-bs.informatik.uni-tuebingen.de/WAPP/

I-80

Improved Prediction of Bacterial Protein Subcellular Localization Using Support Vector Machine and Insights Gained from Comparative Analysis of Bacterial Proteomes

Jennifer L. Gardy[1], Fei Chen[2], Christopher J. Walsh, Matthew R. Laird, Sébastien Rey, Martin Ester, Fiona S.L. Brinkman
[1]jlgardy@sfu.ca, Molecular Biology & Biochemistry, Simon Fraser University; [2]fchen@cs.sfu.ca, Computing Science, Simon Fraser University

A novel support vector machine using frequent subsequences was developed to classify bacterial proteins by localization with precision of 86-100%. The SVMs are incorporated into an upcoming PSORT-B release which was used to analyze multiple proteomes, revealing that the proportion of proteins per localization does not change with genome size.

I-81

Long Short Term Memory networks for protein localization prediction

T. Thireou[1], M. Reczko[2]
[1]thireou@ics.forth.gr, Bioinformatics Lab, Institute of Computer Science, Foundation for Research and Technology - Hellas (FORTH); [2]reczko@ics.forth.gr, Bioinformatics Lab, Institute of Computer Science, Foundation for Research and Technology - Hellas (FORTH)

A new neural network learning algorithm called Long Short Term Memory (LSTM) is applied for the sequence-based prediction of the subcellular localization of proteins. LSTM networks can learn to bridge long gaps between sequence features using constant error flow. A small LSTM-network outperforms other methods in predicting mitochondrial targeting peptides.

I-82

Computational Analysis of Type-1 Polyketide Synthases

Gitanjali[1], D. Mohanty[2], Rajesh S. Gokhale
[1]yadavg@nii.res.in, NII; [2]deb@nii.res.in, NII

Polyketide synthases (PKSs) are large multi-enzymatic, multi-domain megasynthases, We have develop knowledge based in silico approaches for correlating the sequence and domain organization of PKSs to their polyketide products and prediction of their substrate specificity. These results will provide guidelines for rational design of 'novel' natural products by genetic manipulations.

I-83

A Self-Organizing Map of Probability Weight Matrices for Motif Identification

Shaun Mahony[1], David Hendrix[2], Daniel S. Rokhsar
[1]shaun.mahony@nuigalway.ie, NCBES, NUI Galway; [2]dhendrix@socrates.berkeley.edu, Dept. of Physics, UC Berkeley

We propose a self-organizing map of probability weight matrices as a method for motif discovery. This approach can be used to simultaneously characterise every feature present in a dataset, thus lessening the chance that weaker signals will be missed. Advantageous performance over other popular motif-finding methods is demonstrated.

I-84

FASE: A New Fold Recognition Method Through Entropy Profiles

Alejandro Sanchez-Flores[1], Lorenzo Segovia[2]
[1]alexsf@ibt.unam.mx, Instituto de Biotecnolgia-UNAM; [2]alexsf@ibt.unam.mx, Biotecnolgia-UNAM

FASE is a entropy profile-based method to identify and group protein families sharing fold. This approach doesn't use sequence profiles per-se but only the entropy values derived from alignments of homologous proteins. FASE is a reliable method to assign fold to proteins with unknown structure.

I-85

Pathway Analysis of Mycoplasma Pneumoniae Nucleotide Metabolism

Mikhail Pachkov[1], Stefan Schuster[2]
[1]pachkov@minet.uni-jena.de, University of Jena; [2]schuster@minet.uni-jena.de, University of Jena

We present pathway analysis of the nucleotide metabolism of Mycoplasma Pneumoniae. The aim is to find out whether annotated and experimentally observed enzymes activities are sufficient to sustain necessary functionality of the purine and pyrimidine metabolic systems.

I-86

Dynamical Monte Carlo Simulations of Protein Folding using Kinetic Rates and a Statistical Energy Function

Andres Colubri[1]
[1]acolubri@uchicago.edu, University of Chicago

We present a hybrid Dynamical Monte Carlo approach to simulate the dynamics of protein folding in which the kinetic rates for elementary folding events (contact breaking and formation) are determined from experimental data, meanwhile a statistical potential which includes the orientational effects of the side-chains and backbone hydrogen bonds is used to generate structures in torsional j-f space.

I-87

Hybrid System NN/HMM for Large-scale GPI-Anchored Protein Prediction

Guylaine Poisson[1], Anne Bergeron[2], Cedric Chauve
[1]poisson@math.uqam.ca, UQAM;
[2]bergeron.anne@uqam.ca, UQAM

A glycosyl phosphatidyl-inositol (GPI) anchor is a C-terminal post-translational modification of proteins. Here, we investigate the problem of correctly annotating GPI-anchored protein for the growing number of sequences in public Databases. We developed a hybrid system based on the tandem use of Neural Network and Hidden Markov Model methods.

I-88

SDPpred: a Method for Prediction of Amino Acid Residues that Determine Differences in Functional Specificity of Homologous Proteins

Olga V. Kalinina[1], Pavel S. Novichkov[2], Andrey A. Mironov, Mikhail S. Gelfand, Aleksandra B. Rakhmaninova
[1]ok81@yandex.ru, Department of Bioengineering and Bioinformatics, Moscow State University; [2], Department of Bioengineering and Bioinformatics, Moscow State University

SDPpred (Specificity Determining Position prediction) is designed for the analysis of protein families, whose members have biochemically similar but not identical interaction partners (e.g., different substrates for a family of transporters). It predicts residues that could be responsible for the proteins' choice of correct interaction partners. SDPpred is available at http://math.belozersky.msu.ru/~psn/.

I-89

Toward Good Contact Predictions in Proteins

Marco Punta[1], Burkhard Rost[2]
[1]punta@cubic.bioc.columbia.edu, Columbia University - New York;
[2]rost@columbia.edu, Columbia University - New York

We introduced a novel method that predicts inter-residue contacts. It combines evolutionary information, 1D predictions, and information from regions around and between the two contacting residues yielding rather accurate predictions. We also proposed a novel method estimating the overall number of contacts in a protein thereby providing information about stability.

I-90

I-Mutant: Predicting Protein Stability upon Mutation

Emidio Capriotti[1], Piero Fariselli[2], Rita Casadio - CIRB/Dept. of Biology - University of Bologna
[1]emidio@biocomp.unibo.it, CIRB - University of Bologna; [2]piero@biocomp.unibo.it, CIRB/Dept. of Biology - University of Bologna

I-Mutant is an efficient tool for designing single-site mutations in proteins (80% accuracy). Our method is based on neural networks and predicts changes of protein stability upon mutation. The network is trained/tested on a data set of experimental free energy changes detected upon protein mutations and derived from ProTherm.

I-91

Dr. Frankenstein: Automated Protein Structure Prediction by Assembly of Fragments of Multiple Fold-Recognition Models

Marcin Feder [1], Janusz M. Bujnicki [2], Michal Boniecki, Marcin Pawlowski, Michal J. Gajda, Michal A. Kurowski, Joanna M. Sasin
[1]marcin@genesilico.pl, International Institute of Molecular and Cell Biology in Warsaw, Poland;
[2]iamb@genesilico.pl, International Institute of Molecular and Cell Biology in Warsaw, Poland

We designed an automated protein structure predictor based on the FRankenstein's Monster approach. Starting from a set of fold-recognition target-template alignments, it generates local sequence shifts, builds preliminary models, identifies the most common and best-scoring fragments, and creates hybrid models. The final model is selected based on the energy evaluation.

I-92

Specificity Prediction of Adenylation Domains in Nonribosomal Peptide Synthetases (NRPS) Using Support Vector Machines (SVM)

Christian Rausch[1], Tilmann Weber[2], Daniel Huson, Wolfgang Wohlleben
[1]rausch@informatik.uni-tuebingen.de, Center for Bioinformatics Tübingen, University of Tübingen;
[2]Department of Microbiology/Biotechnology, University of Tübingen

A new SVM-based approach to predict the substrate specificity of aryl- and amino-acid activating adenylation domains of NRPS. The residues 8Å around the substrate Phenylalanine bound in the crystal structure of Gramicidin Synthetase A and corresponding positions of other sequences with known specificity were used to construct normalized feature vectors

I-93

Analyzing and Enhancing mRNA Translational Efficiency in an E Coli In Vitro Expression System

Voges[1], Watzele[2], Cordula Nemetz, Sabine Wizemann, Bernd Buchberger
[1]Dieter.Voges@biomax.de, Biomax Informatics AG, Lochhamerstr. 11, D-82152 Martinsried; [2]Manfred.Watzele@roche.com, Roche Diagnostics GmbH, Roche Applied Science, Nonnenwald 2, D-82372 Penzberg

The efficiency of translation initiation was analyzed in an E coli in vitro experiment. Sequence varied in codons 2 to 14. Expression amount depended on mRNA secondary structure and G+C content. Using a derived model for the prediction of expression score translation efficiency was optimized by rational mutagenesis

I-94

Gene Prediction Analysis Based on Spectral Representation of DNA Sequences and Fuzzy KNN

Daniel Kotlar [1], Yizhar Lavner [2], Zohar Idelson
[1]danny-k@actcom.co.il, Tel-Hai academic college; [2]yizhar_l@kyiftah.org.il, Tel-Hai academic college

We present a new gene prediction algorithm, based on spectral representation of DNA sequences in a three-dimensional complex space. A fuzzy K-Nearest Neighbor (KNN) algorithm is applied to discriminate between coding and non-coding sequences. Over 91% of correct identifiction was achieved for S. Cerevisiae genome.

I-95

Secondary Structure Prediction Using Cascade-Correlation Neural Networks

Matthew J. Wood[1], Jonathan. D. Hirst[2]
[1]pcxmjw@nottingham.ac.uk, University of Nottingham; [2]jonathan.hirst@nottingham.ac.uk, University of Nottingham

A number of popular methods use back-propagation neural networks to predict protein secondary structure. However, this learning algorithm can be slow to learn. We compare it to the cascade-correlation algorithm which achieves predictive accuracies comparable to those obtained by back-propagation, in a shorter time.

I-96

On the Performance of Promoter Motif Finding Programs

Tatiana Tatarinova[1], Nickolai Alexandrov[2]
[1]ttatarinova@ceres-inc.com, Ceres, inc; [2]nicka@ceres-inc.com, Ceres, inc

We developed a benchmark for motif-finding programs using gene expression data and compared performance of several programs, including MEME, AlignACE, and our new promoter motif discovery algorithm, TATA. We have investigated conservation of promoter motifs between two species, Arabidopsis thaliana and Oryza sativa.

I-97

Detection of T-cell Epitopes and Modeling of DQ2 and DQ7 Molecules Using Computational Methods

Aggeliki Kosmopoulou[1], Athanasios Staurakoudis[2], Metaxia Vlassi, Maria Sakarellos-Daitsiotis, Constantinos Sakarellos
[1]me00371@cc.uoi.gr, University of Ioannina; [2]astavrak@cc.uoi.gr, University of Ioannina

The aim of this study is the homology modeling of the MHC class-II DQ2 and DQ7 molecules and the prediction of MHC class-II binding T-cell epitopes of La/SSB autoantigen, based on the fact that DQ2 and DQ7 are strongly associated with SS and SLE, applying computational methods.

I-98

Statistical Analysis of Domains in Interacting Proteins

Thomas Nye[1], Carlo Berzuini[2], Walter Gilks, Madan Babu, Sarah Teichmann
[1]thomas.nye@mrc-bsu.cam.ac.uk, MRC Biostatistics Unit, Cambridge; [2]carlo.berzuini@mrc-bsu.cam.ac.uk, MRC Biostatistics Unit, Cambridge

We present a method for the analysis of large-scale datasets of physical interactions between proteins. By incorporating information about protein structure, it is possible to use these datasets to learn which structural elements of proteins are "sticky", and furthermore make simple predictions about the structure of protein complexes.

I-99

Seeking the Vertebrate Secretome

Carlos P. Sosa[1], Erick W. Klee[2], S. Ekker & L. B. Ellis
[1]cpsosa@us.ibm.com, IBM & University of Minnesota;
[2]klee0025@umn.edu, University of Minnesota

We are developing csP, a technique that uses protein domain classification instead of signal sequence identification to predict secreted proteins. Protein domains reported to be more prevalent in the mouse secretome than in other mouse proteins by Grimmond and co-workers [Genome Research 13: 1350 (2003)] are used as our reference.

I-100

GPCR Ligand Determination Using SVM

Osman Sezerman[1], Altuna Akalin[2], Zeynep Kasap, Ersen Kavak
[1]ugur@sabanciuniv.edu, Sabanci University;
[2]altuna@sabanciuniv.edu, Sabanci University

G Protein Coupled Receptors (GPCRs) are cell membrane proteins that play a major role in signal transduction pathway. They regulate many physiological processes upon binding of a ligand. We predict the ligand specifity for Class A, B and C type of GPCRs using SVMs with 87.5 %, 100 % and 100% accuracy respectively.

I-101

STRING - Predicting Protein Networks from Genomic Context and External Data

Lars Juhl Jensen[1], Christian von Mering[2], Peer Bork
[1]jensen@embl.de, EMBL; [2]mering@embl.de, EMBL

The web-based database STRING (http://string.embl.de) is a large, pre-computed resource that allows any protein of interest to be placed into a high-confidence network of functionally associated protein partners. For 100+ genomes, the tool evaluates genomic context, protein interaction data, gene co-expression, co-mentioning of genes in literature, and certain database annotations.

I-102

Viruses have Microsatellites too

Milo Thurston[1], Sarah Turner[2], John Burden, Rosemary Hails, Dawn Field
[1]mith@ceh.ac.uk, CEH Oxford; [2]sltu@ceh.ac.uk, CEH Oxford

We have developed open-source software to perform a survey of microsatellites in small genomes (bacteria, organelles, plasmids and viruses). Our data show that long polymorphic repeats are common and are distributed in taxonomically significant patterns. We have investigated 23 baculovirus genomes, which contain long, hypermutable microsatellites, in more detail

I-103

Splice Site Prediction Using SVM with the Oligo Kernel

Leila Taher[1], Burkhard Morgenstern[2], Peter Meinicke
[1]ltaher@techfak.uni-bielefeld.de, University of Bielefeld;
[2]bmorgen@gwdg.de, University of Goettingen

Using support vector machines and a recently introduced kernel, we investigate which oligomer lengths are optimal for splice site prediction. The oligo kernel allows us to represent splice sites by the occurrences of oligomers of any length, and also to model the positional variability of the oligomer occurrences.

I-104

Protein Structure Prediction by Great Deluge Search on an Extended-Atom Model

Yuri Bykov[1], Jonathan D. Hirst[2], Edmund K. Burke
[1]yxb@cs.nott.ac.uk, The University of Nottingham;
[2]jonathan.hirst@nottingham.ac.uk, The University of Nottingham

We present the application of the Great Deluge local search algorithm to protein structure prediction using an extended-atom model. The experiments were done with a 28-residue fragment of a real protein. Results produced in one hour on a PC Pentium 4 2.4GHz matched the native state with RMSD 5.8-7.8 Å.

I-105

In Silico Approaches to RNA Editing

Claudia L. Kleinman[1], Gertraud Burger[2]
[1]cl.kleinman@umontreal.ca, Université de Montréal; [2]Gertraud.Burger@UMontreal.CA, Université de Montréal

RNA editing is the modification of precursor RNAs through the insertion, deletion or specific substitution of nucleotides, to yield functional RNA species. We will review the status of research in this area, and discuss several strategies aiming at the discovery of cis-acting elements and the prediction of potential editing sites.

I-106

Diagnosis of Metabolic Inherited Diseases Using Support Vector Machine Analysis of Metabolome Data

Zaccaria Paolo[1]
[1]paolo.zaccaria@biocrates.at, Biocrates GmbH

A support vector machine has been tested for analysis of mass-spectrometry data, applied to prediction of newborn metabolic disorders. First results show that machine learning techniques could provide a promising approach to discover new classification rules for diagnosis of metabolic disorders

I-107

Computational Identification of MicroRNA Targets in C. Elegans

John Tsang[1], John Kim[2], Gary Ruvkun
[1]tsang@fas.harvard.edu, Graduate Program in Biophysics, Harvard University; [2]jkim@molbio.mgh.harvard.edu, Dept. of Molecular Biology, Massachusetts General Hospital and Dept. of Genetics, Harvard Medical School

MicroRNAs (miRNA) are small RNA molecules with important regulatory functions. Hundreds of miRNAs have been identified but little is known about their regulatory targets. Here we report a computational method to identify targets of highly conserved miRNAs in C. elegans that utilizes information from multiple genomes and protein localization data.

I-108

Artificial Neural Network and Hidden Markov Model for GPI-Anchored Protein Predictions

Guylaine Poisson[1], Anne Bergeron[2], Cedric Chauve , Bouchra Moumen
[1]poisson.guylaine.2@courrier.uqam.ca, UQAM; [2]bergeron.anne@uqam.ca, UQAM

A glycosyl phosphatidyl-inositol (GPI) anchor is a C-terminal post-translational modification of protein. Here, we investigate the problem of correctly annotating GPI-anchored protein for the growing number of sequences in public Databases. We developed a dual approach based on the use and comparison of Neural Network and Hidden Markov Model approaches.

I-109

Prediction of Protein-protein Binding Sites Using Support Vector Machines

James Bradford[1], David Westhead[2]
[1]bmbjrb@bmb.leeds.ac.uk, University of Leeds; [2]westhead@bmb.leeds.ac.uk, University of Leeds

We have applied an increasingly popular machine-learning approach, the support vector machine (SVM), to the prediction of protein-protein binding sites. Our method classifies surface patches as part of or outside a binding site, and has a high success rate with broad specificity, being able to predict both transient and obligomeric binding sites.

I-110

Mathematical Prediction of Intensity of Prolactin-induced Signal Cascade in Rat Target-tissues

Bogorad R.L.[1], Kilikovsky V.V.[2], Smirnova O.V.
[1]rbogorad@yahoo.com, Biological department, Moscow state university; [2]vkilikov@yandex.ru, Medical information science and cybernetic department, Russian state medical university

An oligomerization of hormone-receptor complexes plays essential role in signalling of many membrane receptors classes. We developed a method for hormone efficacy estimation in the case of hormone-induced oligomerisation of different receptor isoforms and verified the results by experimental data. Prolactin has served an object for verification of the method

I-111

Computational Identification of Transcription Factor Binding Sites with Variable-Order Markov Models

We present a variable-order Markov model as a natural extension of fixed-order Markov models including the position weight matrix model for the recognition of transcription factor binding sites.

J-1

ThurGood: Evaluating Assembly-to-Assembly Mapping

Hagit Shatkay[1], Jason Miller[2], Clark Mobarry, Michael Flanigan,Shibu Yooseph,Granger Sutton
[1]shatkay@cs.queensu.ca, School of Computing, Queens' University; [2]jason.miller@celera.com, Celera/Applied Biosystems

The alignment of large genomic sequences is the focus of much recent research. However, current methods for validating whole-genome alignment remain wanting. We introduce ThurGood, a suite of tools for evaluating genome-alignment. We demonstrate its use in evaluating several methods for assembly-to assembly mapping, which were recently used to align multiple versions of the human genome.

J-2

Yeast Mutation Rates in Promoters and Coding Sequences: Uniformity and Functional Biases

Chen-Shan Chin[1], Jeff Chuang[2], Hao Li
[1]cschin@genome.ucsf.edu;
[2]jchuang@genome.ucsf.edu,

We study local neutral mutation rates in S. cerevisiae genome with the sequences from four Saccharomyces species. Our analysis indicates that the neutral mutation rate is uniform in S. cerevisiae and we identify functional selection that contributes to the observed bias for both coding regions and promoters

J-3

Using Complexity Measures to Detect and Characterise 16S rRNA Chimeras

Kevin Ashelford[1], Andrew Weightman[2], Nadia A. Chuzhanova, John C. Fry, Antonia J. Jones
[1]ashelford@cardiff.ac.uk, Cardiff University;
[2]weightman@cardiff.ac.uk, Cardiff University

Systematic bacteriology relies on 16S rRNA data, yet chimeras, formed from the concatenation of gene fragments during PCR, give a false impression of diversity and phylogeny. We are developing a new algorithm which will analyse the complexity of RNA to detect irregularities in secondary structure and so identify chimeras.

J-4

PARALIGN - a Tool for Rapid and Sensitive Sequence Similarity Searches

Torbjørn Rognes[1], Per Eystein Sæbø[2]
[1]torbjorn.rognes@labmed.uio.no, Centre for Molecular Biology and Neuroscience, Rikshospitalet University Hospital and University of Oslo, Norway;
[2]p.e.sabo@labmed.uio.no, Centre for Molecular Biology and Neuroscience, Rikshospitalet University Hospital and University of Oslo, Norway

PARALIGN is a sequence database similarity search tool equal to Smith-Waterman in sensitivity. A speed similar to BLAST has been attained by exploiting parallel processing technology in modern microprocessors. Executables are free of charge for academic use, and online searches powered by a large Linux cluster are available at www.paralign.org

J-5

Combining Sequence and Structural Information in Multiple Sequence Alignment

Orla O'Sullivan[1], Prof. Des Higgins[2], Prof. Cedric Notredame
[1]orla.osullivan@ucd.ie, Conway Institute of Biochemical and Biomedical Research, University College Dublin;
[2]des.higgins@ucd.ie, Conway Institute of Biomedical and Biochemical Research, Unversity College Dublin

3D-Coffee is based on the multiple alignment method T-Coffee. 3D-Coffee generates alignments using a combination of sequence and structural information. In this poster we investigate the use of structural information in generating multiple sequence alignment using 2 structural alignment programs, SAP and FUGUE, in conjunction with 3D-Coffee

J-6

Finding Optimal Pairs of Patterns

Hideo Bannai[1], Heikki Hyyro[2], Ayumi Shinohara, Masayuki Takeda, Kenta Nakai, Satoru Miyano
[1]bannai@ims.u-tokyo.ac.jp, University of Tokyo;
[2]helmu@cs.uta.fi, PRESTO, Japan Science and Technology Corporation (JST)

We present an O(N^2) time algorithm for finding the optimal pair of substring patterns combined with any Boolean function, for discriminating between two sets of strings. The algorithm is applied to datasets of moderate size to find sequence elements that cooperate or compete with each other in fulfilling their functions.

J-7

Microbase: A Grid-based System for Microbial Genome Comparisons

Anil Wipat[1], Yudong Sun[2], Matthew Pocock, Pete Lee, Paul Watson and Keith Flanagan
[1]anil.wipat@ncl.ac.uk, Computing Science, University of Newcastle; [2]yudong.sun@ncl.ac.uk, Computing Science, University of Newcastle

The rapid developments in genome sequencing technology, are moving comparative genome analysis beyond the local capability of many laboratories. The Microbase project is developing a novel Grid-based system to support comparative genomics applications by providing access to pre-computed datasets and systems to permit the execution of remotely conceived and user-defined computations.

J-8

Protein Family Comparison Using Statistical Models and Predicted Structural Information

Richard Y. Chung[1], Golan Yona[2]
[1]rc238@cs.cornell.edu, Cornell University;
[2]golan@cs.cornell.edu, Cornell University

Profile-profile comparisons are used to assess similarity between families of proteins. We present a simple yet effective scheme to augment an existing information theory-based profile-profile comparison method with structural information. We show that predicted structural information, in conjunction with primary amino acid sequence, increases the sensitivity of these protein family comparisons.

J-9

Rényi Entropy of DNA Sequences

Susana Vinga[1], Jonas S Almeida[2]
[1]svinga@itqb.unl.pt, ITQB/UNL;
[2]almeidaj@musc.edu, MUSC

Entropy estimation of DNA sequences provides a measure of their randomness level. Rényi continuous quadratic entropy here proposed generalizes Shannon's formalism allowing more flexibility and extraction of new features. The asymptotic behavior of this new measure is deduced and the results for artificial and real DNA are presented.

J-10

HapAnalyzer: Minimum Haplotype Analysis System for Association Studies

Ho-Youl Jung[1], Jung-Sun Park[2], Yun-Ju Park, Kuchan Kimm, and InSong Koh
[1]hyjung@ngri.re.kr, National Genome Research Institute, National Institute of Health;
[2]pj518@hosanna.net, National Genome Research Institute, National Institute of Health

HapAnalyzer is an analysis system that provides minimum analysis methods for the SNP-based association studies. It consists of Hardy-Weinberg equilibrium (HWE) test, linkage disequilibrium (LD) computation, haplotype reconstruction, and SNP (or haplotype)-phenotype association assessment. It is well suited to a case-control association study for the unrelated population Availability: The HapAnalyzer is freely available from http://www.ngri.re.kr/HapAnalyzer

J-11

BlockSampler: a Methodology for Phylogenetic Shadowing in Prokaryotes

kathleen Marchal[1], Gert Thijs[2], Pieter Monsieurs, Sigrid De Keersmaecker, Jos Vanderleyden, Bart De Moor
[1]kathleen.marchal@esat.kuleuven,ac.be, ESAT/SCD;
[2]Gert.thijs@esat.kuleuven.ac.be, ESAT/SCD

We developed "BlockSampler", a strategy for phylogenetic shadowing in prokaryotes. BlockSampler uses Gibbs sampling to identify motif seeds in the set of orthologous sequences. These seeds are subsequently extended to identify conserved "blocks". We demonstrated the efficiency of our methodology on orthologous sequences from closely related gamma proteobacterial species.

J-12

Evaluation of Scoring Functions and Background Models for Transcription Factor Binding Site Prediction

Markus Friberg[1], Peter von Rohr[2], Gaston Gonnet
[1]friberg@inf.ethz.ch, ETH Zurich, Institute of Computational Science; [2]vonrohr@inf.ethz.ch, ETH Zurich, Institute of Computational Science

Many scoring functions and background models for transcription factor binding site (TFBS) prediction have been proposed in the literature. Here we compare how different scoring functions and background models perform on both real and simulated data sets, given the same search method and TFBS representation

J-13

EMBOSS - European Molecular Biology Open Software Suite

A.J. Bleasby[1], P.M. Rice[2]
[1]ableasby@rfcgr.mrc.ac.uk, Rosalind Franklin Centre for Genomic Research; [2]pmr@ebi.ac.uk, European Bioinformatics Institute

EMBOSS is an open source suite for bioinformatics with a focus on sequence analysis and application integration. EMBOSS is available from http://www.emboss.org.

J-14

Iterative Alignment Algorithms for Multiple Alignment

Iain Wallace[1], Des Higgins[2]
[1]iain.wallace@ucd.ie, UCD; [2], UCD

Various iterative algorithms were implemented using ClustalW and Perl. The algorithms were benchmarked using the HOMSTRAD database. Despite the relative simplicity of the algorithms, they improved the performance of many multiple aligment programs on even the hardest of test cases.

J-15

Benchmark Structural Alignment Validation using RMSD

Gordon Blackshields[1], Des Higgins[2]
[1]gordon.blackshields@ucd.ie, Conway Institute; [2]des.higgins@ucd.ie, Conway Institute

Several protein alignment benchmark Databases exist that are based on superposition of related protein structures. Validation of these superposed structures is critical to validation of the database. This project uses simple Root Mean Square Deviation (RMSD) calculation between aligned residues on a position-wise basis throughout the alignment as a validator.

J-16

MUSCLE: Faster and More Accurate Multiple Sequence Alignment

Robert C. Edgar[1]
[1]bob@drive5.com, (none)

MUSCLE is the most accurate multiple sequence alignment program (up to 5% better than T-Coffee on benchmarks) with speed similar to CLUSTALW. With appropriate options, MUSCLE achieves accuracy equal to T-Coffee and aligns 5,000 sequences of length 350 in 10 minutes. MUSCLE is a free download at http://www.drive5.com/muscle

J-17

Alternative Slicing Using ESTs and QUASAR

Alexander Herrmann[1], Prof. Dr. Knut Reinert[2]
[1]alexander.herrmann@mdc-berlin.de, MDC-Berlin; [2]reinert@inf.fu-berlin.de, FU-Berlin

Identification of the alternative splice events is based on the Sequence Comparison between mRNA/DNA gene's sequence and ESTs. Usually the comparison takes place by BLAST by means of linear scan of EST data bank. A quick search for the alternative splice events could be achieved by indexation of EST data bank by means of gapped q-gram QUASAR algorithm.

J-18

MODERNA (Algorithm for Motif Design of RNA) - An Application for the Design of New and Recombinant Selenoproteins

Anke Busch[1], Rolf Backofen[2]
[1]busch@inf.uni-jena.de, Friedrich-Schiller-University Jena; [2]backofen@inf.uni-jena.de, Friedrich-Schiller-University Jena

MODERNA is an algorithm for Motif Design of RNA (e.g. SECIS-elements) with both structure and sequence constraints within the coding region of an mRNA. Nevertheless, a certain similarity to the original protein is kept. It can be used e.g. for recombinant expression of several selenoproteins in E.coli

J-19

Uses of Bit-Vector Representation of Amino Acids for Sequence Alignement

Delalin[1], Mephu Nguifo[2]
[1]delalin@cril.univ-artois.fr, CRIL; [2]mephu@cril.univ-artois.fr, CRIL

We propose a method of pairwise sequence alignment based on a bit-vector representation of amino acids. An amino acid is represented by a bit-vector where each bit represent an attribute which characterize it and a function of similarity calculus is employed instead of an ASCII character and a substitution matrix.

J-20

A Model for a Context Independent Repair Mechanism

Alexander Roth[1], Gina Cannarozzi[2], Markus Friberg, Peter von Rohr, Gaston Gonnet
[1]*alexande@inf.ethz.ch, ETH Zurich;*
[2]*cannaroz@inf.ethz.ch, ETH Zurich*

We investigate a theoretical model for correcting pairing mistakes in DNA. Organisms are likely to have evolved mechanisms to minimize phenotypic mutations. The model use codon bias to calculate optimal correction rates. The results predict that mismatches in DNA are corrected in a biased manner

J-21

Statistical Significance of Contextual Alignment

Anna Gambin[1], Boguslaw Kluge[2]
[1]*aniag@mimuw.edu.pl, Warsaw University;*
[2]*b.kluge@zodiac.mimuw.edu.pl, Warsaw University*

The contextual alignment model is an extension of the classical alignment, in which the cost of a substitution depends on the surrounding symbols. We discuss the statistical significance of contextual alignment. Our main result is that the distribution of gap-free local contextual alignment scores follows the Extreme Value Distribution.

J-22

QAlign: Quality-Based Multiple Alignments by Complementary Algorithms

Michael Sammeth[1], Jens Stoye[2], Dag Harmsen
[1]*micha@sammeth.net, Genome Informatics, Technical Faculty, Bielefeld University;*
[2]*stoye@techfak.uni-bielefeld.de, Genome Informatics, Technical Faculty, Bielefeld University*

QAlign enhances the investigation and results of multiple sequence analysis by integrating complementary alignment strategies (progressive, simultaneous, iterative and consistency-based), a layout editor and tools deriving phylogenetic trees in a graphical 'multiple alignment environment'. The software is freely available at http://www.ridom.de/qalign/, where distributions for various operating systems are provided.

J-23

Analysis of Gene Regulatory Regions by Means of DNA Composition

Nora Pierstorff[1], Bernhard Haubold[2], Thomas Wiehe
[1]*nora.pierstorff@uni-koeln.de, Univ. Koeln, Germany;*
[2]*bernhard.haubold@fh-weihenstephan.de, Univ. Appl. Sci. Weihenstephan, Germany*

"Shustring" is a suffix-tree based method to determine the distributions of unique substrings and their close variants in a genome. Comparison of expected and observed distributions yields characteristic properties in intergenic, promoter and coding regions. These signatures help to determine gene regulatory regions and to identify putative protein binding sites.

J-24

?-Scanps: A Chained Iterative Protein Sequence Search Procedure

Greg Machray[1], Geoff Barton[2], Mike Ferguson
[1]*greg@compbio.dundee.ac.uk, University of Dundee;*
[2]*geoff@compbio.dundee.ac.uk, University of Dundee*

?-scanps is an Intermediate Search Sequence extension of iscanps, an iterative Needleman-Wunsch search method. It has been developed in order to improve the prediction of relationships between distantly related proteins by identifying 'stepping stones' between query and target sequences.

J-25

Nuclear Genome of Oryza Sativa (Japonica Cultivar-group) Contains Multiple DNA Insertions Covering the Whole Plastid Genome

Y.Yu.Akbarova[1], I.A.Shahmuradov[2], V.V.Solovyov, J.A.Aliyev
[1]*yagut@pgenomics.org, Institute of Botany;*
[2]*ilham@cs.rhul.ac.uk, Royal Holloway,University of London*

BLAST comparison of 12 assembled chromosomes O.sativa (japonica) and plastid DNA revealed plastid-related 900 fragments of nuclear genome, ranging from 100 bp to ~100 kb and totaling ~860 kb, which cover the whole plastid genome several times and include intact nuclear copies of 37 different known protein-coding genes.

J-26

ESPD: a Pattern Detection Model Underlying Gene Expression Profiles

Home-Bo Weng[1]
[1]WengHomeBo@hotmail.com, Department of Computer Science and Engineering, State University of New York at Buffalo

DNA arrays permit rapid, large-scale screening for patterns of gene expression and simultaneously yield the expression levels of thousands of genes for samples. Furthermore, most of the genes collected may not necessarily be of interest and uncertainty about which genes are relevant makes it difficult to construct an informative gene space.

J-27

Cyclin-dependent Kinases of Leishmania: Sequence Analysis and Genomics

Nahla OM Ali[1], Muntaser E. Ibrahim[2], Karen M. Grant; Jeremy C. Mottram
[1]dr_nahla2004@yahoo.com, Faculty of Veterinary Medicine, University of Khartoum, Sudan; [2]muntaser26@hotmail.com, Institute of Endemic Diseases, University of Khartoum, Sudan

A number of cdc2-related kinase genes have been isolated from trypanosomatids. Searching the Databases have led to the identification of CRK3 homologue from Sudanese strain of Leishmania donovani and a novel cyclin from L. mexicana. Web-based resources were used in sequence analysis. The functional interaction of these two was then facilitated.

J-28

PSI-PRALINE: A Novel Algorithm for Multiple Sequence Alignment

Victor A. Simossis[1], Jaap Heringa[2]
[1]vsimoss@cs.vu.nl, VU Amsterdam; [2]heringa@cs.vu.nl, VU Amsterdam

We introduce a new multiple sequence alignment strategy PSI-PRALINE. The algorithm uses PSI-BLAST to increase the information for each sequence in the set to be aligned and allows quality control and filtering of the PSI-BLAST hits. Preliminary results show that the use of the homology-derived information dramatically improves alignment quality.

J-29

Using Quasi Consensus Sequences for Comparison of Profile Hidden Markov Models and Alignment of Protein Sequences

Robel Kahsay[1], Guoli Wang[2], Li Liao, Roland Dunbrack, and Guang Gao
[1]kahsay@mail.capsl.udel.edu, University of Delaware; [2]Fox Chase Cancer Center

A method is proposed for sensitive detection of distant relationships among protein families and for prediction of structural alignment via comparison of hidden Markov models based on quasi consensus sequences. The method gives better homology detection, yields improved alignments, and runs significantly faster, in comparison to a state-of-the-art profile-profile method.

J-30

PRC, the Profile Comparer

Madera[1]
[1]mm238@mrc-lmb.cam.ac.uk, MRC LMB

PRC is a tool for comparison and alignment of two profile hidden Markov models. PRC uses all available transition information and treats both models on an equal footing. PRC can read SAM, HMMER and PSI-BLAST files. Source code distributed under the GPL can be downloaded from http://supfam.org/PRC/

J-31

Binary DNA Tracts are Promoter Elements

Gad Yagil[1]
[1]lcyagil@weizmann.ac.il, The Weizmann Institute

Long DNA tracts composed of only two bases are highly over-represented in sequenced genomes. W-tracts prevail in bacteria and archea. Purine-pyrimidine tracts prevail in eukaryotic chromosomes. Intergenic regions are the richest genic subregion, with promoter regions the richest of all. A role in transcription control is proposed. see BMC genomics (2004) 5:19.

J-32

Exploring Differences in Biosequences by Principal Component Analysis (PCA)

Steinar Thorvaldsen [1], Tor Flå[2]
[1]steinart@math.uit.no, University of Tromsø;
[2]tor@math.uit.no, University of Tromsø

We consider the amino acid compositions of predicted proteins, and we use principal component analysis (PCA) as multivariate methods to explore relevant information from related biosequences. PCA- and PLS-regression both with real valued variables (like growth temperature), and categorical variables (like EC-numbers), are also examined.

J-33

Distinguishing Chromosome I from II Using Genomic Signatures: a General Rule for Bacteria

Hull, K[1], Young, J.P.W[2]
[1]khh103@york.ac.uk, University of York;
[2]jpy1@york.ac.uk, University of York

Many bacteria harbour large second replicons additional to their main chromosome and plasmids. These replicons carry some essential genes and have a GC-content to match the main chromosome, yet have plasmid-like replication systems. We have used dinucleotide frequencies and multivariate statistics to find a common rule for distinguishing these replicons

J-34

BUSSUB: a Virtual Amplicon Retrieval Software

JP. Sanchez-Merino[1], G. Lopez-Campos[2], I. Spiteri, F. Martin-Sanchez
[1]jpsanchez@iscii.es, ISCIII; [2]glopez@isciii.es, ISCIII

BUSSUB is a new tool for retrieving sequences between two defined flanking regions. It is a useful web/desktop application with wide range applications within molecular biology such as PCR design or phylogenetic studies. The software works using FASTA format for input and output files, which simplifies the interoperability with other sequence analysis software or applications.

J-35

Statistical Score for Assessing Quality of Multiple Sequence Alignments

Virpi Ahola[1], Tero Aittokallio[2], Mauno Vihinen, Esa Uusipaikka
[1]virahola@utu.fi, Department of Statistics, University of Turku; [2]tero.aittokallio@utu.fi, Department of Mathematics, University of Turku

The poster introduces a statistical score for assessing the quality of multiple sequence alignments. The method is based on the calculation of observed significance level for a type of Z-score. Comparisons revealed that the quality score is useful in evaluating multiple sequence alignment programs.

J-36

DPCONTACTS: A Novel Method to Align Protein Structures Using only Residue Contact Information

Michael Hurley[1], Dr. Jon Ison[2]
[1]mhurley@rfcgr.mrc.ac.uk, RFCGR, Hinxton, Cambridge; [2]jison@rfcgr.mrc.ac.uk, RFCGR, Hinxton, Cambridge

Residues involved in folding form can show evolutionary conservation of residue contacts. We describe a novel pairwise structure-based sequence alignment algorithm based on residue contact information. This algorithm is called DPCONTACTS and usesa double dynamic procedure to align proteins solely on the basis of contact information

J-37

Analysis of Structure and History of Complex Repetitive Genome Regions

Amy M. Hauth[1], Gertraud Burger[2], B. Franz Lang
[1]amy.hauth@umontreal.ca, Centre Robert Cedergren, Centre de Recherche en Bioinformatique et en Sciences G&eacute;nomiques de l'Université de Montréal; [2]gertraud.burger@umontreal.ca, Centre Robert Cedergren, Centre de Recherche en Bioinformatique et en Sciences Génomiques de l'Université de Montréal

We study repetitive genome regions having both direct and inverted repeats. Our research interest focuses on regions exhibiting complex similarity patterns with the goal of characterizing internal similarities as well as inferring their history of formation via duplication, mutation and recombination events (http://megasun.bch.umontreal.ca/People/ahauth/tools)

J-38

Identification of SNPs and Establishiment of SNP Markers in Rice

Lee Gang-seob[1], Kim Yong-hwan[2], Yoon Ung-han, Lee Jung-sook, Hyun Do-yoon, Hahn Jang-ho and Kim Ho-il
[1]kangslee@rda.go.kr, National Institute of Agricultural Biotechnology; [2]yghnkim@rda.go.kr, National Institute of Agricultural Biotechnology

To produce a set of SNP markers from whole genome of rice, we compared the genome sequences between indica and japonica rice. SNPs in Xa1 gene region were investigated with comparing sequences from 16 varieties (six resistant, eight susceptible and two unknowns

J-39

Global Propensity of Amino Acids in Prion Proteins

Zarrin Minuchehr[1], Majid Erfani[2], Shiva Fatollahi Rad, Tolue Mahdavi, Mandana Pourian, Fatameh Sadat Sabet, Atiya Taghi, Bahram Goliaei
[1]minuchhr@nrcgeb.ac.ir, National Research Center for Genetic Engineering and Biotechnology, Tehran Iran; [2]erfani_m@yahoo.com, Tarbiat Modarress Univeristy Tehran Iran

Prions are transmissible pathogens that cause fatal Neurodegenerative diseases, in this study we determined the propensity of amino acids in prion proteins. The results showed that Glu, Asp and Leu show low and Trp and Gly show a strikingly high propensity.

J-40

Multivariate Analysis of the 2 Oxoglutarate (2OG) and Fe(II)-dependant Oxygenases to Study Distinct Functional Patterns

Ashwin Sivakumar[1], Liisa Holm[2]
[1]ashwin.sivakumar@helsinki.fi, Structural Genomics group, University of Helsinki; [2]holm@ebi.ac.uk, Structural Genomics group, University of Helsinki

This family contains members of the 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily including the Ethylene forming enzymes of both plant and bacterial origin and other proteins of biological significance. Here we do a detailed multivariate study of the family in the quest for improved annotations and demarcation of functional diversities within this important biological family of proteins.

J-41

IHMMune-align: Hidden Markov Model-Based Alignment Tool for Immunoglobulin Gene Sequences

Harald R. Malming[1], Andrew M. Collins[2], Mike E. Bain, Katherine L. Jackson, Bruno A. Gaeta
[1]z3104176@student.unsw.edu.au, School of Computer Science and Engineering, UNSW; [2]a.collins@unsw.edu.au, School of Biotechnology and Biomolecular Sciences, UNSW

IHMMune-align is an alignment tool designed specifically for modelling the antibody generation process. The program uses a combination of dynamic programming and hidden Markov models to identify the constituent germline gene segments of a mature immunoglobulin gene sequence as well as nucleotides inserted and deleted at the gene segment junctions.

J-42

Mixed Approach for Multiple Alignment Based on Pyramidal Classification

Laure Vescovo[1], Jean-Christophe Aude[2], Géraldine Polaillon, Jean-Loup Risler
[1]laure.vescovo@supelec.fr, Supélec, Plateau de Moulon, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette Cedex; [2]jean-christophe.aude@cea.fr, CEA-DBJC-SBGM, CEA Saclay, 91191 Gif-sur-Yvette Cedex

ClustalW still remains the most used algorithm for its compromise between computing needs and accuracy. We propose to develop a progressive algorithm of the Clustal type with the advantages of a mixed strategy, i.e. combining local and global alignments. For this purpose, we use pyramidal classification to compute the guide structure.

J-43

Protein Annotation by Secondary StrucTure Based Alignments

Constantin Bannert[1], Jens Stoye[2]
[1]Bannert@CeBiTec.Uni-Bielefeld.de, Genome Informatics, Bielefeld University; [2]Stoye@CeBiTec.Uni-Bielefeld.de, Genome Informatics, Bielefeld University

PASSTA is a software tool that helps biologists in deciding whether a protein sequence query is related to a protein with known structure, and how well the query can be represented by sequence fragments from PDB proteins. We present the basic version of the tool and a few usage examples.

J-44

Align-m - a New Algorithm for Multiple Alignment of Highly Divergent Sequences

Ivo Van Walle[1], Ignace Lasters[2], Lode Wyns
[1]ivwalle@vub.ac.be, Vrije Universiteit Brussel;
[2]ignace.lasters@algonomics.com, AlgoNomics

Align-m uses a truly multiple local alignment strategy that requires O(N2L2) time and memory. It was developed primarily for aligning highly divergent sequences, for which it frequently has significantly better average accuracy than ClustalW, T-Coffee and DiAlign, on a testset covering the entire SCOP database. Align-m is available from http://bioinformatics.vub.ac.be

J-45

3DMA - A Fast and Accurate Multiple Structure Alignment Algorithm

Azat Badretdinov[1], Tom Oldfield[2], Paul Flook, Lisa Yan
[1]azat@accelrys.com, Accelrys, Inc;
[2]oldfield@ebi.ac.uk, EBI

We present a new structure alignment method, 3DMA, which uses a fast algorithm to determine the initial seed alignment based on a hash table. Multiple structure alignment involves an iterative addition of input protein chains based on suboptimal pairwise alignment. We provide details of the altorithm and of work to validate this method.

J-46

Reconciling the Numbers: ESTs Versus Protein-Coding Genes

Anton Nekrutenko[1]
[1]anton@bx.psu.edu, PennState

The number of expressed sequences greatly surpasses the estimated number of protein-coding genes in mammalian genomes. An evolutionary approach reveals that only 9-14% of human and mouse expressed sequences are able to code for proteins. Clustering of these sequences using cross-species relationships suggests that millions of expressed sequences may correspond to only ~20,000 distinct protein-coding transcripts.

J-47

The Extension of Group-to-group Sequence Alignment Algorithm under a Piecewise Linear Gap Cost

Shinsuke Yamada[1], Osamu Gotoh[2], Hayato Yamana
[1]shinsuke@yama.info.waseda.ac.jp, Department of Computer Science, Graduate School of Science and Engineering, Waseda University, Japan;
[2]o.gotoh@aist.go.jp, Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Japan; Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Japan

Prrn is one of the programs for multiple sequence alignment. We extend the group-to-group sequence alignment algorithm of Prrn so that it can deal with a piecewise linear gap cost. This poster shows the extension of the algorithm and evaluates alignment accuracy using BAliBASE benchmark.

J-48

The Relationship between Multiple Sequence Alignment and the Quality of Protein Comparative Models

Domenico Cozzetto[1], Anna Tramontano[2]
[1]domenico.cozzetto@uniroma1.it, Dept. Biochemical Sciences University La Sapienza;
[2]anna.tramontano@uniroma1.it, Dept. Biochemical Sciences University La Sapienza

The quality of a comparative model depends upon the structural divergence and the quality of the (multiple) sequence alignment between target and template. We describe how to evaluate the accuracy of a comparative model given the number and similarity distribution of the sequences in the multiple sequence alignment.

J-49

Analysis of Primate microRNA Genes by Phylogenetic Shadowing

E. Berezikov[1], V. Guryev, R.H.A. Plasterk. E. Cuppen
[1]berezikov@niob.knaw.nl, Hubrecht Laboratory

MicroRNAs (miRNAs) are a novel class of non-coding RNAs that regulate expression of genes at posttranscriptional level. To reveal conservation patterns in miRNA genes, we have sequenced genomic regions encompassing 122 miRNAs in 10 primate species (approach know as phylogenetic shadowing) and revealed several interesting features, including the discovery of primate-specific miRNA genes.

J-50

Sequence Similarity Searching at the EMBL-European Bioinformatics Institute (EBI)

Nicola [1], Rodrigo Lopez[2], Karyn Duggan, Asif Kibria, Adam Lowe, Gulam Patel, Sharmila Pillai, Emmanuel Quevillon, Stephen Robinson, Ville Silventoinen, Brendan Vaughan
[1]nharte@ebi.ac.uk, EBI; [2]rls@ebi.ac.uk, EBI

There are five main similarity search algorithms available at the EBI: FastA, WU-BLAST2, NCBI-BLAST2, MPsrch and ScanPS. These tools can be used to search against a large number of biological Databases. They are available from the EBI website at http://www.ebi.ac.uk/services/ and via the mail server.

J-51

ReHAB: A Tool for Finding New Hits in BLAST Searches

Joe Whitney[1], Chris Upton[2]
[1]aquinas@uvic.ca, University of Victoria, Victoria, BC Canada; [2]cupton@uvic.ca, University of Victoria, Victoria, BC Canada

ReHAB (Recent-Hits-Acquired-from-BLAST) compares results from PSI-BLAST searches performed with different NR Databases and highlights new hits. ReHAB imports queries from Databases or files into its own mySQL database, automates search runs and saves new hits. Results are presented in easily comprehended table that uses colors to hightlight new hits.

J-52

GeMoDA: Generic Motif Discovery Algorithm

Kyle L. Jensen[1], Mark P. Styczynski[2], Isidore Rigoutsos, Gregory N. Stephanopoulos
[1]kljensen@mit.edu, MIT; [2]marksty@mit.edu, MIT

GeMoDA is an exhaustive motif discovery algorithm that returns, from sequential data, the set of all patterns meeting some statistical criterion. Various similarity and clustering metrics can be supplied by the user to allow a wide range of discovery criteria. Applications include DNA motif and protein secondary structure motif identification.

J-53

Align-m - a New Algorithm for Multiple Alignment of Highly Divergent Sequences

Ivo Van Walle[1], Ignace Lasters[2], Lode Wyns
[1]ivwalle@vub.ac.be, Vrije Universiteit Brussel; [2]ignace.lasters@algonomics.com, AlgoNomics NV

Align-m uses a non-progressive local approach, from which a global alignment is constructed. Performance is compared with ClustalW, T-Coffee and DiAlign on a testset covering the entire SCOP classification with sequence identities ranging from 0-50%. In general, Align-m aligns the same or more residues correctly but significantly less residues incorrectly.

J-54

Greedy Motif Search: Local Multiple Alignment Using Similarity

Wei-Mou Zheng[1]
[1]zheng@itp.ac.cn, Inst. Theor. Phys., Academia Sinica

The greedy algorithm is for motifs described by the weight matrix models. The block of the highest cumulative similarity score of BLOSUM is used to derive a modified blosum matrix. Blocks of high total scores using this similarity matrix gives the optimal. Techniques to reduce computing are also discussed.

J-55

Towards Feature Based Profile HMMs

Thomas Plötz[1], Gernot A. Fink[2]
[1]tploetz@techfak.uni-bielefeld.de, Bielefeld University, Technical Faculty; [2]gernot@techfak.uni-bielefeld.de, Bielefeld University, Technical Faculty

In order to improve the classification performance for remote homology detection we present approaches for feature bases enhancements of Profile HMMs. the discrete emissions of the model are replaced by feature vectors calculated for protein data. First promising results of an evaluation on the SCOP database are presented.

J-56

Distinguishing Genomic Sequences from Different Species Based on Sequence Feature

Xiao Sun[1], Jianming Xie[2], Jing Fu, Xueying Xie, Zuhong Lu
[1]xsun@seu.edu.cn, Key Laboratory of Molecular and Biomolecular Electronics (Southeast University); [2]xjm@seu.edu.cn, Key Laboratory of Molecular and Biomolecular Electronics (Southeast University)

We present a new approach to compare or distinguish genomic sequences by using the sequence feature, base-base correlation (BBC). We analyzed 80 large genomic sequences of seven species and found that BBCs of sequences from the same genome are very close. BBC can be used as a genomic signature.

J-57

Searching for Significant Protein Residues Using Entropy Based Clustering of Multiple Sequence Alignments

Boris Reva[1], Chris Sander[1]
[1]Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, NY, 10021, USA

We have developed a new method for determining the specificity residues in protein families from multiple sequences alignments. Validation tests on 12 protein-protein and protein-DNA/RNA complexes showed that the method accurately identifies functionally important residues from sequence information alone, without the use of 3D structure or functional annotation.

J-58

Maximal Information Content Seeds for Sequence Alignment

Michael Brudno[1], Pavel S. Novichkov[2], Andrey A. Mironov[2]
[1]brudno@CS.Stanford.EDU, Stanford University, Department of Computer Science; [2]Moscow State University, Department of Bioengineering &amp; amp; Bioinformatics

Using a variable length approach we finds seeds for local alignment with a pre-specified frequency cutoff, irregardless of their length. A trial comparison of two bacterial genomes suggests that our algorithms can reduce the number of seeds used for BLAST by a factor of four, while considering short, rare words that would usually be missed.

J-59

IBP-BLAST: Using Logistical Networking to Distribute BLAST Databases over the Internet

Ravi Kosuri[1], Jay Snoddy[2], Stefan Kirov[2], Erich Baker[3]
[1]Department of Computer Science, Baylor University, Waco, TX, 97356 USA; [2] Graduate School in Genome Science and Technology, University of Tennessee-Oak Ridge National Laboratories, Oak Ridge, TN 37831, USA; [3]Erich_Baker@Baylor.edu, Department of Computer Science, Baylor University, Waco, TX, 97356 USA

We describe a methodology to distribute BLAST Databases over a wide area network by leveraging the distributed storage technologies developed for the field of logistical networking. This approach allows limited resource clients access to complete remotely maintained Databases, query caching, batch processing, and interoperability with locally developed software.

J-60

Indexing Biological Sequences to Support Hypertension Research

Susan Fairley[1], Ela Hunt[2], Rob Irving and Anna Dominiczak
[1]sf@dcs.gla.ac.uk, Department of Computing Science, University of Glasgow; [2]ela@dcs.gla.ac.uk, Department of Computing Science, University of Glasgow

Hypertension research makes extensive use of the rapidly growing volume of sequence data. Indexing can provide a means of structuring the data that enables the sequences to be queried efficiently. We are developing the suffix sequoia index as a persistent index to support sequence analysis in hypertension research.

K-1

Protein Binding Sites Seen From a Ligand-Perspective

Pal, Debnath[1], Weiss, Manfred S.[2], Suehnel, Juergen
[1]dpal@mbi.ucla.edu, UCLA Molecular Biology Institute;
[2]msweiss@embl-hamburg.de, EMBL Hamburg Outstation

A new method to detect putative ligand binding sites in proteins is described, that utilizes structural features of the recently discovered nest/egg motif.

K-2

A Model Structure of the Gap-junction Transmembrane Domain Specifying C-alpha Positions

Sarel J. Fleishman[1], Nir Ben-Tal[2], Vinzenz M. Unger, Mark Yeager
[1]sarel@post.tau.ac.il, Tel-Aviv University;
[2]bental@ashtoret.tau.ac.il, Tel-Aviv University

A set of methods for predicting the orientations of transmembrane alpha helices was developed and applied to predicting the gap junction's structure. The model provides a structural basis for understanding the different physiological effects of almost 30 mutations and polymorphisms, revealing an intimate relationship between molecular structure and disease.

K-3

An Evolutionarily Conserved Network of Amino Acids Mediates Voltage Gating in Potassium Channels

Sarel J. Fleishman[1], Nir Ben-Tal[2], Ofer Yifrach
[1]sarel@post.tau.ac.il, Tel-Aviv University;
[2]bental@ashtoret.tau.ac.il, Tel-Aviv University

A novel technique for detecting correlated amino-acid positions in proteins was applied to study voltage-dependent gating of potassium channels. Viewed on the structure of the Kv channel a network of intercorrelated positions disclosed correlations between residues in the voltage sensor and the pore, including regions that are involved in gating

K-4

Comparison of Two Different Definitions of Protein-protein Interaction Interface

Bingding Huang[1], Michael Schroeder[2]
[1]bhuang@mpi-cbg.de, Technical University of Dresden;
[2]ms@mpi-cbg.de, Technical University of Dresden

We compare two different definitions of protein-protein interaction interface.Our comparison results show that these two types of definitions are strongly correlated and the interface residues obtaining from these two definitions are nearly identical. The statistic derived from our comparison can be useful for the prediction of protein-protein interaction sites.

K-5

Evolution of Tetratricopeptide Repeats

Manjunatha R. Karpenahalli[1], Johannes Söding [2], Jörg Martin and Andrei N. Lupas
[1]manju@tuebingen.mpg.de, Max-Planck-Institute for Devlopmental Biology;
[2]johannes.soeding@tuebingen.mpg.de, Max-Planck-Institute for Devlopmental Biology

Tetratricopeptide repeat (TPR) is formed of two or more stacked aa-hairpins. TPR-like domains might have arisen from the repetition of protein fragments. We have identified several promising aa-hairpins in globular proteins that are similar in sequence and structure, from which we are interested to engineer perfectly repetitive TPR proteins.

K-6

Computing Structural Differences on Prion Proteins due to Single Nucleotide Polymorphisms Using the STING MILLENNIUM Suite

Diana M. Oliveira[1], Ana Carolina L. Pacheco[2], Sonia M. P. Oliveira, Nilo B. Diniz, Daniel A. Viana, João J. S. Gouveia, Elton J. R. Vasconcelos, Michely C. Diniz, Marianna C. Albuquerque, Thiago D. Ferreira, Gregory B. Clark, Jennifer Tsai, Bhooma Thiruvahindrapuram
[1]diana.oliveira@utoronto.ca, University of Toronto;
[2]nugen-l@lcc.uece.br, Universidade Estadual do Ceara

We have considered the problem of computing structural differences on prion proteins (PrP) using the STING MILLENIUM Suite to provide a compilation of newly predicted PrP models, computed from candidate gene polymorphisms (SNPS, single nucleotide polymorphisms) and represented as accountable for potential conformational changes in PrP proteins.

K-7

Target Selection Informatics Resource for Structural Genomics

Ana Rodrigues[1], Guy G Dodson[2], Roderick E Hubbard
[1]rodrigues@ysbl.york.ac.uk, University of York;
[2]ggd@ysbl.york.ac.uk, University of York & NIMR

An informatics target selection resource for structural genomics projects is presented. The resource facilitates the selection and prioritization of candidate proteins for structural determination, by enabling structural biologists to select targets from their genomic sequences of interest, and according to their own research needs URL: http://www.ysbl.york.ac.uk/~rodrigues/

K-8

Modeling Side Chain Conformations Using Contact Surfaces and Solvent Accessible Surface

Eran Eyal[1], Vladimir Sobolev[2], Rafael Najmanovich, Brendan J. McConkey, Marvin Edelman
[1]eran.eyal@weizmann.ac.il, Weizmann Institute of Science; [2]vladimir.sobolev@weizmann.ac.il, Weizmann Institute of Science

Contact surface area and chemical properties of atoms form the core of a scoring function to concurrently predict conformations of amino acid sidechains. The program (http://sgedg.weizmann.ac.il/sccomp.html) combines accuracy and speed. Most atoms prefer intramolecular surface contact over solvent contact. This might be the driving force for maximizing protein packing.

K-9

Interface Cores in Sandwich-like Proteins

Vladimir Potapov[1], Vladimir Sobolev[2], Marvin Edelman, Alexander Kister, Israel Gelfand
[1]vladimir.potapov@weizmann.ac.il, Weizmann Institute of Science; [2]vladimir.sobolev@weizmann.ac.il, Weizmann Institute of Science

We define protein-protein interface and domain cores for immunoglobulins containing about half the residues and surface areas of the full interfaces. Residues of the two cores are structurally connected, imparting rigidity at the interface core. The rule of positional connectivity extends generally to sandwich-like proteins interacting in a sheet-sheet fashion.

K-10

Protein Structure and Evolutionary History Determine Sequence Space Topology

Boris Shakhnovich[1], Charles DeLisi[2], Eric Deeds, Eugene Shakhnovich
[1]borya@bu.edu, Boston University;
[2]delisi@bu.edu, Boston University

In this work we focus mostly on the effect of the physical constraints imposed on the structure encoded by sequences of the gene family. We show that variability in these constraints represents difference in potential for sequence diversity of gene families.

K-11

Cross-Strand Disulphides in Cell Entry Proteins: Poised to Act

Merridee A Wouters[1], Ken K Lau[2], Phillip J Hogg
[1]m.wouters@victorchang.unsw.edu.au, Victor Chang Cardiac Research Institute;
[2]k.lau@victorchang.unsw.edu.au, Victor Chang Cardiac Research Institute

Cross-strand disulphides (CSDs) link adjacent beta-strands. A search for CSDs in a representative set of the PDB showed their dihedral strain energies are higher than disulphides in general. Conspicuously, CSDs are over-represented in molecules involved in cell entry and there is evidence suggesting their involvement in cell entry events.

K-12

Protein-Protein Docking Simulations with Local Backbone Flexibility

maxim totrov[1]
[1]max@molsoft.com, molsoft llc

Induced fit remains a principal obstacle for the accurate protein-protein docking. Internal Coordinate Mechanics methodology was adapted to efficiently handle full flexibility of a segment of the polypeptide chain and tested in docking with fully flexible binding loop. The method may allow accurate simulations for systems involving highly flexible regions.

K-13

A Multipoles-based Method for the Comparison of Protein Structures

Apostol Gramada[1], Philip E. Bourne[2]
[1]agramada@sdsc.edu, San Diego Supercomputer Center, University of California San Diego; [2]bourne@sdsc.edu, School of Pharmacology and San Diego Supercomputer Center, University of California at San Diego

We present a method for the comparison of protein structures using a hierarchical set of shape descriptors. It is based on the use of a multipolar representation of the quantitative property of interest. Possible metrics for protein similarity calculation are discussed and illustrated with test calculations.

K-14

MMD: A Macro-Molecular Docking Program for Locating Domain-Domain Interactions

Qingjuan Gu[1], William W. Reenstra[2], John J. Rux
[1]qingjuan@wistar.upenn.edu, Wistar Institute; [2]reenstra@mail.med.upenn.edu, University of Pennsylvania

MMD is a macro-molecular docking program specifically designed to take advantage of distributed computing resources so that multiple docking solutions can be evaluated in parallel on a wide range of hardware. The ready availability of inexpensive PC's and open-source software makes these types of studies feasible on a limited budget.

K-15

Prediction of Oligosaccharide 3D Structure using Genetic Algorithms

Abraham Nahmany[1], Francesco Strino[2], Jimy Rosen, Per-Georg Nyholm, Graham Kemp
[1]avi2001@hotmail.com, Gothenburg University; [2]frs@interfree.it, Gothenburg University

In the present study, we have implemented a system called GLYGAL for the prediction of oligosaccharides 3D structures. GLYGAL performs conformational searches using a parallel genetic algorithm. The searches are performed in the torsion angle conformational space and energy calculations are performed using the MM3(96) force field.

K-16

Visualization and Presentation of 3-dimensional Catalytic-mechanism Models: The Case of Alpha-amylase

Nozomi Nagano[1]
[1]n.nagano@aist.go.jp, National Institute of Advanced Industrial Science and Technology (AIST)

To elucidate detailed catalytic functions of enzymes, 3-dimensional models of the catalytic mechanisms should be constructed and presented. To present catalytic-mechanism models, we have devised a method for visualizing protein structures and their interactive functions. By its use, such a model for alpha-amylase will be presented.

K-17

A Web-Based Protein Comparison System Using Structural Classification Information

Su-Hyun Lee[1], Jin-Hong Kim[2], Geon-Tae Ahn, and Myung-Joon Lee
[1]suhyun@sarim.changwon.ac.kr, Changwon National University, South Korea; [2]avenue@ulsan.ac.kr, University of Ulsan, South Korea;

WS4E, a web-based protein structure comparison system using structural classification information, searches common substructures among proteins in PSAML database. For fast comparison, we developed an algorithm for constructing the compatibility graphs and applying SCOP's HMM to decreasing the number of candidate protein structures to be compared in the PSAML database.

K-18

Hybrid Molecular Dynamics: Coupling Atomistic and Hydrodynamics Descriptions on the GRID

Gianni De Fabritiis[1], Rafael Delgado Buscalioni[2], P.V. Coveney and S. Barsky
[1]g.defabritiis@ucl.ac.uk, Chemistry Department, Centre for Comptational Science, University College of London; [2]r.delgado-buscalioni@ucl.ac.uk, Chemistry Department, Centre for Comptational Science, University College of London

We are presenting a coupled molecular dynamics and hydrodynamics model referred as hybrid molecular dynamics which opens a larger window on non-equilibrium MD simulations. We developed the hybrid MD deploying the coupling models (MD and Hydrodynamics) in a GRID environment using RealityGrid framework (UK e-science project).

K-19

Conformational Studies on the Exo-polysaccharide of the Burkholderia Cepacia Using Genetic Algorithms and Molecular Mechanics

Francesco Strino[1], Abraham Nahmany[2], Jimmy Rosen , Graham Kemp Per-Georg Nyholm
[1]frs@interfree.it, Gothenburg University;
[2]avi2001@hotmail.com, Gothenburg University

Conformational studies were performed on the repeating unit of the exo-polysaccharide of Burkholderia cepacia which is believed to play a role in colonization of these bacteria in the lungs of cystic fibrosis patients. The conformational search using genetic algorithm (GLYGAL) coupled to molecular mechanics MM3(96), shows a well-defined linear conformation.

K-20

Inference in Graphical Models for Side-chain Prediction

Chen Yanover[1], Ori Shachar[2], Yair Weiss
[1]cheny@cs.huji.ac.il, School of Computer Science and Engineering ; [2]orish@cs.huji.ac.il, School of Computer Science and Engineering

Inference in graphical models is a widely studied problem in computer science. We apply various inference algorithms to the side chain prediction problem - both for finding the M lowest energy configurations and characterizing the variability in configurations - and show excellent results. An extension to protein-peptide binding prediction gives promising results.

K-21

Development of a Scoring Function for Protein-protein Docking Based on Physical Energy Protentials and Machine Learning

Oliver Martin[1], Dietmar Schomburg[2]
[1]Oliver.Martin@uni-koeln.de, University of Cologne;
[2]D.Schomburg@uni-koeln.de, University of Cologne

A scoring function for protein-protein docking based on up to 21 different physico-chemical energy potentials has been developed using the GRID method (Goodford 1985). Based on these potentials, machine learning algorithms are trained to distinguish near-native complex structures from false solutions as generated by rigid-body FFT docking algorithms.

K-22

BioPySDS: an Object-oriented Interface to Manage Protein Dynamics in an Evolutionary Framework

Alessandro Pandini[1], Giancarlo Mauri[2], Laura Bonati
[1]alessandro.pandini@unimib.it, DISAT - Università degli Studi di Milano-Bicocca;
[2]giancarlo.mauri@unimib.it, DISCO - Università degli Studi di Milano-Bicocca

A new approach to computational investigation of protein function is presented: different levels of information (structural, dynamical and evolutionary) are combined and compared to gain deeper insight in the function of protein superfamilies. Design, implementation and application of the new object oriented program BioPySDS are presented.

K-23

Automatic Structural Modelling for Peptide/MHC Complexes

Kaas Q.[1], Chiche L.[2], Lefranc M.-P.
[1]kaas@ligm.igh.cnrs.fr, Laboratoire d'ImmunoGénétique Moleculaire, Université Montpellier II, UPR CNRS 1142 IGH, 141 rue de la Cardonille, F-34396 Montpellier, France; [2]chiche@cbs.cnrs.fr, Centre de Biochimie Structurale, UMR 5048 CNRS INSERM Université Montpellier I, Faculté de Pharmacie, 15 avenue Charles Flahault, F-34093 Montpellier, France

We have implemented a specific structural modelling method for peptide/MHC complexes. The less conserved regions are modelled with a "dead-end" search algorithm in the phi/psi dihedral angle space and a progressively more complex protein representation.

K-24

Regulation of Protein-Gated Electron Transfer by Inter-subunit Hydrogen Bonding and packing

Ilan Samish[1], Oksana Kerner, David Kaftan, Avigdor Scherz
[1]ilan.samish@weizmann.ac.il, Weizmann Insititute of Science

Protein gating in membrane proteins play key functional roles. Combining in-silico and in-vivo combinatorial mutagenesis we demonstrate that in photosynthetic reaction-centers the phenomenon is regulated by weak inter-subunit H-bonds and packing at the protein core. This conserved mechanism and adjustment to working temperature may be relevant to other membrane proteins.

K-25

"Life Core", the Program for Classification of 3D Structures of Macromolecules

Andrei Alexeevski[1], Mikhail Gribkov[2], Sergei Spirin[3]
[1]aba@belozersky.msu.ru, Belozersky Institute; [2]lone_strider@mtu-net.ru, MEPhI; [3]sas@belozersky.msu.ru, Belozersky Institute

We developed a program "Life Core" based on an original algorithm. The program identifies a maximal set of commonly disposed atoms (a geometric core) in a family of 3D structures of macromolecules, calculates similarity values for each pair of structures, and identifies structural subfamilies. The program is available at http://www.dpidb.belozersky.msu.ru:8080/

K-26

Membrane Protein Folding: Centrality of Backbone-Mediated Interhelical Hydrogen Bonds

Eran Goldberg[1], Avigdor Scherz[2], Ilan Samish
[1]eran.goldberg@weizmann.ac.il, Weizmann Institute of Science; [2]avigdor.scherz@weizmann.ac.il, Weizmann Institute of Science

Statistical analysis of membrane protein structures shows that weak, backbone-mediated interhelical hydrogen bonds are laterally clustered in the central, buried, and conserved parts of transmembrane proteins and correlate well to the amino acid packing scale, thus providing new insight into membrane protein folding, maintenance of functional flexibility and structure prediction.

K-27

Analysis of Conformational Changes on Protein-Protein Complexes

Raul Mendez[1], Giacomo De Mori[2], Didier Croes, Shoshana J. Wodak
[1]raul@scmbb.ulb.ac.be, SCMBB; [2]giacomo@scmbb.ulb.ac.be, SCMBB

An exhaustive analysis of protein - protein conformational changes is presented on protein complex structures deposited at the PDB. Secondary structure elements changing upon association are identified using a non-standard structural alingment method. A detailed description on the different conformational changes over the different protein complexed families is provided.

K-28

Modelling the Electrostatics of Permeation in Voltage-gated K+ Channels

Binbin Liu[1], David R Westhead[2], Mark R Boyett
[1]bmbbl@bmb.leeds.ac.uk, School of Biomedical Sciences, University of Leeds.; [2]d.r.westhead@leeds.ac.uk, School of Biochemistry and Microbiology, University of Leeds

As a subfamily of the Shaker family, voltage-gated K+ channels function as important cellular regulators. In this study, the electrostatics of permeation in mammalian voltage-gated K+ channel was modelled using the finite difference Poisson-Boltzmann equation. The results were used to explain the effect of the charged residues, H509 and K540 on channel electrophysiology

K-29

Molecular Dynamics Refinement of HPPK-DHPS Homology Models

T de Beer[1], F Joubert[2], AI Louw
[1]tjaart@tuks.co.za, University of Pretoria; [2]fjoubert@postino.up.ac.za, University of Pretoria

Hydroxymethylpterin pyrophosphate-dihydropteroate synthase (HPPK-DHPS) is involved in sulfadoxine/dapsone resistance in malaria causing P.falciparum. We have made and refined homology models of HPPK-DHPS through Molecular Dynamics in order to try and elucidate a mechanism of resistance. Our models suggests that loop movement are involved in causing resistance to sulfadoxine and dapsone.

K-30

Identifying the Stabilizing Residues in (a/b)8 Barrel Proteins Based on Hydrophobicity, Long-range Interactions and Sequence Conservation

M. Michael Gromiha[1], Gerard Pujadas[2], Csaba Magyar, S. Selvaraj and Istvan Simon
[1]michael-gromiha@aist.go.jp, CBRC, AIST, Tokyo, Japan; [2], Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Spain

We proposed a new consensus approach for locating the stabilizing residues in (a/b8 barrel proteins based on long-range interactions, hydrophobicity and conservation of amino acid residues. We have identified 957 stabilizing residues with the abundance of eight segments per domain. These stabilizing residues show good agreement with experimental observations.

K-31

Detections of the Mirror Sites in PP2A Alpha Protein

Victoria Dominguez Del Angel[1], Alphonse Garcia[2], Bernard Caudron
[1]victoria@pasteur.fr, Institut Pasteur;
[2]agarcia@pasteur.fr, Institut Pasteur

The Protein phosphatase 2A (PP2A) is a holoenzyme with pleiotropic functions. Based on the analysis of its amino acid composition and a 3D representation, we developed an in silico approach that detects small protein fragments likely to interact with the AC core of PP2A. This strategy may be extended to detect small binding fragments in other protein families

K-32

Automatic Identification of Key Patterns Within Multiple Protein Structures

Craig Lucas[1], Andrew Bulpitt[2]
[1]craigl@comp.leeds.ac.uk, School of Computing, University of Leeds; [2]andyb@comp.leeds.ac.uk, School of Computing, University of Leeds

We present an algorithm that searches for multiple key patterns between protein structures, optimised for comparing multiple structures with common function but different fold. We demonstrate the algorithm running on three proteins, showing that finding patterns by comparing all three together is preferable to comparing two of the three separately

K-33

Surveying Protein Domain Interactions to Aid their De Novo Prediction

Stephen J Littler[1], Simon J Hubbard[2]
[1]mjfi9sl2@stud.umist.ac.uk, UMIST;
[2]simon.hubbard@umist.ac.uk, UMIST

With the ever widening sequence to structural information gap the importance of structural genomics has become paramount. Recently, it was revealed that 1009 putative domain combinations remained unsolved in structure. We aim to survey known intra-chain domain interactions to reveal pathways to aid their prediction.

K-34

The Eukaryotic Linear Motif Structural Filter

Allegra Via[1], Pal Puntervoll[2], Rune Linding, Christine Gemund, Sophie Chabanis-Davidson, Morten Mattingsdal, Scott Cameron, David M.A. Martin, Gabriele Ausiello,Anna Costantini, Simona Panni, Andreas Zanzoni, Gianni Cesareni, Francesca Diella, Giulio Superti-Furga, Lucjan Wyrwicz, Chenna Ramu, Caroline McGuigan, Rambabu Gudavalli, Ivica Letunic, Peer Bork, Leszek Rychlewski, Bernhard Kuster, Manuela Helmer-Citterich, William N. Hunter, Rein Aasland, Toby J. Gibson
[1]allegra@cbm.bio.uniroma2.it, Centre for Molecular Bioinformatics, Department of Biology, University of Rome Tor Vergata, Rome, Italy;
[2]pal.puntervoll@mbi.uib.no , Department of Molecular Biology, University of Bergen, Norway

ELM (Eukaryotic Linear Motifs - http://elm.eu.org) is a web resource for investigating candidate short functional motifs in eukaryotic proteins. The server provides core functionality including false positives filtering by cell compartment, phylogeny, globular domain clash, proteins disorder, secondary structure, and solvent accessibility

K-35

Down-Regulation of the Catalytic Activity of the EGF Receptor via Direct Contact between the Kinase and C-Terminal Domains

Meytal Landau[1], Sarel J. Fleishman [2], Ben-Tal Nir
[1]meytalc@post.tau.ac.il, Tel-Aviv University;
[2]sarel@post.tau.ac.il, Tel-Aviv University

The ErbBs are unique among RTKs, as their catalytic elements are constitutively ready for phospho-transfer. The absence of conformational regulation raises a fundamental dilemma: namely, by what mechanism is spurious activation avoided? Our studies, using various computational tools suggest a novel molecular regulation mechanism.

K-36

Side-Chain Entropy at Protein-ligand Interfaces

Christian Cole[1], Jim Warwicker[2]
[1]c.cole@umist.ac.uk, UMIST;
[2]jim.warwicker@umist.ac.uk, UMIST

Understanding Protein-ligand interactions are important for the therapeutic modulation of the associated response. Protein-ligand interfaces have different characteristics to protein-protein interfaces. However, we show that, using the Astex validation dataset, the conformational entropy of protein-ligand interfaces have similarity to protein-protein interfaces.

K-37

Structure and Sequence Characteristics of Oligomeric Proteins

Hannes Ponstingl[1], Janet Thornton[2]
[1]hpo@ebi.ac.uk, EMBl-EBI; [2]thornton@ebi.ac.uk, EMBl-EBI

The three-dimensional crystal structures of oligomeric proteins are studied with respect to physico-chemical and geometric properties and sequence variation. Feature combinations are evaluated for the prediction of protein-protein interaction sites on the subunit surface

K-38

Substrate Diversity of Single-domain Proteins

Irene Nobeli[1], Pilar Ortega[2], Ruth Spriggs, Richard George and Janet Thornton
[1]nobeli@ebi.ac.uk, EBI-EMBL; [2]ortega@ebi.ac.uk, EBI-EMBL

The diversity in the structure of cognate substrates binding to homologous single-domain proteins has been studied, both within E.coli and across all organisms. The conservation of structural scaffolds in the small molecules interacting with members of a homologous protein superfamily is assessed, and promiscuous superfamilies are highlighted.

K-39

???????????? 1 Subunit Overexpressed in Escherichia coli

Hannah Hong Xue[1], ManKit Tse[2], DaoWei Zhu, Hui Zheng and ShengXiang Lin
[1]hxue@ust.hk, HKUST; [2]bctmk@ust.hk, HKUST

The human GABAA receptor belongs to the ligand gated ion channel super family (LGICS) that mediates most of the rapid synaptic transmission in CNS. Recently, we have successfully crystallized the reported recombinant membrane proximal domain of the receptor for X ray diffraction structural studies.

K-40

Benchmarking Protein Structure Alignment Statistics Against the CATH Database

Michael Sierk[1], William Pearson[2]
[1]mls5w@virginia.edu, University of Virginia; [2]wrp@virginia.edu, University of Virginia

We present a large-scale analysis of several structure alignment methods' ability to detect homologs in the CATH domain database. The programs vary considerably in their performance, and the statistical estimates reported by the programs overstate the significance matches by orders of magnitude compared to the actual distribution of errors.

K-41

Detection of DNA-binding Structure Motifs Using Charge Clustering

Shandar Ahmad[1], Akinori Sarai[2]
[1]shandar@bse.kyutech.ac.jp, Kyushu Institute of Technology, Iizuka, Fukuoka, Japan; [2]sarai@bse.kyutech.ac.jp

We present a method of clustering charged regions in proteins and applied it to detect DNA-binding motifs. Cubic grids in a protein structure are clustered to obtain the regions of higher charge density. Final representation of these grids appears as a cluster tree diagram called Qgrid pattern from which binding motifs are detected.

K-42

Conservation of Collective Vibrational Dynamics of Proteins: Application to the Globin Family Fold

S. Maguid[1], S. Fernandez-Alberti[2], J. Echave
[1]smaguid@unq.edu.ar, Universidad Nacional de Quilmes; [2]seba@unq.edu.ar, Universidad Nacional de Quilmes

We propose a method to explore the global dynamics of structurally similar proteins but sequential and functionally different. Patterns of the common collective dynamics were identified and described as representative vectors for these motions. Minimum subspaces that expand the collective vibrational motions associated to the family fold are defined.

K-43

Analysis of Intermediate Resolution Structures

Matthew Baker[1], Wah Chiu[2]
[1]mbaker@bcm.tmc.edu, Baylor College of Medicine; [2]wah@bcm.tmc.edu, Baylor College of Medicine

To facilitate the analysis of structures from electron cryomicroscopy, we have developed the Analysis of Intermediate Resolution Structures toolkit (AIRS). AIRS provides a mechanism to localize structures, recognize secondary structure elements, identify folds and model large macromolecular assemblies. This toolkit has been successfully used on numerous datasets, including HSV-1 and RDV

K-44

Bayesian Segmental Semi-Markov Models for Protein Secondary Structure Prediction

Wei Chu[1], Zoubin Ghahramani[2], David L Wild
[1]chuwei@gatsby.ucl.ac.uk, University College London; [2]zoubiN@gatsby.ucl.ac.uk, University College London

We develop a segmental semi-Markov model for protein secondary structure prediction which incorporates multiple sequence alignment profiles. By incorporating the information from long range interactions in beta-sheets, this model is also capable of carrying out inference on contact maps.

K-45

PiBase: A Database of Protein Interfaces

Fred P. Davis[1], Andrej Sali[2]
[1]fred@salilab.org, UCSF; [2]sali@salilab.org, UCSF

PiBase is a database of protein interfaces which integrates high resolution structural data from the Protein DataBank with lower resolution protein interaction data stored in BIND and DIP. The goal is to centralize interface information for structural biologists as well as computational methods used to predict the structure of protein interactions and protein assemblies. PiBase is available at http://salilab.org/pibase.

K-46

Molecular Recognition at Protein-Protein Interfaces - CDK/cyclin Homologue Interactions

Xueping Quan[1], Dietlind Gerloff[2]
[1]s0240137@sms.ed.ac.uk, ICMB, University of Edinburgh; [2]ICMB, University of Edinburgh

The molecular interactions between the cyclins and their CDK (CDK = cyclin dependent kinase) partners are amongst the most relevant regulatory events in eukaryotes. The aim of this project is to combine bioinformatics, 3-D structural modeling and biophysical techniques to investigate, and probe, the characteristics and principles of known cyclin-CDK protein-protein interfaces in order to discover new ones.

K-47

Evolution of Specificity in the Packing of Protein Tertiary Contacts

Stephen A. Cammer[1]
[1]scammer@ucsd.edu, Univ. of California, San Diego

Preferred puzzle pieces of residue packing in proteins are identified across distinct protein folds from a database of 1452 representative structures. An all-atom based RMSD comparison reveals specificity in packing achieved through distinct evolutionary paths.

K-48

Identification of Metacaspase of T.cruzi: an Ancient Caspase-like Protein

Veronika Stoka[1], Gregor Kosec[2], Dusan Kordis, Gregor Guncar, Juan Jose Cazzulo, Vito Turk
[1]veronika.stoka@ijs.si, J. Stefan Institute; [2]gregor.kosec@ijs.si, J. Stefan Institute

T.cruzi metacaspase-3 (tcMC-3) is a caspase-like cysteine protease. It was cloned in pFastBacHT vector, expressed in baculovirus insect cells and purified by affinity chromatography. It is a protein of 45 kDa. The structure-function relationship was attempted by modeling on the known three-dimensional structures of caspases and the validity of the model was confirmed by site-directed mutagenesis.

K-49

Prediction of Ligand Receptor Specificity in Signalling Protein-Protein Complexes by Computational Docking and Global Energy Optimization of Flexible Interfaces

Juan Fernandez-Recio[1], Tom L. Blundell[2]
[1]juan@cryst.bioc.cam.ac.uk, University of Cambridge; [2]tom@cryst.bioc.cam.ac.uk, University of Cambridge

The ICM-DISCO protein-protein docking method was successfully evaluated in a recent blind experiment (CAPRI). The method provides an accurate description of the association energetics. In addition, analysis of the docking landscapes generated from these pseudo-Brownian rigid-body simulations indicates the existence of preferred protein-binding areas on protein surfaces.

K-50

Adaptive Evolution Selects between Kinetic and Thermodynamic Protein Folding

Ezhilkani Subbian[1], Yukihiro yabuta[2], Dr.Ujwal Shinde
[1]subbiane@ohsu.edu, Oregon Health and Sciences University; [2]yabyab@macmail.com, Oregon Health and Sciences University

Intracellular and extracellular serine proteases, two subfamilies within subtilases have highly conserved sequences, structures and catalytic activities but fold through significantly different pathways and mechanisms. Our results suggest surface residues influence protein folding, and their positive selection dictates folding pathways, mechanisms, and the choice between kinetic versus thermodynamically stable folds.

K-51

A Structural Annotation Pipeline for the Malaria Genome

Fourie Joubert[1], Tjaart de Beer[2], Yolandi Joubert, Corne Schriek
[1]fjoubert@postino.up.ac.za, University of Pretoria; [2]tjaart@tuks.co.za, University of Pretoria

A pipeline for the strucural annotation of the P. falciparum genome is being constructed. A series of structure-related analysis is performed on a Linux cluster, with results being submitted to a PostgreSQL database. Once completed, it will be extended to other genomes.

K-52

Structural Analysis of Residue Interaction Graphs

Arye Shemesh[1], Gil Amitai[2], Einat Sitbon, Maxim Shklar, Dvir Netanely, Ilya Venger, Shmuel Pietrokovski
[1]arye.shemesh@weizmann.ac.il, Weizmann Institute of Science; [2]gil.amitai@weizmann.ac.il, Weizmann Institute of Science

We present a graph theory approach for studying protein structures. Structures were transformed to residue interaction graphs (RIGs), where nodes are residues and edges connect interacting residues. Centrality measures of RIG nodes identify various protein functional sites. Our method can analyze single structures, is independent of sequence conservation, and does not rely on prior training.

K-53

Optimized Protein Representations From Information Theory

Julian Mintseris[1], Zhiping Weng[2]
[1]julianm@bu.edu, Boston University; [2]zhiping@bu.edu, Boston University

We address the long-standing problem of representing a protein by an alphabet of functionally similar atom/residue groups. Using information theory we can obtain an optimized protein representation from protein monomer or protein interface datasets that are in agreement with each other and with general concepts of protein energetics.

K-54

Conformational Change During Enzyme Catalysis

Alex Gutteridge[1]
[1]alexg@ebi.ac.uk, European Bioinformatics Institute

The role of conformational change in substrate binding, catalysis and product release is reviewed for 11 enzymes. Crystal structures of the apo, substrate-bound and product-bound states are analysed to measure the extent of global conformational changes and the movements of the functional regions involved in catalysis.

K-55

THESEUS: A Parallel Threading Core

Patrick May[1], Thomas Steinke[2], Michael Meyer
[1]patrick.may@zib.de, Zuse Institute Berlin; [2]steinke@zib.de, Zuse Institute Berlin

THESEUS is a parallel implementation of a protein threading algorithm for structure prediction. THESEUS is designed on the basis of a fast branch-and-bound search for the optimal threading through a library of structural templates build on SCOP domains based on a core model and a pairwise scoring function.

K-56

A Searchable Database for Comparing Protein-ligand Binding Sites for the Discovery of Structure-function Relationships

Nicola D. Gold[1], Richard M. Jackson[2]
[1]n.d.gold@leeds.ac.uk, University of Leeds; [2]jackson@bmb.leeds.ac.uk, University Of Leeds

It may be possible to infer common functional roles for proteins with similar ligand binding sites. We have therefore developed a Web accessible database of ligand binding sites combined with a similarity searching method based on geometric hashing to identify binding sites with similar atomic arrangements and properties.

K-57

Modelling the Interaction of Human Estrogen Receptor Alpha with Organic Polychlorinated Compounds

Ermanna Rovida[1], Pasqualina D'Ursi[2], Paola Fossa, Luciano Milanesi
[1]ermanna.rovida@itb.cnr.it, CNR - Institute of Biomedical Technologies; [2]pasqualina.dursi@itb.cnr.it, CNR - Institute of Biomedical Technologies

The organic polychlorinated compounds are endocrine disrupters known to bind and activate estrogen receptors. We present the molecular models of the complexes of the estrogen receptor alpha with DDT and its metabolites (DDD, DDE) and with PCB obtained with docking simulation. Details of the binding interactions are also described.

K-58

Gold STING: Studying Protein Stability and Folding by Looking at an Extensive DB of the Structure Descriptors

Goran Neshich[1], Walter Rocchia[2], Adauto L. Mancini, Michel E. B. Yamagishi, Paula R. Kuser, Renato Fileto and Roberto H. Higa
[1]neshich@cnptia.embrapa.br, EMBRAPA/CNPTIA; [2] Walter ROCCHIA <w.rocchia@sns.it>, NEST-INFM, Scuola Normale Superiore, Piazza dei Cavalieri, 7, I - 56126 Pisa -Italy

Gold STING Suite is an interactive, integrated toolset that computes physicochemical properties of macromolecular structures and provides a powerful visualization environment for analysis. STING complements modeling tools by calculating physicochemical properties from amino acid sequence and protein structure data. A key feature of STING is the capability to compare two structurally aligned proteins.

K-59

Protein Structure Active Site Signatures

Richard J. Morris[1], Rafael Najmanovich, Fabian Glaser, Roman Laskowski, Janet Thornton
[1]rjmorris@ebi.ac.uk, EBI

Methods are presented that are under development to aid protein function prediction from structure. Specifically we focus here on algorithms to locate and to describe a protein's active site with a small number of parameters. An active site similarity metric is presented that allows for rapid comparisons against a large signature database.

K-60

Using Local Sequence-structure Relationships in Proteins to Improve Fold Recognition

Oliver Sander[1], Ingolf Sommer[2], Thomas Lengauer
[1]osander@mpi-sb.mpg.de, Max-Planck-Institut für Informatik; [2]sommer@mpi-sb.mpg.de, Max-Planck-Institut für Informatik

Representatives of local structure were defined by clustering 7-residue fragments based on structural similarity. In the second step an SVM classified has been constructed to predict the local structure class from sequence information. In the future we will use these local structure predictions to improve fold recognition accuracy

K-61

"Phase Switch" Algorithm for Simulation of Intermolecular Binding

J.E. Magee[1], J. Warwicker[2], L. Lue
[1]j.magee@umist.ac.uk, UMIST; [2]jim.warwicker@umist.ac.uk, UMIST

We present a novel Monte Carlo algorithm for studying binding between molecules, which allows for direct evaluation of free energy differences. We provide proof of concept by studying square-well homopolymer chains, and discuss application to simple heteropolymer protein models and to binding of RNA with eIF4E.

K-62

Sidechain Prediction in the Recognition Helices of Homeodomains

Trevor Siggers[1], Barry Honig[2]
[1]tws11@columbia.edu, Columbia University; [2]bh6@columbia.edu, Columbia University

Gene expression requires transcription factors to correctly recognize and bind to specific DNA sequences. We aim to understand how the homeodomains recognize their cognate DNA by applying side-chain modeling to the protein-DNA interface. We rebuild the protein-DNA interface of several homeodomains, demonstrating the utility of the approach.

K-63

mRNA Openers - Computationally Designed Modulators of mRNA Secondary Structure which Manipulate Gene Expression

Joerg Hackermueller[1], Nicole-Claudia Meisner[2], Manfred Auer, Markus Jaritz
[1]joerg.hackermueller@pharma.novartis.com, In Silico Sciences, Informatics and Knowledge Management @ NIBR, Novartis Institute for Biomedical Research Vienna, Austria; [2]nicoleclaudia.meisner@pharma.novartis.com, Innovative Screening Technologies, Discovery Technologies, Novartis Institute for Biomedical Research Vienna, Austria

Recognition of RNA secondary structure motifs is of central importance for regulatory RNA protein interactions. The presented method allows to explain measured RNA protein affinities by quantitatively describing RNA secondary structure motif availability. Computationally designed secondary structure modulators boost or inhibit regulatory RNA protein interactions as predicted.

K-64

SDPMOD: A Comprehensive Comparative Modeling Server for Small Disulphide-bonded Proteins

Lesheng Kong[1], Bernett Teck Kwong Lee[2], Joo Chuan Tong, Tin Wee Tan, Shoba Ranganathan
[1]lesheng@bic.nus.edu.sg, National University of Singapore; [2]bernett@bic.nus.edu.sg, National University of Singapore

SDPMOD is a comprehensive comparative modeling server specifically designed for small disulphide-bonded proteins. SDPMOD provides fully automated comparative modeling service as well as more advanced options which allow user to choose template and adjust target-template alignment. SDPMOD is freely accessible to academic users via the web interface at http://proline.bic.nus.edu.sg/sdpmod.

K-65

Computational Approaches Towards Protein Aggregation

Jennifer A. Siepen[1], David R. Westhead[2]
[1]bmbjas@bmb.leeds.ac.uk, University of Leeds; [2]d.r.westhead@leeds.ac.uk, University of Leeds

Potential models of amyloid fibrils have been constructed using a simple modeling technique, where monomers are arranged into a continuously hydrogen bonded beta-sheet resembling that of a fibril. Simulated x-ray diffraction patterns validated our models. This method can be applied to other proteins providing insight into protein aggregation.

K-66

The Application of Multiobjective Genetic Algorithms to Protein-Ligand Docking

Sally Mardikian[1], Valerie J. Gillet[2], Richard M. Jackson, David R. Westhead
[1]bmbsam@bmb.leeds.ac.uk, University of Sheffield/University of Leeds; [2]V.Gillet@sheffield.ac.uk, University of Sheffield

A docking tool is currently being developed that employs a Multiobjective Genetic Algorithm (MOGA). MOGAs optimise at least two different objectives during a given search. Docking objectives that may be optimised are: individual non-bonded energy terms (e.g. vanderwaals interactions), and individual scoring functions within a consensus scoring function study

K-67

Structural Features of Human Galactose-1-phosphate Uridylyltransferase and of its Galactosemia-related Mutant Q188R: an Homology Modelling Study

Anna Marabotti[1], Angelo M. Facchiano[2]
[1]amarabotti@isa.cnr.it, Institute of Food Science - CNR, Avellino, Italy and CRISCEB - Second University of Naples, Naples, Italy; [2]angelo.facchiano@isa.cnr.it, Institute of Food Science - CNR, Avellino, Italy and CRISCEB - Second University of Naples, Naples, Italy

We have created by homology modelling methods the three-dimensional models of human galactose-1-phosphate uridylyltransferase, of the galactosemia-linked mutant Q188R and of heterodimers composed by different combinations of wild type and mutant subunits, and we discuss differences in their structural features to highlight the molecular basis of this genetic disease.

K-68

Homology Model and Substrate Recognition of Yeast Prk1p Kinase

Alvin Ng[1]
[1]ngyj@imcb.a-star.edu.sg, IMCB

Prk1p is a serine/threonine kinase involved in actin cytoskeleton organization in yeast Saccharomyces cerevisiae. Pan1p is required for normal organization of the actin cytoskeleton during budding. Prk1p phosphorylates Pan1p on threonine residues located on multiple repeats of LxxQxTG. A homology model was created to further understand the substrate recognition of the LxxQxTG motif.

K-69

Structural Bioinformatics of Proteasome-related HslV Protease

M. Kamran Azim[1], Sajid Noor and S. Rizwan Ali
[1]mkamranazim@yahoo.co.uk, H.E.J. Research Institute of Chemistry, University of Karachi, Karachi, Pakistan

HslVU is an ATP-dependent, two-component bacterial protease involved in the turnover of short-lived regulatory proteins. Sequence analyses and homology modeling of HslV protease component revealed that regardless of conserved active sites considerable mutations at different positions might responsible for divergent protomer interactions, HslV-HslU interfaces

K-70

Flexible Docking of Pemphigus Vulgaris Peptides to MHC Class II DR and DQ Alleles

Joo Chuan Tong[1], Shoba Ranganathan[2], Jeff Bramson, Daria Kanduc, Animesh Sinha, Shoba Ranganathan
[1]victor@bic.nus.edu.sg, National University of Singapore; [2]shoba@bic.nus.edu.sg, Research Institute for Biotechnology, Macquarie University

Pemphigus Vulgaris (PV), a potentially life-threatening form of autoimmune blistering skin disorder, is characterized by the presence of pathogenic autoantibodies directed against a 130-kDa transmembrane glycoprotein, desmoglein-3 (Dsg3). This study investigates the susceptibility and non-susceptibility of MHC class II alleles to PV using a novel MHC-peptide docking protocol.

K-71

Investigating Protein Structure Evolution by Fold Usage on Genomes

Sanne Abeln[1], Charlotte M. Deane[2]
[1]abeln@stats.ox.ac.uk, Department of Statistics, University of Oxford, UK; [2]deane@stats.ox.ac.uk , Department of Statistics, University of Oxford, UK

Here we use different measures of fold usage on completed genomes in order to explore protein structure evolution. We show that there is not a clear relationship between the number of families per fold, the number of fold duplications per genome and the number of genomes a fold occurs on.

K-72

Constant Time Nucleic Acid Backbone Construction

Philippe Thibault[1], François Major[2]
[1]thibaup@iro.umontreal.ca, Université de Montréal; [2]francois.major@umontreal.ca, Université de Montréal

We present a constant time nucleic acid backbone construction algorithm. The ribose between two phosphate atoms is constructed using a numerical approximation method. X-ray crystallography and NMR ribose conformations were built with a precision of less than 1 Å of RMSD in constant time.

K-73

Conformational Dynamics of Glutamate Receptors: Simulation Studies

Philip Biggin[1], Yalini Arinaminpathy[2], Mark Sansom
[1]phil@biop.ox.ac.uk, Oxford University; [2]pathy@biop.ox.ac.uk, Oxford University

The precise mechanism of partial agonism in ionotropic glutamate receptors remains unclear. We present results from molecular dynamics simulations of the ligand-binding core (S1S2) which suggest how partial agonists can induce single-channel currents of the same magnitude as full agonist but with a reduced frequency as recently observed experimentally.

K-74

Simulation of the Interaction of Gluten Peptides with HLA-DQ2 Molecule to Investigate the Molecular Basis of Coeliac Disease

Angelo M. Facchiano[1], Susan Costantini[2], Mauro Rossi, Giovanni Colonna
[1]angelo.facchiano@isa.cnr.it, Institute of Food Science, CNR, Avellino, Italy; [2]susan.costantini@unina2.it, CRISCEB - Research Center of Computational and Biotechnological Sciences, Second University of Naples, Naples, Italy

To investigate the molecular basis of celiac disease, we simulated the complex of specific gluten peptides with HLA-DQ2 molecule, the most frequent in celiac patients, and analysed the binding site structural features, thus giving molecular details about the HLA-DQ2 preference for negatively charged residues in specific anchor positions.

K-75

Simulation and Modelling Studies of the NMDA Receptor: A Mammalian Neurotransmitter Receptor

S. L. Kaye[1], M. S. Sansom[2], P. C. Biggin
[1]samantha@biop.ox.ac.uk, University of Oxford; [2]mark@biop.ox.ac.uk, University of Oxford

Ionotropic glutamate receptors (iGluR) are ligand-gated ion channels which mediate excitatory synaptic transmission. Crystal structures of the ligand binding domain of two types of iGluR have been solved. This poster describes the results of 30ns molecular dynamic simulations of one of these structures (NR1), in complex with agonists and an antagonist.

K-76

Molecular Dynamics Simulations of the BRCT Region of BRCA1 Reveal Long-Range Effects of Mutations on a Protein-Protein Interaction Site

Craig Gough[1], Tadashi Imanishi[2], Takashi Gojobori
[1]cgough@jbirc.aist.go.jp, Japan Biological Information Research Center; [2]imanishi@jbirc.aist.go.jp, Japan Biological Information Research Center

The fundamental motions of wild-type BRCA1-BRCT and three cancer-associated mutants were investigated using molecular dynamics and quasiharmonic analysis. In the mutants, a functionally important BACH1 helicase-binding loop spatially distant from the mutation sites exhibited motions of much larger amplitude than those observed in this loop in the wild type.

K-77

Refinement of Unbound Protein Docking Studies with Biological Knowledge

Philipp Heuser[1], Pascal Benkert[2], Davide Bau, Dietmar Schomburg
[1]philipp.heuser@uni-koeln.de, CUBIC; [2]pascal.benkert@uni-koeln.de, CUBIC

Two entirely automated docking post-filters are applied to 33 unbound docking examples. One filter uses information about domains interacting homologous to those present in the unbound proteins. The other one analyses the interface of each conformation for the conservation of Phe, Met and Trp and their polar neighbouring residues.

K-78

Predicting Local Structural Candidates from Sequence by the "Hybrid Protein Model" Approach

Cristina Benros[1], Alexandre G. de Brevern[2], Serge Hazout
[1]benros@ebgm.jussieu.fr, EBGM, INSERM E0346, university Paris 7; [2]debrevern@ebgm.jussieu.fr, EBGM, INSERM E0346

We are developing a structural profile-based method for local protein structure prediction from sequence. The aim is to propose long local structural candidates. The prediction scheme is based on a novel clustering method, named "Hybrid Protein Model". The principle of our strategy is to carry out a sequence - structure alignment.

K-79

Prediction of Factors Determining Changes in Thermostability in Protein Mutants

Parthiban Vijayarangakannan[1], Dietmar Schomburg[2]
[1]parthi@uni-koeln.de, CUBIC; [2]D.Schomburg@uni-koeln.de, CUBIC

The prediction of protein thermostability is a key research area in protein structural bioinformatics. Despite vast exploration, there are still uncovered hidden factors determining the structural stability of proteins. This approach involves the mean force potentials based on structural data such as radial distribution of heavy atoms, torsion angles, etc. to predict the changes in thermostability in protein mutants

K-80

Electrostatic Features within JAVAProtein Dossier: AminoAcid and Surface AminoAcid Electrostatic Potential Calculation and Representation

Walter Rocchia[1], Adauto L. Mancini[2], Michel E. B. Yamagishi, Paula R. Kuser, Renato Fileto, Christian Baudet, Roberto H. Higa and Goran Neshich
[1]w.rocchia@sns.it, NEST-INFM and Scuola Normale Superiore, Piazza dei Cavalieri, 7, I - 56126 Pisa -Italy; [2]adauto@cnptia.embrapa.br, Núcleo de Bioinformática Estrutural, Embrapa/Informática Agropecuária, Campinas, Brazil

JPD provides a collection of physicochemical parameters describing protein structure, stability, function and interaction. Particular attention is paid to the Electrostatic Potential, obtained solving the Poisson Boltzmann Equation. "Amino-acid Electrostatic Potential" and "Surface Amino-acid Electrostatic Potential" are calculated over all PDB content and made searchable. JPD is available at http://www.cbi.cnptia.embrapa.br.

K-81

Analysis of Forces Leading the Helix Forming. Insight into the Contribution of Side Chains

Gelena Kilosanidze[1], Alexey Kutsenko[2], Vladimir Tumanyan
[1]Gelena.Kilosanidze@mtc.ki.se, Microbiology and Tumor Biology Center, Karolinska Institute; [2]Alexey.Kutsenko@mtc.ki.se, Microbiology and Tumor Biology Center, Karolinska Institute

We analysed the energy of olygopeptide fragments in alpha-helical conformation. The result is that the "side chains - backbone" interaction makes essentially higher contribution into the van der Waals energy of a polypeptide structure than the "side chains - side chains" interaction. The former modulates a course of an energy profile.

K-82

Large Scale Surface Comparison for the Identification of Functional Similarities in Unrelated Proteins

Fabrizio Ferrè[1], Gabriele Ausiello[2], Andreas Zanzoni, Manuela Helmer-Citterich
[1]fabrizio@cbm.bio.uniroma2.it, Centre for Molecular Bioinformatics - Dept. of Biology - University of Rome Tor Vergata; [2]gabriele@cbm.bio.uniroma2.it, Centre for Molecular Bioinformatics - Dept. of Biology - University of Rome Tor Vergata

A systematic local comparison of protein surfaces is performed aiming at functional annotation based upon stringent shape and residue similarity criteria. Protein surface patches and the results of the comparison are stored in the SURFACE database. Significant similarities are found between proteins sharing low sequence or structural homology (structural genomics).

K-83

Assessment of the Chemical Diversity of Related Ligand Binding Sites

Gareth R Stockwell[1], Janet M Thornton[2]
[1]gareth@ebi.ac.uk, European Bioinformatics Institute; [2]thornton@ebi.ac.uk, European Bioinformatics Institute

A protocol for assessing the chemical diversity of a group of related ligand-binding sites is described. This method is used to identify common interaction patterns, and to determine the extent to which different strategies have evolved for the recognition of given ligand types.

K-84

ClusConf: Automated Clustering of Conformers in Protein Structure Databases

Francisco S Domingues[1], Jörg Rahnenführer[2], Thomas Lengauer
[1]doming@mpi-sb.mpg.de, Max-Planck-Institut für Informatik; [2]rahnenfj@mpi-sb.mpg.de, Max-Planck-Institut für Informatik

We describe ClusConf, an automated method to cluster ensembles of conformers. ClusConf was applied to the sets of alternative structural models available in SCOP (Structural Classification of Proteins). As a result, hierarchical clustering dendrograms and PAM clusters are provided for each SCOP species level.

K-85

Comparison of X-ray and NMR Structures

Galzitskaya [1], Garbuzynskiy[2], Melnik, Lobanov,Finkelstein
[1]ogalzit@vega.protres.ru, Insitute of Protein Reserach; [2]sergeu@alpha.protres.ru, Insitute of Protein Research

What is the difference between X-ray and NMR-resolved protein structures? A comparison of structures of 60 proteins determined by both NMR and X-ray show statistically reliable differences in the number of contacts per residue and the number of main-chain hydrogen bonds.

K-86

Stripping Down the Kinesin Molecular Motor: a Combined Informatics and Simulation Approach

Barry Grant[1], Leo Caves[2], Rob Cross
[1]grant@ysbl.york.ac.uk, York Systems Biology Laboratory, Depertment of Chemistry, University of York; [2]lsdc1@york.ac.uk, Depertment of Biology, University of York

Structural informatics and molecular modelling were used to probe sequence-structure-function relationships in the kinesin molecular motor. This study provides information on conformational changes, allosteric modulation, protein-protein interactions and evolutionary relationships. The work illustrates the powerful interplay of informatics and simulation with structural, biochemical and biophysical experiments on this important mechanoenzyme.

K-87

SWISS-MODEL Repository: Annotated Three-dimensional Protein Structure Homology Models

Juergen Kopp[1], Torsten Schwede[2]
[1]Juergen.Kopp@unibas.ch, Biozentrum Basel & SIB;
[2]Torsten.Schwede@unibas.ch, Biozentrum Basel & SIB

The aim of the SWISS-MODEL Repository is to provide access to an up-to-date collection of annotated three-dimensional protein models generated by automated homology modeling, bridging the gap between sequence and structure Databases. As of April 2004, the Swiss-Model repository contained 370,134 models.
The repository is accessible at http://swissmodel.expasy.org/repository/

K-88

Developing a Protein-ligand Docking Algorithm

Peter Oledzki[1], Richard Jackson[2], Paul Lyon
[1]bmbpo@bmb.leeds.ac.uk, University of Leeds;
[2]jackson@bmb.leeds.ac.uk, University of Leeds

A program Flexligdock is being developed to provide a flexible ligand docking tool for small molecule docking to proteins. The program is based on the Q-fit docking suite and uses a probabilistic sampling method in conjunction with the GRID molecular mechanics force field to generate and score solutions.

K-89

ARP/wARP: New Algorithms for Version 6.1

Serge X Cohen[1], Petrus H Zwart[2], Serge X. Cohen, Petrus H. Zwart, Richard J. Morris, Francisco J. Fernandez, Olga Kirillova, Mattheos Kakaris, Marouane Ben Jelloul, Gerrit Langer, Venkataraman Parthasarathy, Victor S. Lamzin & Anastassis Perrakis
[1]s.cohen@nki.nl, NKI; [2]zwart@lsx9b.nsls.bnl.gov, NSLS/BNL

We report on significant methodological developments in automated mode building that enable the new version of ARP/wARP (Version 6.1) to construct protein models with diffraction data extending to 2.6 Å

K-90

Conserved Features in 3D Structures of Homeodomain - DNA Complexes

Anna Karyagina[1], Anna Ershova[2], Andrei Alexeevski, Sergei Spirin
[1]anna@iab.ac.ru, Institute of Agricultural Biotechnology;
[2]anna@iab.ac.ru, Institute of Agricultural Biotechnology

All available 3D structures of homeodomain - DNA complexes were analysed. The geometric core and the conserved hydrophobic core of homeodomains were determined. The conserved contacts between homeodomain and DNA, including hydrogen bonds, water mediated contacts, and hydrophobic clusters on the interface, were described. Rules of homeodomain - DNA interaction are suggested.

K-91

CluD, a Program for the Determination of Hydrophobic Clusters in 3D Structures of Protein and Protein-Nucleic Acids Complexes

Andrei Alexeevski[1], Sergei Spirin[2], Daniil Alexeevski, Oleg Klychnikov, Anna Ershova, Mikhail Titov, Anna Karyagina
[1]aba@belozersky.msu.ru, Belozersky Institute, Moscow State University; [2]aba@belozersky.msu.ru, Belozersky Institute, Moscow State University

A program CluD for detecting hydrophobic clusters in a given 3D structure is elaborated (http://math.belozersky.msu.ru/~mlt/HF_page.html). In the program non-polar atomic groups are used as elementary units of hydrophobicity. This allows to include into consideration also DNA bases. The program was tested on 62 structures of homeodomains and homeodomain-DNA complexes.

K-92

Macromolecular Docking of Electron-transfer Proteins

JL Pellequer[1], J Alric[2], P Parot, S-w W Chen
[1]jlpellequer@cea.fr, CEA; [2]alric@luminy.univ-mrs.fr, Univ. Méditerranée-CEA

Protein-protein docking of electron-transfer proteins remain an important challenge to the field of computational biology. Using computational graphics with rigid-body docking as well as experimental evidence, we present a model of the reaction center tetraheme cytochrome subunit from Rubrivivax gelatinosus with two in-vivo electron transfer proteins: cytochrome c8 and HiPIP.

K-93

An Investigative Approach into Dimensionality Reduction Techniques for Protein Flexibility Modeling

Shirley Hui[1], M. Usman Shakeel[2]
[1]s2hui@uwaterloo.ca, University of Waterloo;
[2]mushakil@uwaterloo.ca, University of Waterloo

Recently, a linear dimensionality reduction technique called Principal Component Analysis was used to reduce the dimensionality of protein motion with good results. This work explores the use of a non-linear technique called Locally Linear Embedding and shows that it outperforms the results of PCA on the HIV-1 Protease model.

K-94

Structural Modeling of Packaged RNA in the Pariacoto Virus Using Molecular Mechanics Algorithms

Dipinder S. Keer[1]
[1]dipinder.keer@biology.gatech.edu, Georgia Institute of Technology

We present a model for the structure of the packaged single-stranded RNA genome in the Pariacoto virus along with the development methodology for such a model. The computational methodology uses molecular mechanics algorithms to generate a model of PaV genomic RNA within the visualized dodecahedral RNA cage.

K-95

Graph Theoretic Properties of Networks Formed by the Delaunay Tessellation of Protein Structures

Todd J. Taylor[1], Iosif Vaisman[2]
[1]ttaylora@gmu.edu, George Mason Univ., School of Computational Sciences; [2]George Mason Univ., School of Computational Sciences

The authors have subjected several sets of real and simplified model protein structures to Delaunay tessellation. The system of contacts defined by residues joined with Delaunay simplex edges can be thought of as a graph and analyzed with techniques from graph theory and the theory of complex networks. Such analysis leads us to assert: 1) protein contact networks have small world character but are technically not small world networks 2) the closed loops of 22-32 residues reported by Berezovsky can be detected with this approach 3)networks formed by native structures and grossly misfolded decoys have different graph properties.

L-1

Inference of Biological Modules in the Escherichia Coli Regulatory Network

Cristhian Avila-Sanchez[1], Sarath Chandra-Janga[2], Agustino Martinez-Antonio, Ricardo Menchaca-Mendez, Julio Collado-Vides
[1]aavila@cifn.unam.mx, Program of Computational Genomics, CIFN-UNAM; [2]sarath@cifn.unam.mx, Program of Computational Genomics, CIFN-UNAM

We present a methodology to infer nuclear biological modules of genes involved in common processes within the regulatory network of E. coli. The methodology consists of using the topological properties of the regulatory network and the information about the Gene Ontology process classification of the E. coli genes.

L-2

Ontology Network of Interactions among Operons in E.coli and B. subtilis

Sarath Chandra Janga[1], Ricardo Menchaca-Mendez[2], Cristhian Avila-Sanchez, Julio Collado-Vides
[1]sarath@cifn.unam.mx, Program of Computational Genomics,CIFN,UNAM; [2]menchaca@cifn.unam.mx, Program of Computational Genomics,CIFN,UNAM

We propose a methodology for the construction of operon interaction networks using experimentally known operons in Escherichia coli and Gene Ontology annotations. We evaluate different properties of the network so generated. Our approach provides a means of studying the organization of biological systems at higher modular level composed of operons.

L-3

IVE: Direct Link of Concrete Biochemical Networks to Dynamic Simulation

Hiroyuki Kurata[1], Kouichi Masaki[2]
[1]kurata@bio.kyutech.ac.jp, Kyushu Institute of Technology; [2]c673078k@bio.kyutech.ac.jp, Kyushu Institute of Technology

We developed the CADLIVE Simulator that directly links concrete biochemical networks to mathematical models having mechanism-related parameters based on the strategy of three layers and two stages. Once a concrete biochemical map is provided, CADLIVE builds a dynamic model, thereby enabling one to simulate and analyze it.

L-4

Quantification of the E. coli mRNA for the Quantitative Dynamic Modeling of Metabolic Pathways

Tetsuo Sato[1], Hayato Kimura[2], Ryoko Morioka, Taku Oshima, Hirotada Mori, Kotaro Minato
[1]tsato@is.naist.jp, Nara Institute of Science and Technology; [2]ha-kimur@bs.aist-nara.ac.jp, Nara Institute of Science and Technology

We introduce our recent progress and findings in measuring E. coli mRNA quantitatively by using Real-Time qPCR method. We applied our method to the dynamic modeling of E. coli metabolic pathways such as glycolysis and implemented the model to the computer simulation based on the measurement.

L-5

m Likelihood Based Model Selection Approach to Topology Classification for Biological Networks

Debra S. Goldberg[1], Frederick P. Roth[2]
[1]debg@hms.harvard.edu, Harvard Medical School; [2]froth@hms.harvard.edu, Harvard Medical School

We fit maximum likelihood models (e.g., power-law, exponential, Poisson, truncated power-law, or combination models) to the observed degree distribution of protein interaction or other biological networks. We objectively compare models for degree distribution using the Bayesian Information Criterion. This may allow quantification of false positives or other interaction classes.

L-6

oring Integrated Biological Networks: From Motifs to Superstructures

Lan V. Zhang[1], Oliver D. King[2], Sharyl L. Wong, Debra S. Goldberg, Amy H. Y. Tong, Guillaume Lesage, Brenda Andrews, Howard Bussey, Charles Boone, Frederick P. Roth
[1]lan_zhang@student.hms.harvard.edu, Harvard Medical School; [2]oliver_king@hms.harvard.edu, Harvard Medical School

To study relationships between different biological interaction types, we searched for significantly enriched network motifs in an integrated S. cerevisiae network. Many of the identified network motifs can be explained in terms of network superstructures. We produced maps of these network superstructures to provide insights and to guide further research.

L-7

Development of Rule-based Algorithms for Predicting and Synthesizing Novel Pathways

Bo Kyeng Hou[1], Dong-Yup Lee[2], Sang Yup Lee
[1]bkher71@kaist.ac.kr, Department of BioSystems and Bioinformatics Research Center, KAIST;
[2]dylee@pse.kaist.ac.kr, Department of BioSystems and Bioinformatics Research Center, KAIST

Presented herein are rule-based algorithms for predicting and synthesizing novel pathways. The plausible pathways are explored by inferring the network connections according to biotransformation rules which are based on enzyme functions and structure information of metabolites. Consequently, the whole network can be reconstructed by recovering the network connectivity.

L-8

nciliation of GC/MS Based Flux Distribution Data in Large-scale Metabolic Networks

Tae Yong Kim[1], Soon Ho Hong[2], Dong-Yup Lee, Sang Yup Lee
[1]kimty@kaist.ac.kr, Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical & Biomolecular Engineering and BioProcess Engineering Research Center, Korea Advanced Institute of Science and Technology;
[2]totenkof@webmail.kaist.ac.kr, Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical & Biomolecular Engineering and BioProcess Engineering Research Center, Korea Advanced Institute of Science and Technology

In this study, the isotope-based flux analysis and traditional metabolic flux analysis are combined to more accurately quantify the intracellular flux distribution in a large metabolic system. As a result, metabolic flux values of 308 intracellular reactions were estimated from 29 GC/MS based fluxes with relatively high accuracy.

L-9

etric Properties of Protein-Protein Interaction Networks

Natasa Przulj[1], Derek G. Corneil[2], Igor Jurisica
[1]natasha@cs.toronto.edu, University of Toronto, Department of Computer Science; [2]dgc@cs.toronto.edu, University of Toronto, Department of Computer Science

The structure of protein-protein interaction (PPI) networks is not well understood. We use local structural properties of these networks to demonstrate that the currently popular scale-free model fails to fit the data in several respects. We show that a random geometric model provides a more accurate model of PPI networks.

L-10

Data Integration to Network Discovery

Keith D. Allen[1], Marie Coffin[2], Laura O'Brien, David Pegram, Brian R Bullard
[1]kallen@paragen.com, Paradigm Genetics;
[2]mcoffin@paradigm.com, Paradigm Genetics

We have developed computational tools to integrate multiple data streams, including biochemical and gene expression profiling, and map these results to a network representation of metabolism, extracted from a human-curated mammalian reaction database. We show how this can be used in mechanism-based biomarker discovery.

L-11

Transition from Stochastic to Deterministic Behavior in Biochemical Systems

Juergen Pahle[1], Ursula Kummer[2], Borut Krajnc, Marko Marhl
[1]juergen.pahle@eml-r.villa-bosch.de, EML Research gGmbH; [2]ursula.kummer@eml-r.villa-bosch.de, EML Research gGmbH

Stochastic effects can influence the dynamic behavior of biochemical systems significantly. Therefore, it is important to take these effects into account during computational simulation whenever necessary. Investigating under which circumstances stochastic influences are substantial, we find that this is highly dependent on the attractice properties of the phase space (divergence).

L-12

Interpreting Biological Networks with Petri Nets

Oliver Shaw[1], L. Jason Steggles[2], Anil Wipat
[1]o.j.shaw@ncl.ac.uk, Computing Science, University of Newcastle; [2]l.j.steggles@ncl.ac.uk, Computing Science, University of Newcastle

Petri nets are a mathematical formalism that have been used to model computational systems but are also proving useful for analysing biological networks. Their Structural properties describe the connectivity of the network, while behavioural properties analyse the network function. We discuss these concepts and describe how they can be use to investigate bacterial networks.

L-13

A New Framework for the Systematic Integration of Models in Biology

O. Margoninski[1], P. Saffrey[2], J. Hetherington; L. Li; A. Warner; A. Finkelstein
[1]omargoni@cs.ucl.ac.uk, University College London - CoMPLEX; [2]P.Saffrey@ucl.ac.uk, University College London - CoMPLEX

We propose a new computational framework for the systematic integration of models in Biology. Our approach supports the integration of a models, implemented on a variety of tools, as well as the information management requirements posed by such a task. We have built a prototype of our framework and tested it by constructing models of metabolic pathways in the liver.

L-14

A Predictive Model of the Cognitive Proteome

J. Douglas Armstrong[1], Seth G Grant[2], Mark Cumiskey, Holger Husi, Alex Howell, Walter Blackstock, Jyoti Choudhary, Tom O'Dell, Andrew Pocklington, Peter Visscher.
[1]jda@inf.ed.ac.uk, University of Edinburgh; [2]sg3@sanger.ac.uk, Wellcome Trust Sanger institute

We have constructed a theoretical model of the protein complex associated with the NMDA receptor in mammalian neurons using information mined from the literature. The network properties of the proteins from the model correlate with known physiological and cognitive functions.

L-15

BioPAX - Biological Pathway Data Exchange Format

BioPAX Group[1], Gary Bader, Erik Brauner, Michael Cary, Robert Goldberg, Chris Hogue, Peter Karp, Joanne Luciano, Debbie Marks, Natalia Maltsev, Eric Neumann, Suzanne Paley, John Pick, Aviv Regev, Andrey Rzhetsky, Vincent Schachter, Imran Shah, Jeremy Zucker, Chris Sander
[1]biopax-info@biopax.org, BioPAX

BioPAX (http://www.biopax.org) is a data exchange format for biological pathways. The first release supports metabolic pathway data and is initially supported by BioCyc, WIT and PharmGKB. Subsequent releases of BioPAX will add support for protein-protein interactions, signal transduction pathways, genetic interactions, and other pathway data types

L-16

Differential Network Expression During Drug and Stress Response

Lawrence Cabusora[1], Christian Forst[2], Andy Fulmer, Procter & Gamble
[1]cabusora@fas.harvard.edu, Harvard University; [2]chris@lanl.gov, Los Alamos National Laboratory

We use gene expression data to identify response networks of known stress responders in M. tuberculosis (TB). This is compared to similar networks constructed from data obtained from subjecting TB to various drugs; We anticipate that this approach will be able to accelerate drug development for tuberculosis in the future.

L-17

Cross-Species Proteome Comparison as a Tool for Cancer Biomarker Discovery

Natasha Levenkova[1], Hsin-Yao Tang[2], David W. Speicher, and John J. Rux
[1]nlevenkov@wistar.upenn.edu, Wistar Institute; [2]tangh@wistar.upenn.edu, Wistar Institute

We developed a set of tools for cross-species proteome comparison that allow identification of species-specific peptides generated by in-silico proteolytic cleavage. We show how these tools can be used to identify potential cancer-associated proteins from proteomic analysis of chimerical mouse models with subcutaneous injections of human melanoma cells.

L-18

Analysis of Combinatorial Transcription Programs in Yeast

Joseph C. Mellor[1], Charles DeLisi[2]
[1]mellor@bu.edu, Boston University; [2]delisi@bu.edu, Boston University

We've developed a systematic approach (expression partitioning algorithm) for analyzing different aspects of combinatorial transcription regulation across the genome, and apply our procedure to the regulation in Saccharomyces cerevisiae to recover many instances of complex and cooperative regulation between groups of regulators and the genes they target.

L-19

Discovering Multiple Flux Distributions and Pathways for the Identification and Prioritization of Antimicrobial Drug Targets

Dong-Yup Lee[1], L. T. Fan[2], Sunwon Park, Sang Yup Lee, Shahram Shafie, Botond Bertok, Ferenc Friedler
[1]dylee@pse.kaist.ac.kr, Bioinformatics Research Center, Korea Advanced Institute of Science and Technology; [2]fan@cheme.ksu.edu, Department of Chemical Engineering, Kansas State University

A unified approach is presented for synergistically, or complementarily, identifying multiple flux distributions and redundant pathways. The comparative results from applying the approach to the models of the virtual host, i.e., human cell, and microbial pathogen demonstrate that the approach would be useful in identifying and prioritizing antimicrobial drug targets.

L-20

A Versatile Petri Net Based Architecture for Modeling and Simulation of Complex Biological Processes

Satoru Miyano[1], Masao Nagasaki[2], Atsushi Doi, Hiroshi Matsuno
[1]miyano@ims.u-tokyo.ac.jp, University of Tokyo; [2]masao@iums.u-tokyo.ac.jp, University of Tokyo

We developed a new enhanced Petri net called Hybrid Functional Petri Net with extension (HFPNe) and implemented this architecture in Genomic Object Net (http://www.genomicobject.net/). Its effectivenes is demonstrated by modeling four biological processes; alternative splicing, frameshifting, Huntington's disease model, p53 modifications which are hard to model with the former architecture

L-21

DBsolve7: New Update Version to Develop and Analyze Models of Complex Biological Systems

Nail Gizzatkulov[1], Alexander Klimov[2], Galina Lebedeva, Oleg Demin
[1]nail@kiam.ru, M.V. Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, 125047 Moscow, Russia; [2]kl_alex2000@mail.ru, Institute for Biosystems Simulation, Moscow, Russia

DBsolve 7.0 is an integrated environment software to create, analyse stoichiometric and kinetic models of complex biological networks based on systems of ordinary differential equations (ODE). DBsolve 7.0 is a successor of previously free available simulator DBsolve 5. Significant changes and improvements have been made to meet current computational Systems Biology requirements.

L-22

Toward a Functional Catalog of Arabidopsis Genome: The Construction and Analysis of Starch Biosynthesis

Tayvich Vorapreeda[1], Papapit Ingkasuwan[2], Aluck Thipayarat, Supapon Cheevadhanarak, Sakarindr Bhumiratana
[1]tayvich@yahoo.com, National Center for Genetic Engineering and Biotechnology, Thailand; [2]jobping@yahoo.com, King Mongkut's University of Technology Thonburi, Thailand

To modify starch an insight understanding of the starch biosynthesis pathway is a prerequisite. Putative starch biosynthesis pathway was reconstructed utilizing the complete genomic information of Arabidopsis thaliana via the comparative genomics approach. The two unknown proteins were discovered, which were predicted to associate with starch synthases of isoform-v.

L-23

SBML Level 3: Proposals for Advanced Model Representations

Andrew Finney[1], Michael Hucka[2]
[1]A.Finney@herts.ac.uk, University of Hertfordshire; [2]mhucka@caltech.edu, California Institute of Technology

The Systems Biology Markup Language (SBML) is an XML-based exchange format for models of biochemical networks. SBML Level 2 is in widespread use, and there are currently several proposals under development for extending SBML to create Level 3. This poster describes two of these proposals: model composition and multi-component species.

L-24

Biomodules: Analyzing a Molecular Network's Module-level Structure, Expression and Functions to Generate Hypotheses

Iliana Avila-Campillo[1], Susanne Prinz[2], Christine Aldridge, Alex Rives, Ajitha Srinivasan, Krassen Dimitrov, Andrew F. Siegel, and Timothy Galitski
[1]iavila@systemsbiology.org, Institute for Systems Biology; [2]sprinz@systemsbiology.org, Institute for Systems Biology

Biological modules are loose associations of preferred molecular interaction partners that perform a collective function. Biomodules (http://labs.systemsbiology.net/galitski/) is a program that identifies modules in molecular networks and analyzes their expression patterns and biological functions. This information is integrated, visualized, and analyzed in an interactive graphical representation

L-25

From Metabolic Pathways Network to Drug Development of Mycobacterium Tuberculosis

Asawin Meechai[1], Supalerk Satitthamajit[2], Sangdao Langsanam, Supapon Cheevadhanarak, and Sakarindr Bhumiratana
[1]asawin.mee@kmutt.ac.th, King Mongkut's University of Technology Thonburi; [2]aom_mc@hotmail.com, King Mongkut's University of Technology Thonburi

We analyzed the topology of the strain-specific-metabolic network and developed the genome-scale metabolic model of Mycobacterium tuberculosis H37Rv. From the results, we have identified a list of 160 protein targets for future development of new drugs for multi-drug-resistant (MDR) strains that have caused millions of deaths worldwide.

L-26

On the Analysis of the Intersection of the Interactome and the Metabolome

Joan Planas-Iglesias[1], Daniel Aguilar[2], Baldomero Oliva
[1]joan.planas@upf.edu, Structural Bioinformatics Laboratory (GRIB/IMIM). Departament de Ciències Experimentals i de la Salut. Universitat Pompeu Fabra; [2]aguilardv@yahoo.es, Institut de Biomedicina i Biotecnologia. Universitat Autònoma de Barcelona

We propose a semi-automated method to study the intersection of metabolome and interactome that aids inferring similar protein-protein relationships in other organisms than the few from which the actual experiments were performed. Those inferred relationships can either be validated via metabolic annotations (KEGG pathways) or via their own biological significance.

L-27

4DiCeS: Simulation Parallelisation and Model Construction

B.E. Oleson[1], M. Moeller[2], K. Prank
[1]oleson@cebitec.uni-bielefeld.de, Int. NRW Graduate School in Bioinformatics and Genome Research; [2]mmoeller@cebitec.uni-bielefeld.de, Int. NRW Graduate School in Bioinformatics and Genome Research

The Four-Dimensional Cell Simulator (4DiCeS) allows the interpretation of information transferred to, within and between cells. The cellular modelling accepts three-dimensional cell geometry changes as well as molecular stochastic effects. The application is designed to allow Java and native code plug-ins for variable cellular model-language, reaction, diffusion and geometry modules.

L-28

Network Architecture and Functional Organisation in a Post-synaptic Signalling Complex

Andrew J. Pocklington[1], Seth G.N. Grant[2], J. Douglas Armstrong, Mark Cumiskey, Alex Howell
[1]apocklin@inf.ed.ac.uk, University of Edinburgh; [2]sg3@sanger.ac.uk, Sanger Institute

The post-synaptic signalling complex associated to the NMDA receptor contains many proteins implicated in synaptic plasticity, rodent behaviour and human cognitive disorder. We investigate the structural and functional organisation of the complex, exploring the use of graph-theoretic techniques to identify signalling pathways and subnetworks related to LTP/LTD and various phenotypes.

L-29

Ontologies for Representing Interactions between Biological Entities

Holger Michael[1], Martin Haubrock[2], Torsten Crass, Edgar Wingender
[1]hom@med.uni-goettingen.de, Dept. of Bioinformatics, UKG, University of Goettingen; [2]martin.haubrock@med.uni-goettingen.de, Dept. of Bioinformatics, UKG, University of Goettingen

CYTOMER is a database originally conceived to model expression sources, (cells, tissues, and organs) at different developmental stages, mainly in human. The database is now being reorganized to serve as a starting point for the development of an ontology-based data model for formally describing the different kinds of interactions between various types of biological entities.

L-30

Modelling and Analysis of E.coli Central Metabolism Using Petri Nets

Nina Kramer[1], Ina Koch[2]
[1]kramer.nina@gmx.de, Technical University of Applied Sciences Berlin, Department of Bioinformatics; [2]ina.koch@tfh-berlin.de, Technical University of Applied Sciences Berlin, Department of Bioinformatics

The high complexity of biochemical pathways makes a clear and unambiguous representation necessary. Petri net theory provides methods for analyzing systems with concurrent processes. We have modelled and analyzed E. coli central metabolism using structural and dynamic properties of Petri nets. Based on these results statements about robustness have been made.

L-31

Clustered Genes that Participate in Differentiation of Mouse Embryonic Stem Cells

Hye Young Kim[1], Min Jung Kim[2], Hyo Chang Kim, Yong Sung Lee, Young Seek Lee, Tae Sung Park, Young Chul Kim, and Jin Hyuk Kim
[1]hykim121@hanyang.ac.kr, Hanyang University;
[2]midung@hanyang.ac.kr, Hanyang University

While differentiations of mouse embryonic stem cells, there are significant changes in gene expression. From cDNA microarray data refined through deconvolution technique, we collected genes which were significantly up- or down-regulated by induction of differentiation, hierarchically clustered those genes, and traced similarities and differences in gene sequences among the clusters.

L-32

Simulating Genetic Networks Made Easy: Network Construction with Simple Building Blocks

Steven Vercruysse[1], Martin Kuiper[2]
[1]stcru@psb.ugent.be, VIB / Ugent;
[2]makui@psb.ugent.be, VIB / Ugent

We present SIM-PLEX, a genetic network simulator with a very intuitive interface by which a user can easily specify interactions as simple "if - then" statements. The simulator is based on the mathematical model of Piecewise Linear Differential Equations. With PLDEs, genetic interactions are approximated as acting in a switch-like manner.

L-33

Modeling Cellular Processes at System Level

Xinghua Lu[1]
[1]lux@musc.edu, Medical University of South Carolina

We have developed a novel probabilistic model (VBCVQ) to identify and infer the states of the cellular processes that control the gene expression. The model is tested on a compendium of gene expression data.

L-34

GEM System: Automatic Generation of Dynamic Cell-wide Metabolic Pathway Model from the Genome

Kazuharu Arakawa[1], Yohei Yamada[2], Kosaku Shinoda, Yoichi Nakayama, Masaru Tomita
[1]gaou@sfc.keio.ac.jp, Institute for Advanced Biosciences, Keio University; [2]skipper@g-language.org, Institute for Advanced Biosciences, Keio University

The Genome-based E-cell Modeling System (GEM System) developed upon the G-language Genome Analysis Environment realizes a fully automatic conversion of genome sequence data into a quantitative in silico cell-wide metabolic pathway model, linking information from major public database such as GenBank, EMBL, SWISS-PROT, KEGG, ARM, Brenda, and WIT.

L-35

An in Silicio and in Vivo Study of Nucleotide Excision Repair

Mari[1], Geverts[2], J.H.J Hoeijmakers, W. Vermeulen, A.B. Houtsmuller.
[1]p.mari@erasmusmc.nl, Erasmus MC;
[2]b.geverts@erasmusmc.nl, Erasmus MC

A Monte Carlo simulation of the chromatin-associated process Nucleotide Excision Repair (NER) is presented. NER is a repair mechanism able to eliminate a variety of DNA-distorting lesions. Analysis of a simple sequential binding model of NER has given some insights on the kinetic properties of this biological system.

L-36

WASPIV: Web-based Assistant System for PPI (protein-protein interaction) Inference and Validation

Mi-Kyung Lee[1], Ki-Bong Kim[2]
[1]mklee@smu.ac.kr, Sangmyung University;
[2]kbkim@smu.ac.kr, Sangmyung University

Our system is a web-based system for retrieving and validating the information on protein-protein interactions (PPI). By using our system, users not only can easily search integrated database composed of several PPI Databases, such as DIP, BIND, GRID, but also can retrieve information on PPI based on functional homology and domain information. Our system is available at http://211.229.180.92:8080/piv

L-37

The Biology of Ageing e-Science Integration and Simulation

Richard J Boys[1], Colin S Gillespie[2], Thomas BL Kirkwood, Carole J Proctor, Daryl P Shanley, Darren J Wilkinson
[1]*richard.boys@ncl.ac.uk, University of Newcastle;*
[2]*c.gillespie@ncl.ac.uk, University of Newcastle*

The BASIS system is a GRID-enabled modelling tool to serve the biology of ageing research community by helping to integrate data and hypotheses from diverse biological sources. A virtual ageing cell is being developed using a set of interacting modules to represent key variables and reaction pathways within the cell.

L-38

Computer Simulation of Nuclear Protein Dynamics

Geverts[1], Mari[2], A.B. Houtsmuller
[1]*b.geverts@erasmusmc.nl, Erasmus MC;*
[2]*p.mari@erasmusmc.nl, Erasmus MC*

A computer modeling environment is presented for the simulation of (i) DNA-transacting processes in the cell nucleus, and (ii) quantitative fluorescence microscopic methods like FRAP (fluorescence recovery after photobleaching), FRET (fluorescence resonance energy transfer) an FCS (fluorescence correlation spectroscopy).

L-39

Finding Major Trends in Microarray Data Using Discriminant Analysis on Gene Ontology Classes

Ilana Saarikko[1], Matej Oresic[2], Riikka Lund, Tero Aittokallio, Riitta Lahesmaa
[1]*ilana.saarikko@btk.utu.fi, Department of Mathematics and Turku Centre for Biotechnology, Univ of Turku;*
[2]*matej.oresic@vtt.fi, VTT Biotechnology*

We present an explorative method for microarray gene expression data which performs discriminant analysis on Gene Ontology (GO) to select most descriptive classes. Method is designed to work without replication. We also present a novel visualisation of the trends of the data combining the GO classes, samples and time points.

L-40

POINT: Prediction of Human Protein-Protein Interactions

Tao-Wei Huang[1], Chi-Ying Huang[2], Cheng-Yen Kao
[1]*d90016@csie.ntu.edu.tw, Department of Computer Science and Information, National Taiwan University;*
[2]*chiying@nhri.org.tw, Division of Molecular and Genomic Medicine, National Health Research Institutes*

POINT is a database of predicting human protein-protein interactions. The POINT project predict human protein-protein interactions from the protein sequence homologs and the ortholog in each of four lower eukaryotic (Mus musculus, Saccharomyces cerevisiae, Caenorhabditis elegans, and Drosophila melanogaster) protein-protein interactions deposited in the DIP. POINT is available at http://insilico.csie.ntu.edu.tw:9999/point/

L-41

Cyclonet - a Database on Cell Cycle Regulation

Fedor Kolpakov[1], Sergey Zhatchenko[2], Igor Deineko, Vadim Valuev, Alexander Kel
[1]*fedor@biosoft.ru, Design Technological Institute of Digital Techniques; Biosoft.Ru/Development OnTheEdge.com; Design Technological Institute of Digital Techniques; Biosoft.Ru/DevelopmentOnTheEdge.com;*
[2]*zha@biosoft.ru, Design Technological Institute of Digital Techniques; Biosoft.Ru/Development OnTheEdge.com; Design Technological Institute of Digital Techniques; Biosoft.Ru/Development OnTheEdge.com*

Using BioUML technology we are developing Cyclonet database - a database on cell cycle regulation in eukaryotes. The database contains information about cell cycle specific genes, protein complexes and their interactions, microarray data, models of cell cycle regulation and results of analyses, literature references and other related resources. Availability: http://cyclonet.biouml.org.

L-42

BioUML - Open Source Extensible Java Workbench for Systems Biology

Fedor Kolpakov[1]
[1]fedor@biosoft.ru, Design Technological Institute of Digital Techniques;Biosoft.Ru/Development OnTheEdge.com; Design Technological Institute of Digital Techniques; Biosoft.Ru/DevelopmentOnTheEdge.com

BioUML workbench consists from runtime environment, meta model providing abstract layer for formal description of biological systems and a set of plug-ins. Various plug-ins provide database access (GeneNet, TRANSPATH, KEGG/pathways. CellML and SBML models), diagram editing, graph layout, query engine, biological pathways simulation, integration with MATLAB and SBW. Availability: http://www.biouml.org.

L-43

Exploring Biochemical Network Model Topologies using an Evolutionary Algorithm

Adrian Heilbut[1], Christopher W.V. Hogue
[1]heilbut@mshri.on.ca, Samuel Lunenfeld Research Institute

An evolutionary algorithm to consruct biochemical network models is applied to search for a variety of interesting networks, including oscillators and systems that exhibit different behaviour under stochastic vs. deterministic dynamics. The approach may be useful for generating hypotheses about natural systems and for the design of synthetic biochemical networks.

L-44

Cosine Transformation to Refine Transcriptional Regulation Relationships using Time-series Microarray Data

jihun kim[1], juhan kim[2], jihoon kim, mingoo kim
[1]djdoc@snu.ac.kr, SNUBI: SNUBiomedical Informatics; [2]vcdent1@snu.ac.kr, SNUBI: SNUBiomedical Informatics

Time-series Microarray data has given us a lead to extract transcriptional regulation information. we propose an approach based on cosinor analysis to fine segmentation of potential regulation (whether a gene is regulator or regulatee), and shows a more detailed way of separating between potential activation or inhibition regulations.

L-45

The Specificity of Topology of Protein-protein Interaction, Interactions in Yeast and its Biological Characterization

Yasuhiro Suzuki[1], Takeshi Hase[2], Sohich Ogisima, So Nakagawa, Hiroshi Tanaka
[1]suzuki@bioinfo.tmd.ac.jp, Tokyo Medical and Dental University; [2]hase@bioinfo.tmd.ac.jp, Tokyo Medical and Dental University

The specificity of topology of the protein-protein interactions of yeast was investigated and its biological significance was examined. Previous studies have been mainly focused on the high layer proteins (highly-connected proteins), but we found that the middle layer classified proteins were topologically and biologically significant.

L-46

Classical and Bayesian Approaches to Reconstructing Genetic Regulatory Networks

Matthew J. Beal[1], Claudia Rangel[2], Francesco Falciani, Zoubin Ghahramani, David L. Wild
[1]beal@cs.tornoto.edu, University of Toronto; [2]rangelc@usc.edu, University of Southern California

We have used state space models to reverse engineer transcriptional networks from highly replicated gene expression profiling time series data obtained from a well-established model of T cell activation. Our models represent the dynamics of T cell activation and provide a methodology for the development of experimentally testable hypotheses.

L-47

Metabolomic Networks in Plants

Ralf Steuer[1], Wolfram Weckwerth[2], Katja Morgenthal
[1]steuer@agnld.uni-potsdam.de, University Potsdam; [2]weckwerth@mpimp-golm.mpg.de, Max Planck Institute of Molecular Plant Physiology

Gas chromatography coupled to mass spectrometry (GC/MS) analysis of plant metabolite mixtures achieved an unmatched sample throughput enabling the construction of dynamic metabolomic networks at a systems level We present a computational approach how these networks can be understood with respect to known reaction pathway structures an provide a systematic comparision of metabolomic networks of potato leaf

L-48

Structural Analysis of Genomic Scale Metabolic Networks

Bhushan Bonde[1], David Fell[2], Mark Poolman
[1]bkbonde@brookes.ac.uk, Oxford Brookes University; [2]dfell@brookes.ac.uk, Oxford Brookes University

A genomic scale model of E.coli metabolic network is being examined to study the effect of genetic regulation of enzyme subsets. A detail analysis of model includes identification of enzyme subsets and dead reactions. We observed that enzymes belonging to the same enzyme subset may share similar patterns of genetic regulation.

L-49

SBML Software and Services

Michael Hucka[1], Andrew Finney[2], Benjamin J. Bornstein, Bruce E. Shapiro, Sarah M. Keating, Benjamin L. Kovitz, Akira Funahashi, Maria J. Schilstra, and Joanne Matthews
[1]mhucka@caltech.edu, Control and Dynamical Systems, California Institute of Technology, Pasadena, CA, USA; [2]afinney@cds.caltech.edu, STRS, University of Hertfordshire, Hatfield, UK

The Systems Biology Markup Language (SBML) is a widely-used XML-based exchange format for computational models of biochemical networks. In this poster, we summarize the software infrastructure currently being developed and made available by our group for using and working with SBML. More information about SBML is available at http://www.sbml.org.

L-50

PATIKA: An Informatics Infrastructure for Cellular Networks

Emek Demir[1], Asli Ayaz[2], Ozgun Babur, Ugur Dogrusoz, Zeynep Erson, Erhan Giral, Gurcan Gulesir, Gurkan Nisanci
[1]emek@cs.bilkent.edu.tr, Bilkent University; [2]ayaz@cs.bilkent.edu.tr, Bilkent University

The PATIKA Project aims for an informatics infrastructure to cope with the inherent complexity of cellular pathway data and provides a tool composed of a central database and a visual editor, built around an extensive ontology and an integration framework. It also features microarray data analysis and pathway inference components.

L-51

Analyzing Fragility and Identifying Targets in Biochemical Reaction Networks by Means of Minimal Cut Sets

Steffen Klamt[1]
[1]klamt@mpi-magdeburg.mpg.de, Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany

We present the new approach of minimal cut sets (MCSs) useful for studying dysfunctions in biochemical reaction networks. Each MCS is a minimal (irreducible) set of reactions whose inactivation provokes network failure and impedes, for example, the continuous synthesis of certain products. Potential applications including target identification are illustrated.

L-52

Mathematical Modelling of Calcium Homeostasis in Yeast Cells

Jiangjun Cui[1], Jaap A. Kaandorp[2]
[1]jcui@science.uva.nl, Section Computational Science, University of Amsterdam; [2]jaapk@science.uva.nl, Section Computational Science, University of Amsterdam

Based on currently available experimental observations, we constructed a mathematical model to describe calcium homeostasis in yeast cells under normal conditions. Simulation results show that tightly controlled low cytosolic calcium ion level can be a natural result under the general mechanism of gene expression feedback control.

L-53

Modeling of Starch Synthesis Metabolism in Tuber Plant: a Step Towards an in Silico Plant Cell

Treenut Saithong[1], Asawin Meechai[2], Supapon Cheevadhanarak and Sakarindr Bhumiratana
[1]treenut.sai@biotec.or.th, National Center for Genetic Engineering and Biotechnology, KMUTT; [2]asawin.mee@kmutt.ac.th, KMUTT

The analysis of starch biosynthesis pathway in amyloplast of potato using metabolic control analysis with Gepasi is the first-step study of the ultimate goal; plant system biology. The results indicated that phosphoglucomutase which functionally converts glucose-6-phosphate to glucose-1-phosphate is the product controlling enzyme of the system pathway.

L-54

Gene Cluster Regulatory Network to Drug Target Identification by Using Transcriptional Profile of Plasmodium Falciparum

Saowalak Kalapanulak[1], Asawin Meechai[2], Supapon Cheevadhanarak, Sakarindr Bhumiratana
[1]saowalak.kal@biotec.or.th, Biochemical Engineering and Pilot Plant Research and Development Unit, King Mongkut's University of Technology Thonburi, Bangkuntien Campas, National Center for Genetic Engineering and Biotechnology, Bangkok 10150, Thailand; [2]Department of Chemical Engineering, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand

A regulatory network of Plasmodium falciparum that interconnects gene clusters to transcription factors to gene clusters was inferred from the systemic analysis of the expression data and upstream sequences of its global genes. This network reveals us a set of promising targets for further development of new drugs against malaria.

L-55

Simulations of the Modulation of Protein Clustering by Lipid Rafts

Dan V. Nicolau, Jr.[1], Francis Clark[2], John Hancock, Kevin Burrage
[1]dan@maths.uq.edu.au, University of Queensland; [2]fc@maths.uq.edu.au, University of Queensland

A model for the selective aggregation of proteins in lipid rafts is presented, based on the random diffusion of proteins on a surface covered with patches representing rafts. This model can be used to infer physico-chemical characteristics of protein motion in the membrane or to validate fundamental lipid raft models.

L-56

PyBioS - an Object-oriented Tool for Modelling and Simulation of Cellular Processes

Christoph Wierling[1], Elisabeth Maschke-Dutz[2], Edda Klipp, Marie-Laure Yaspo, Hans Lehrach, and Ralf Herwig
[1]wierling@molgen.mpg.de, Max-Planck Institute for Molecular Genetics; [2]maschke@molgen.mpg.de, Max-Planck Institute for Molecular Genetics

PyBioS is an object-oriented, web-based environment that is designed for applications in Systems Biology. It supports the formulation of hypotheses, e.g. for in silico knock-out experiments, quantitative and qualitative effects on the systemic level using high levels of automation through database interfaces (e.g. KEGG) and experimental data platforms (e.g. Microarrays).

L-57

Completion of the Yeast Transcriptional Regulation Network by Protein-protein Interactions

Smidtas Serge[1], Schachter Vincent[2], Képès F., Bourgine P.
[1]sergi@sergi5.com, CNRG Genoscope, CNRS UMR 8030; [2]vs@genoscope.cns.fr, CNRG Genoscope, CNRS UMR 8030

Global and local structure of the yeast heterogeneous protein-protein and protein-dna network. Topological properties of the heterogeneous network: completion, distinct layers, network motif searches. From topological to dynamical network motif properties: translation, characterization of motif functional role. Exemple of a regulatory control loop with function in adapting response to environmental conditions.

L-58

Localized Estimation Algorithm to Gene Regulatory Network from Microarray Data Using Linear System of Differential Equations

Dougu Nam[1], Sangsoo Kim[2]
[1]dunam@kribb.re.kr, NGIC, KRIBB; [2]sskimb@kribb.re.kr, NGIC, KRIBB

From time series microarray data, we try to estimate the genetic network using high dimenetional linear system of stochastic difference equations. By localizing the original maximum likeihood estimation algorithm, we could estimate the underlying system more accurately.

L-59

"The Inferelator": A Modular Computational System for Inferring Gene Regulatory Networks from Diverse Systems-biology Data Sets, with Application to the Archaeon Halobacterium NRC-1

Richard[1], Vestienn[2], Lee Hood, David Reiss, Nitin Baliga
[1]rbonneau@systemsbiology.org, Institute for Systems Biology; [2]thorsson@systemsbiology.org, Thorsson

We describe a modular system for the inference of global gene regulatory networks. We apply our system to infer much of the global regulatory network for the Halophilic archaeon Halobacterium NRC-1, using complementary data types including but not limited to: microarray expression, ChIP-chip analysis and 3-dimensional protein structure prediction.

L-60

Computational Genome-scale Assessment of the Impact of Codon Choice on Translational Efficiency in Escherichia coli

Timothy E. Allen[1], Bernhard O. Palsson[2]
[1]teallen@ucsd.edu, University of California, San Diego; [2]palsson@ucsd.edu, University of California, San Diego

The theoretical range in the translational efficiency for every protein in Escherichia coli was computed under two assumptions: localized tRNA molecules and non-diffusion limited tRNAs. The results suggest that systems analyses of macromolecular interactions will have to account for the actual spatial organization of the translational apparatus within the cell.

L-61

Proposal of Metabolic Network Formal Description to Enable Comparison between Species

Nicolas Parisey[1], Mazière Pierre[2], Beurton-Aimar Marie, Molina Franck
[1]Nicolas.Parisey@etud.u-bordeaux2.fr, Laboratoire de Physiologie Mitochondriale, Université Victor Segalen, Bordeaux; [2]pierre.maziere@ibph.pharma.univ-montp1.fr, Center for Pharmacology and Health Biotechnology, UMR 5160, Faculté de Pharmacie, Montpellier

Formal description of biological processes is often a bottleneck for computational approaches. Therefore, we have used a formal description sheme, BioY, to describe and to store descriptions of several species TCA Cycle. This work is part of a project to describe and to simulate in silico mitochondrial metabolism.

L-62

Kinetic Model of E.coli Krebs Cycle

Zobova E.A.[1], Lebedeva G.V.[2], Demin O.V.
[1]zobova@genebee.msu.su, A.N.Belozersky Institute of Physico-Chemical Biology, Moscow State University; [2]lebedgv@yandex.ru, A.N.Belozersky Institute of Physico-Chemical Biology, Moscow State University

Kinetic model of E.coli Krebs cycle functioning during aerobic growth on acetate or fatty acids has been developed. Model includes the enzymes of full Krebs cycle, glyoxylate shunt and isocitrate dehydrogenase kinase/phosphatase description. The model was applied for the investigation of Krebs cycle regulatory properties.

L-63

The Kinetic Model of Lysine of E.coli Biosynthetic Pathway

Anastasia Lavrova[1], Sergei Mironov[2], Oleg Demin
[1]ijgty@mail.ru, Moscow State University, Biophysics Department; [2], Physico-Technical Instituit, Dolgoprudnyi, Russia

We have developed model of lysine biosynthetic pathway, which consists of (mini) models of each enzymes. Using this model we have found possible ways of inhibition of diaminopimelate biosynthesis, which is one of intermediates of lysine biosynthesis pathway and essential component of cell wall of bacteria.

L-64

Kinetic Model of Mitochondrial Adenine Nucleotide Translocase

Eugeniy Metelkin[1], Oleg Demin[2]
[1]emetelkin@yandex.ru, A.N. Belozersky Institute of Physico-Chemical Biology; [2]demin@genebee.msu.su, A.N. Belozersky Institute of Physico-Chemical Biology

Adenine nucleotide translocase (ANT) is one of the main components of oxidative phosphorylation in mitochondria. We developed kinetic models of ANT and found that only two binding sites mechanism allowed us to describe biphasic behavior. Using available experimental data we estimated kinetic parameters.

L-65

Mathematical Modelling of Biosynthesis of Valine and Isoleucine in E.coli

Tatiana Plusnina[1], Demin Oleg[2]
[1]plusn@yandex.ru, Moscow State University; [2]demin@genebee.msu.su, Belozersky Institute of Physico-Chemical Biology

The kinetic model of biosynthesis of the branched-chain amino acids valine and isoleucine in E.coli was developed. It was described by the system of the ordinary differential equations and validated according to experimental data. The model was used for choosing of the optimal strategy for increasing of amino acids production

L-66

CUBIC Metabolic Pathway Hunter Tool - CMPHT

Syed Asad Rahman[1], Dietmar Schomburg[2], Rainer Schrader
[1]*asad.rahman@uni-koeln.de, CUBIC;*
[2]*D.Schomburg@uni-koeln.de, CUBIC*

Comparative study of metabolic pathways can reveal hidden facts. The goal of this work is to develop robust methods to analyze the underlying metabolic network in different organism. This will open the door for better drug as well as help us to cater the industrial needs in an efficient manner.

L-67

Quantitative Modelling Strategy for the Discovery of Cell-Type and Disease Specific Metabolic Pathways

Kumar Selvarajoo[1], Patrick Tan[2]
[1]*kumars@bii.a-star.edu.sg, Bioinformatics Institute;*
[2]*cmrtan@nccs.com.sg, National Cancer Centre*

We implemented a novel computational approach that uses metabolite concentrations and generic metabolic networks to simulate and compare the metabolic phenotypes of different cell types and species. Our study illustrates how computer simulations can be used for biological pathway discovery and to elucidate complex metabolic phenotypes associated with human disease

L-68

Kinetic Simulation of Signal Transduction in Hippocampal Long-term Potentiation with Dynamic Modeling of PP2A

Shinichi Kikuchi[1], Kenji Fujimoto[2], Noriyuki Kitagawa, Taro Fuchikawa, Michiko Abe, Kotaro Oka, Kohtaro Takei, Masaru Tomita
[1]*kikuchi@sfc.keio.ac.jp, Institute for Advanced Biosciences, Keio University;* [2]*, Institute for Advanced Biosciences, Keio University*

We modeled and analyzed signal transduction system of hippocampal LTP. Especially, we showed that another mechanism could introduce LTP bistable behavior by adding dynamic reactions of PP2A. Results of sensitivity analysis showed evidence of PP1 activation module rather than protein kinase C reported in the conventional

L-69

Stochastic Spatial Simulation of Gene Regulation Networks Using Cellular Automata

Joerg R. Weimar[1]
[1]*J.Weimar@tu-bs.de, AG Bioinformatik, Institut für Software, Technical University Braunschweig Germany*

We model gene regulation phenomena using cellular automata where each site contains at most one macromolecule, which interact locally. Stochastic fluctuations and spatial localization effects are included, and localized phenomena can be investigated. The rules of the probabilistic block-cellular automata are derived from the reactions to be modeled.

L-70

Mathematical Modeling of the Complex Cooperative Behavior in Populations of Plant Pathogen Agrobacterium Tumefaciens

Andrew Goryachev[1], Low Hong Sang[2], V. Sarathy, C. Pang, L. H. Zhang
[1]*andrewg@bii-sg.org, Bioinformatics Institute;*
[2]*lowhs@bii.a-star.edu.sg, Bioinformatics Institute*

Biologically faithful mathematical model of the entire quorum sensing network is presented for the plant pathogen Agrobacterium tumefaciens. We perform functional analysis of the network by considering its structural perturbations corresponding to several knock-out mutants. We demonstrate direct consequences of the topology changes on efficiency and performance of the network.

L-71

Efficient Simulation Methods for Stochastic Biological Systems

Tianhai Tian[1], Kevin Burrage[2]
[1]*tian@maths.uq.edu.au, University of Queensland;*
[2]*kb@maths.uq.edu.au, University of Queensland*

We present an efficient simulation method for stochastic chemical kinetics: the binomial tau-leap methods. We also discuss stepsize conditions for restricting reaction numbers in a chosen time interval. Numerical results indicate that these methods can be applied to a wide range of chemical reaction systems.

L-72

4DiCeS: Advanced Hybrid Model - an Efficient Algorithm on Molecular Reaction

M Moeller[1], H Wagner[2], B Oleson, K Prank
[1]mark.moeller@gmx.de, University of Bielefeld; [2]hwagner@Mathematik.Uni-Bielefeld.de, University of Bielefeld

We present an algorithm for chemical reactions basing on two levels of stochastic modelling: An exact modeling of reaction channels with small numbers of reactants, and a more effective modeling for the others, basing on the transition of certain portions of substrates

L-73

Inference of Local Gene Networks from Gene Expression Time Course

Mukesh Bansal[1], Giusy Della Gatta[2], Diego di Bernardo
[1]bansal@tigem.it, TIGEM; [2]dellagatta@tigem.it, TIGEM

A new algorithm that can infer the network of gene-gene interaction to which the gene of interest belongs, without the prior knowledge of the gene involved in the network, using the perturbation of only one gene and by measuring the gene expression profiles of all genes at multiple time point.

L-74

Simulation of the Notch Genetic Network during Early Development of the Vertebrate Inner Ear

Adrian L. Garcia-Lomana[1], Gina Abelló[2], Berta Alsina, Jordi Villà-Freixa
[1]alopez@imim.es, UPF; [2]ginaabello@hotmail.com, UPF

We have simulated the Notch genetic network responsible for early development of the vertebrate inner ear by a limited set of differential equations. We have obtained a dynamical model that reproduces cellular lateral inhibition and boundary formation. Parameters space is quantitatively explored for the design of new bench experiments

L-75

Robustness of Key Components of Homeostatic Gene Networks to a Wide Range of Mutations

Alexander V. Ratushny[1], Vitaly A. Likhoshvai[2], Elena V. Ignatieva, Nikolay A. Kolchanov
[1]ratushny@bionet.nsc.ru, Institute of Cytology and Genetics SB RAS; [2]likho@bionet.nsc.ru, Institute of Cytology and Genetics SB RAS

Here we present a theoretical research of functioning of a gene network controlling cholesterol homeostasis in animal cells. It was investigated effects of typical known from the literature and hypothetical mutations in the gene network and analyzed a mutational portrait of the system by using the computer dynamic model developed earlier.

L-76

Analysis of Gene Expression Data from a Bone-related Cell Line

William Newell[1], Teresa Gracia[2], Beatrice Vayssiere, Georges Rawadi, Teresa Garcia
[1]william.newell@proskelia.com, Proskelia SAS; [2]teresa.garcia@proskelia.com, Proskelia SAS

Genome-wide gene expression has been measured in a mouse pluripotent cell line treated with Wnt3a, and analyzed using several numerical and methods. The results are visualized using different graphical browsing techniques, enabling rapid evaluation of key biological processes. Methods for visualizing regulated pathways are also demonstrated.

L-77

Extension of a Multi-level Discrete Formalism to Simulate Regulatory Networks within a Large Number of Cells

Aitor González González[1], Claudine Chaouiya[2], Denis Thieffry
[1]gonzalez@ibdm.univ-mrs.fr, Laboratoire de Génétique et Physiologie du Développement; [2]chaouiya@ibdm.univ-mrs.fr, Laboratoire de Génétique et Physiologie du Développement

To simulate intra- and inter-cellular regulatory networks within small cell number, we have been using a multilevel discrete formalisation method. Here we extend this approach to ease the simulation of larger number and of different kinds of regulation. Finally we illustrate it with an example from the literature.

L-78

Interaction Networks in Time and Space - a Model of the Mitotic Cell Cycle

Ulrik de Lichtenberg[1], Lars Juhl Jensen[2], Søren Brunak, Peer Bork
[1]ulrik@cbs.dtu.dk, Center for Biological Sequence Analysis; [2]jensen@embl.de, EMBL

Large-scale protein interaction data sets provide important, although static, information on dynamic processes such as the mitotic cell cycle, which has been extensively studied at the transcriptional level. Here we provide a dynamic, system-level view of the cell cycle through integration of data from interaction screens, Microarrays, and subcellular localization.

L-79

A Method for Reconstruction of Small-scale Genetic Networks

Tra T. Vu[1], Jiri Vohradsky[2]
[1]vuthitra@biomed.cas.cz, Inst. of Microbiology, CAS; [2]vohr@biomed.cas.cz, Inst. of Microbiology, CAS

An algorithm for reconstruction of weight matrix of nonlinear differential model of genetic network from experimental time series using Levenberg-Marquardt Backpropagation algorithm and Numerical Computation approach (called LMBP_NC algorithm) is presented. The approach has been applied to reconstruction of alternative pathways of a lambda phage genetic network controlling lytic and lysogenic state of the cell.

L-80

On the Generalisation of the Metabolic Thermodynamic Theory of the Cell Cycle

A Kummer[1], R Ocone[2]
[1]ari.kummer@hw.ac.uk, Heriot-Watt University; [2]r.ocone@hw.ac.uk, Heriot-Watt University

In this work we present the generalisation of the thermodynamic-like theory of the cell cycle developed by ourselves to include the situation of dissipative systems. We also show that the theory still holds when the system is not at equilibrium

L-81

Computational Methods for Metabolite Screening

Priti Talwar[1], Thomas Lengauer[2], Joerg Rahnenführer, Vidya Velagapudi, Christoph Wittmann , Elmar Heinzle
[1]ptalwar@mpi-sb.mpg.de, Max-Planck-Institut für Informatik; [2]lengauer@mpi-sb.mpg.de, Max-Planck-Institut für Informatik

We are developing computational methods for protein function prediction by identification of metabolite signals reflective of changes in stable isotopic carbon assimilation, flux coupling and metabolic flux distribution. Here we present preliminary results from analysis of 59 gene knockout mutants involved in regulation of central carbon metabolism in Saccharomyces cerevisiae.

L-82

Incremental Refinement of Metabolic Network Models

Esa Pitkänen[1], Ari Rantanen[2], Juho Rousu, Esko Ukkonen
[1]esa.pitkanen@cs.helsinki.fi, Department of Computer Science, University of Helsinki; [2]ari.rantanen@cs.helsinki.fi, Department of Computer Science, University of Helsinki

We present a novel computational method for generation of metabolic networks for metabolic reconstruction or engineering purposes. The method explores the neighborhood of an user-given network by inserting or deleting a few enzymatic reactions. In particular, any evaluation method can be used to assess the quality of generated networks.

L-83

Design Considerations for Gene Expression Probabilistic Graphical Models with Hidden Variables

William Amadio[1], Jonathan Yavelow[2], Christopher Holcombe
[1]amadio@rider.edu, Rider University; [2]yavelow@rider.edu, Rider University

This poster examines the implications for PGM design of some long-standing statistical results. We present guidelines for the PGM user that, to date, have not been articulated in the gene expression PGM literature, but which are apparent in some recently published models.

L-84

The Network Inference Testbed: An Environment for the Testing and Training of Algorithms for the Inference of Regulatory Networks

Ronald Taylor[1]
[1]*ronald.taylor@pnl.gov, Pacific Northwest National Laboratory*

The Network Inference Testbed (NIT) is being created at Pacific Northwest National Laboratory as an interactive environment for the evaluation of algorithms used in the reconstruction of the structure of regulatory networks. The NIT compares and trains genetic network inference methods on artificial networks and simulated gene expression perturbation data

L-85

Cell++ - Simulating Biochemical Pathways

Matthew Yip[1], John Parkinson[2]
[1]*mlkyip@hotmail.com, Hospital for Sick Children;*
[2]*jparkin@sickkids.ca, Hospital for Sick Children*

Cell++ is a novel spatial/temporal simulation environment aimed at modelling biochemical pathways. Unlike traditional pathway models based on differential equations, Cell++ is capable of exploring the influence of cell architecture and organisation on the behaviour of biochemical pathways.

L-86

A Fast Reconstruction Algorithm for Gene Networks

Ilaria Mogno[1], Lorenzo Farina[2], Timothy Gardner
[1]*mogno@bu.edu, Universita' di Roma La Sapienza;*
[2]*mogno@bu.edu, Universita' di Roma La Sapienza*

Our Fast Algorithm infers causal connections between targeted genes. We assume that the dynamic can be approximated by a linear systems of differential equations, and we deal with the problem of sparse network reconstruction. The algorithm is tested and validated in a series of in numero experiments and applied to real data.

L-87

Finding Modules Using Network Motifs as Building Blocks

Ricardo Menchaca-Mendez[1], Sarath Chandra Janga[2], Cristhian Avila-Sanchez and Julio Collado-Vides
[1]*menchaca@cifn.unam.mx, Program of Computational Genomics, CIFN-UNAM;*
[2]*sarath@itzamna.cifn.unam.mx, Program of Computational Genomics, CIFN-UNAM*

We present a methodology for detection of functional modules from the known E.coli regulatory network by using the identified motifs as building blocks. Gene ontology (process) information was used to evaluate the biological significance of each module. This approach provides a means of improving the gene ontology annotations currently available.

L-88

Statistical Approaches for Inferring Non-exclusive Gene Groupings from Gene Expression Data

Sudhakar[1], Rajagopalan[2]
[1]*g0203685@nus.edu.sg, Jonnalagadda;*
[2]*chergs@nus.edu.sg, Srinivasan*

Gene products are known to participate in multiple pathways within a network. Several methods have been proposed to infer nonexclusive gene groups from microarray data. In this paper, we propose a new framework to compare these methods. Using this, we critically evaluate the objectives and underlying assumptions of the methods.

L-89

ViCe: a VIrtual CEll Formalized in the pi-calculus

Roberto Marangoni[1], Michele Curti[2], Davide Chiarugi, Pierpaolo Degano
[1]*marangon@di.unipi.it, Department of Informatics, Pisa;*
[2]*curtim@di.unipi.it, Department of Informatics, Pisa*

Relying on the formalism provided by an enhanced pi-calculus, we specified VICE, our VIrtual CEll. This hypothetical prokaryote is based on a very simplified genome and possess a complete set of metabolic pathways. The results of our experimentation in silico shows that VICE reproduces some features typical of real organisms.

L-90

Production Distances in Metabolic Networks

Esa Pitkänen[1], Ari Rantanen[2]
[1]esa.pitkanen@cs.helsinki.fi, Department of Computer Science, University of Helsinki;
[2]ari.rantanen@cs.helsinki.fi, Department of Computer Science, University of Helsinki

We define a new distance measure for metabolite pairs in a metabolic network. Results for pairwise distances in a model of S. cerevisiae are presented, and compared against previous works. Our results indicate that metabolic networks are not as robust as suggested in previous small world studies

L-91

Dynamical Complexity Reduction in Biochemical Reaction Networks

Juergen Zobeley[1], Dirk Lebiedz[2], Julia Kammerer (IWR, University of Heidelberg), Ursula Kummer (EML Research gGmbH)
[1]juergen.zobeley@eml-r.villa-bosch.de, EML Research gGmbH; [2]lebiedz@iwr.uni-heidelberg.de, IWR, University of Heidelberg

A dynamical complexity reduction method for the simulation of large and complex biochemical reaction networks is presented. The capabilities of the method which allows for the automated identification of the key features of dynamical systems are illustrated in a simulation study of a Peroxidase-Oxidase reaction network model.

L-92

Distinguishing Transcriptional and Post-transcriptional Expression Regulation on a Genomic Scale

Andreas Beyer[1], Jens Hollunder[2], Thomas Wilhelm
[1]beyer@imb-jena.de, IMB-Jena; [2]hollund@imb-jena.de, IMB-Jena

Correlation of protein and mRNA abundance in different cell compartments and functional modules is analyzed for yeast on a genome-wide scale. This analysis reveals the significance of transcriptional and post-transcriptional expression regulation. A correlation between mRNA levels and the mutation-rate is found, suggesting that higher transcription protects against mutation.

L-93

SCIpath - An Integrated Environment for Systems Biology Analysis and Visualisation

Manish Patel[1], Sylvia Nagl[2], Amos Folarin, Ken Edwards
[1]manish.patel@ucl.ac.uk, University College London;
[2]snagl@medsch.ucl.ac.uk, Univeristy College London

SCIpath is an extendable suite of visualisation and analysis programs that aims to promote understanding of Systems Biology. Currently, the suite draws upon Data Mining of high throughput technologies (e.g. Microarrays) and builds useful visualisations built from these mining techniques.

L-94

Inferring Pathways from Gene Lists Using a Literature-derived Network of Biological Relationships

Dilip Rajagopalan[1], Pankaj Agarwal[2]
[1]dilip.2.rajagopalan@gsk.com, GlaxoSmithKline;
[2]pankaj.agarwal@gsk.com, GlaxoSmithKline

We present an approach to interpretation of gene lists derived from omics experiments. We search a network of gene relationships for subnetworks consisting largely of genes from the gene list. The genes in the subnetwork and the relationships between them help uncover pathway information contained in the gene list.