

ISMB 2005 Posters

1 RNA and Protein Structural Biology

Poster A-2 (*There will also be an oral presentation of this poster.*)

A Conserved Sparse Dicondon Framework Which Correlates Sequence and Structure: Implications for Gene Finding

David Halitsky (Cumulative Inquiry, Inc); Arthur Lesk (Dept of Haemaology, CIMR); Jacques Fresco (Princeton University)

Abstract: Analysis of di-codon pairs in mRNA sequences can identify structurally similar features in the encoded proteins via a sparse signal characterized by number and order of certain dicodons occurring within codon subsequences of specific lengths. The signal reliably detects structurally similar features with virtually no underlying sequence similarity.

Poster A-3

De Novo Assembly of Transmembrane Helices of Polytopic Membrane Proteins Using Sequence Conservation Patterns

Yungki Park (Center for Bioinformatics, Saarland University); Volkhard Helms (Center for Bioinformatics, Saarland University)

Abstract: A novel two-step method for modeling structures of transmembrane helix bundle proteins was developed: generation of libraries of folds and specification of the best fold based on sequence conservation patterns. For a broad spectrum of test proteins, it consistently generated model structures within CA RMSDs of 3–5 Å.

Poster A-4

Protein-Protein Docking Methods Used to Study Complex Protein Interactions

Dana Haley-Vicente (Accelrys); Tim Glennon (Accelrys)

Abstract: Understanding the protein-protein interactions is important for insights into signal transduction pathways. Here we have applied protein-protein docking, Evolutionary trace, fold, hydrophobic, and electrostatics analysis to determine and understand the interaction between a regulator of G-protein signaling protein and the alpha subunit of G-proteins.

Poster A-5

Comprehensive Laboratory information Management system (CLAM): the Structural Module

Tjaart de Beer (University of Pretoria); Fourie Joubert (University of Pretoria)

Abstract: The aim of this project is to construct an Open Source, web based functional genomic information system called CLAM (Comprehensive Laboratory information Management). CLAM will contain modules for genotyping, proteomics, genetics, phylogenetics, microarray, comparative genomics and structural biology data analysis. This poster will focus on the structural module in CLAM.

Poster A-6

Metal binding sites: pre-organized scaffolds in the unbound state

Mariana Babor (Weizmann Institute of Science); Harry Greenblatt (Weizmann Institute of Science);

Marvin Edelman (Weizmann Institute of Science); Vladimir Sobolev (Weizmann Institute of Science)

Abstract: Protein metal binding sites in the unbound state, and their rearrangements upon metal binding were analyzed. More than 40% of the metal binding sites show a capacity for flexibility, but in the vast majority of cases, part of the first coordination shell is already in place in the pre-bound form.

Poster A-7

Functional Prediction of Protein Mutants Using a Four-Body Potential

Majid Masso (George Mason University); Iosif Vaisman (George Mason University)

Abstract: Studies exploring single point mutants of HIV-1 protease and T4 lysozyme suggest that prediction of mutant enzyme catalytic activity is realizable by employing supervised learning in conjunction with mutant attribute vectors, based on a four-body statistical potential, that characterize constituent amino acid environmental changes from wild-type.

Poster A-8 (*There will also be an oral presentation of this poster.*)

Enzyme Mechanism Annotation and Classification

Daniel Almonacid (Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge); Gemma Holliday (Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge); Peter Murray-Rust (Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge); Janet Thornton (EMBL-EBI, Wellcome Genome Campus); John Mitchell (Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge)

Abstract: MACiE is a unique database containing enzyme-catalysed reaction mechanisms. Reaction steps as well as overall reactions are included. Data mining of this database has already provided a better insight into the nature's catalytic diversity. Our ongoing work addresses the evolution and classification of enzymes.

Poster A-9

Automated creation of in silico analogue ligand libraries from a lead molecule template

Wolf Cochrane (University Pretoria); Fourie Joubert (University Pretoria)

Abstract: LIGLIB (LIGand LIBrary Builder) is an open source tool that allows users to create an in silico library of molecules that are analogues of a lead chemical compound given as input to the software. The software is available as a plug-in to chimera.

Poster A-11 (*There will also be an oral presentation of this poster.*)

A novel approach to structural alignment using realistic structural and environmental information

Yu Chen (Bioinformatics program, University of Michigan); Gordon Crippen (College of Pharmacy, University of Michigan)

Abstract: We find a new structural alignment approach using realistic structural as well as environmental information. Statistics are defined to measure the goodness of alignments in structure cores. With this method, we can distinguish structures in different oligomeric states, and can flexibly align multiple domain proteins without domain splitting.

Poster A-12

Identifying Functional Signatures from Structural Alignments

Kai Wang (University of Washington); Ram Samudrala (University of Washington)

Abstract: We developed a method called Functional Signature from Structural Alignments (FSSA), to estimate the log odds of a residue being functionally important versus structurally important. The FSSA signatures can be used to interpret the functional importance of each residue, or classify proteins into functional categories.

Poster A-13

Doing a double take: function based target selection for structural genomics

Iddo Friedberg (The Burnham Institute); Phillip Lord (University of Manchester); Andrei Osterman (The Burnham Institute); Adam Godzik (The Burnham Institute)

Abstract: Structural genomics target selection schemes usually favor proteins predicted to have new folds. Here we argue that more targets should be selected within a given fold, to provide accurate templates not only for fold space, but also for function space, which is more finely grained.

Poster A-14

A new set of docking potentials for efficient discrimination between native and non-native conformations of protein complexes

Dror Tobi (Department of Computational Biology School of Medicine, University of Pittsburgh); Ivet Bahar (Department of Computational Biology School of Medicine, University of Pittsburgh)

Abstract: We generated putative docked complexes for a set of 63 non-redundant complexes, which were used in a linear programming algorithm to generate coarse-grained Docking Potentials. The resulting set of potentials show promising results for discriminating the native complex among decoys generated with the unbound form of the interacting proteins.

Poster A-16

Structural identification and prediction of amphipathic alpha-helices

Mamta Bajaj (School of Biological Sciences, University of Nebraska-Lincoln); Hideaki Moriyama (Department of Chemistry, University of Nebraska-Lincoln); Etsuko Moriyama (School of Biological Sciences and Plant Science Initiative, University of Nebraska-Lincoln)

Abstract: We developed a new method for identifying amphipathic alpha-helices based on PDB coordinate information, and identified 26 amphipathic alpha-helices that are not annotated as amphipathic alpha-helices in the PDB. Based on this dataset, we developed a new prediction method for amphipathic alpha-helices from primary structure information.

Poster A-17

Accurate Recognition of Protein-DNA Interaction Using Optimized Potential with Multi-body Consideration

Zhijie Liu (Computational Systems Biology Laboratory, Department of Biochemistry and Molecular Biology, University of Georgia); Ying Xu (Computational Systems Biology Laboratory, Department of Biochemistry and Molecular Biology, University of Georgia)

Abstract: A knowledge-based potential that considers the distance-dependent multi-body interactions was developed to quantitatively evaluate the binding affinity between protein and DNA. The potential achieved significant agreement between predictions and the experimental data, and succeeded in identification of DNA binding motifs of transcription factors in the genome-scale.

Poster A-18

Few strict rules determine permissible arrangements of strands in the Sandwich Proteins

Yih-Shien Chiang (Department of Health Informatics, SHRP, University of Medicine and Dentistry of New Jersey); Tatyana Gelfand (Department of Mathematics, Rutgers University); Thanasis Fokas (Department of Applied Mathematics and Theoretical Physics, University of Cambridge); Alexander Kister (Department of Health Informatics, SHRP, University of Medicine and Dentistry of New Jersey); Israel Gelfand (Department of Mathematics, Rutgers University)

Abstract: Analysis of the arrangements of strands in beta-sandwich proteins has led to propose a set of rules, which determine the main principles of the packing of strands in structures. These constraint rules allow one to determine all permissible motifs of the sandwich-like proteins. Keywords : Protein prediction, supersecondary structure.

Poster A-19

A Protein Structure Comparison system using 3D LRA

Chan-Yong Park (Electronics and Telecommunications Research Institute); Sung-Hee Park (Electronics and Telecommunications Research Institute); Dae-Hee Kim (Electronics and Telecommunications Research Institute); Seon-Hee Park (Electronics and Telecommunications Research Institute); Chi-Jung Hwang (Chung Nam University)

Abstract: The protein structure comparison using the LRA(Locally Relative Angle) is a algorithm of the efficient protein structure representation. The algorithm consists of two parts. The indexing part stores the LRA with all the proteins and the retrieval part adds the comparison process that compares the LRA of a query.

Poster A-20

Secondary Structure in the Target as a Confounding Factor in Synthetic Oligomer Microarray Design

Vladyslava Ratushna (Virginia Polytechnic Institute and State University); Jennifer Weller (George Mason University); Cynthia Gibas (Virginia Polytechnic Institute and State University)

Abstract: Prediction and thermodynamic analysis of secondary structure formation in a genome-wide set of transcripts from *Brucella suis* 1330 demonstrates that the properties of the target molecule have the potential to strongly influence the rate and extent of hybridization between transcript and tethered oligonucleotide probe in a microarray experiment.

Poster A-21

Protein Loop Modeling using Genetic Algorithms

Chiuan-Jung Chen (Department of Computer Science and Information Engineering, National Taiwan University); Chen-hsiung Chan (Department of Computer Science and Information Engineering, National Taiwan University); Cheng-Yan Kao (Department of Computer Science and Information Engineering, National Taiwan University)

Abstract: We have developed a robust protein loop modeling algorithm using genetic algorithm for conformation search. Using RMSD as fitness function, the prediction accuracy reaches 0.59 Å for 60 loops with length of 8 residues. Our loop modeling algorithm can evaluate the strengths of various scoring functions.

Poster A-22

A Fast Similarity Search System for Protein 3D Structure Databases Using Spatial Topological Relationships and Rtree Index

Sung-Hee Park (Database & Bioinformatics Laboratory, Chungbuk National University); Keun Ho Ryu (Database & Bioinformatics Laboratory, Chungbuk National University); David Gilbert (Bioinformatics Research Centre, Department of Computing Science, University of Glasgow)

Abstract: We develop a prototype system for fast similarity search for protein 3D structure databases based on spatial index and topological relationship patterns. Our approach can rapidly generate a small candidate set to be subsequently used in more accurate and slow alignment methods. The sever will be available <http://dmlab.chungbuk.ac.kr/simsearch.jsp>

Poster A-23 (*There will also be an oral presentation of this poster.*)

On the importance of being left-handed

Marian Novotny (Uppsala University); Gerard Kleywegt (Uppsala University)

Abstract: The handedness of helices has not received much attention in the past. Therefore, an extensive survey of left-handed helices was undertaken to analyse their frequency, length, amino acid composition and possible structural or functional role. The survey suggests that left-handed helices are rare, but structurally or functionally significant.

Poster A-24

The Energetics and Stability of Transmembrane Helix Packing: a Density of States Simulation

Zhong Chen (Dept. of Biochemistry and Molecular Biology, University of Georgia); Ying Xu (Dept. of Biochemistry and Molecular Biology, University of Georgia)

Abstract: Packing of transmembrane helices was successfully modeled by Wang-Landau simulations that calculate the density of states, from which the stabilities of different packing topologies were obtained. Contrary to common belief, helix-lipid interactions seem to be as important as helix-helix interactions for structure formation of some membrane proteins.

Poster A-25

In silico structure-based design of a potent, mutation resilient, small peptide inhibitor against Rifampicin-resistant tuberculosis

Deepak Bunger (Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh); Gita Subba Rao (All India Institute of Medical Sciences (AIIMS), New Delhi)

Abstract: Mycobacterium tuberculosis RNA polymerase (RNAP) is a key enzyme involved in the replication of the bacterium and is a potential target for therapeutic intervention following infection. We present here the design of a peptide inhibitor of RNAP. The designed peptide has the potential of being a novel and promising drug candidate against Rifampicin-resistant M.tuberculosis.

Poster A-26

An Improved Fully-Connected Hidden Markov Model for Rational Vaccine Design

Chenhong Zhang (Department of Computer Science, University of Saskatchewan); Anthony Kusalik (Department of Computer Science, University of Saskatchewan); Mik Bickis (Mathematical Sciences Group, University of Saskatchewan)

Abstract: The predictive accuracy of a rational vaccine design program based on a fully-connected HMM is improved via a biochemistry-based matrix initialization heuristic and a topology reduction heuristic. With the combination of approaches, the program outperforms HMMER on two alleles tested, HLA-A*0201 and HLA-B*3501.

Poster A-27

A graph theoretical approach for the Identification of Protein Domains

Frank Emmert-Streib (Stowers Institute for Medical Research); Arcady Mushegian (Stowers Institute for Medical Research)

Abstract: We propose a graph theoretical approach for the problem of protein domain identification by representing its three-dimensional structure as graph. The domains of the protein are then identified as partitions of the graph. These partitions are obtained by maximizing an objective function corresponding to the mutual maximization of cycle distributions.

Poster A-28

Homology modeling of the AdoMetDC domain from the bifunctional malarial enzyme S-adenosylmethionine decarboxylase/Ornithine decarboxylase.

Gordon Wells (University of Pretoria, Department of Biochemistry); Lyn-Marie Birkholtz (University of Pretoria, Department of Biochemistry); Fourie Joubert (University of Pretoria, Bioinformatics and Computational Biology Unit); Rolf Walter (Bernhard Nocht Institute for Tropical Medicine, Department of Biochemical Parasitology); Abraham Louw (University of Pretoria, Department of Biochemistry)

Abstract: The AdoMetDC domain of the bifunctional malarial enzyme S-adenosylmethionine decarboxylase/Ornithine decarboxylase was modeled based on the human and potato crystal structures. From this and related site-directed mutagenesis a number of novel properties can be predicted, which may aid the discovery of novel parasite-specific inhibitors.

Poster A-29

Predicting lipid accessible surface areas of transmembrane residues

Zheng Yuan (Institute for Molecular Bioscience and ARC Centre in Bioinformatics, The University of Queensland); Shane Zhang (Institute for Molecular Bioscience and ARC Centre in Bioinformatics, The University of Queensland); Mellisa Davis (Institute for Molecular Bioscience and ARC Centre in Bioinformatics, The University of Queensland); Mikael Boden (School of Information Technology and Electrical Engineering, The University of Queensland); Rohan Teasdale (Institute for Molecular Bioscience and ARC Centre in Bioinformatics, The University of Queensland)

Abstract: A Support vector regression approach has been used to predict lipid accessible surface areas (LASAs) of transmembrane residues. Based on a non-redundant dataset of 59 transmembrane helix proteins, we achieve a correlation coefficient 0.66 between predicted and observed LASAs by Jackknife tests. The mean absolute error can decrease to 19.6 squared Armstrong. Tested on 14 beta-barrel membrane proteins, the correlation coefficient and mean absolute error are 0.70 and 19.2 squared Armstrong, respectively. This approach is useful for prediction transmembrane domain arrangement.

Poster A-30

High-throughput exploration of functional residues in protein structures

Gabriele Ausiello (Centre for Molecular Bioinformatics, Dept. of Biology, Uni. of Tor Vergata, Rome); Andreas Zanzoni (Centre for Molecular Bioinformatics, Dept. of Biology, Uni. of Tor Vergata, Rome); Daniele Peluso (Centre for Molecular Bioinformatics, Dept. of Biology, Uni. of Tor Vergata, Rome); Allegra Via (Centre for Molecular Bioinformatics, Dept. of Biology, Uni. of Tor Vergata, Rome); Manuela Helmer-Citterich (Centre for Molecular Bioinformatics, Dept. of Biology, Uni. of Tor Vergata, Rome)

Abstract: pdbFun (pdbfun.uniroma2.it), a server for mass functional analysis of protein structures at residue level, integrates different databases and methods for 3D functional annotation together with a local structural comparison algorithm. pdbFun permits fast, detailed and high-throughput exploration of the whole PDB reorganized as an annotated residues DB.

Poster A-31

Sequence conservation and secondary structure - finding structural traits by using conservation as a magnifying glass.

Einat Sitbon (Weizmann Institute of Science); Shmuel Pietrokovski (Weizmann Institute of Science)

Abstract: What are structurally distinguishing features of conserved protein sequence regions? We found beta-strands abundant, alpha-helices rare, and certain combinations of secondary structures specific, in conserved sequence regions. These findings are relevant to basic science of protein structure, and to protein function prediction.

Poster A-32

Disparity in the nuclear localization signal of Stat1 and Stat3: Use of molecular modeling and visualization techniques for comparative analysis of relevant aspects of the crystal structures

Agnes Tan (Institute of Molecular and Cell Biology)

Abstract: Stat proteins possess distinct functions in the cytoplasm and in the nucleus. We have studied the crystal structures of Stat1 and Stat3 in view of differences in the nuclear localization signal. We have correctly predicted the relative importance of Arg214 in Stat3, and elucidated the packing of Leu407/411.

Poster A-33

Optimal relationship between average conformational entropy and average energy of residue interactions for fast protein folding

Oxana Galzitskaya (Institute of Protein Research, Russian Academy of Sciences); Sergiy Garbuzynskiy (Institute of Protein Research, Russian Academy of Sciences)

Abstract: Based on the known experimental data and using theoretical modeling of protein folding, we demonstrate that there exists optimal relationship between the average conformational entropy and the

average energy of contacts per residue for fast protein folding. Our result is in agreement with the experimental folding rates for 59 proteins.

Poster A-34

Computational analysis of RNA binding proteins based on composition, sequence and structural information

Parthiban Vijaya (Cologne University BioInformatics Center / International Max Planck Research School); Michael Gromiha (Computational Biology Research Center (AIST)); Abhinandan Madenhalli (Cologne University BioInformatics Center); Dietmar Schomburg (Cologne University BioInformatics Center)

Abstract: RNA binding proteins are involved in key roles in the regulation of gene expression. Critical analyses of protein-RNA complexes at sequence/structural level are needed to understand the RNA-protein interactions and related molecular processes. Statistical methods were developed to analyse complexes which facilitate better understanding of biological processes.

Poster A-35

B-cell epitope predictions based on three-dimensional structural information of protein antigens.

Pernille Haste Andersen (Center for Biological Sequence Analysis, BioCentrum-DTU, Technical University of Denmark,); Morten Nielsen (Center for Biological Sequence Analysis, BioCentrum-DTU, Technical University of Denmark,); Ole Lund (Center for Biological Sequence Analysis, BioCentrum-DTU, Technical University of Denmark,)

Abstract: We have constructed a data set of conformational B-cell epitopes and used it to the analyse the structural characteristics of B-cell epitopes. A well-performing B-cell epitope predictor has been developed, which includes characteristics of the three-dimensional structure of a pathogenic protein.

Poster A-36 (*There will also be an oral presentation of this poster.*)

Route as trees: The parsing view on protein folding

Julia Hockenmaier (University of Pennsylvania); Aravind Joshi (University of Pennsylvania); Ken Dill (University of California at San Francisco)

Abstract: Protein folding is a parallel, hierarchical process. Therefore, folding routes should be viewed as trees, not linear pathways. The Cocke-Kasami-Younger parsing algorithm is an efficient, accurate technique to search all folding route trees. It predicts the Plaxco et al. result that folding speed is inversely correlated with native contact order.

Poster A-37

Protein Structure from Contact Maps: An Hierarchical Approach

Alan Ableson (Queen's University); Jim Davies (Queen's University); Tony Kuo (Queen's University); Eduardo Zuviria (Queen's University); Janice Glasgow (Queen's University)

Abstract: One approach to prediction of protein structure from sequence is to predict a contact map and structural features, and then reconstruct the protein from its predicted contact map. This poster proposes a method for structure determination from contact maps using the experience embedded in the PDB.

Poster A-38

A probabilistic approach to the prediction of non-covalent residue contacts

David Cook (University of Glasgow); Pawel Herzyk (University of Glasgow)

Abstract: Here we present a new approach to the prediction of non-covalent residue contacts in proteins. Preliminary results demonstrate that incorporating multiple scales derived from a hidden site class model into a correlated mutation algorithm is able to improve the accuracy over a single scale/matrix model.

Poster A-39

How old is your fold?

Sanne Abeln (University of Oxford, Department of Statistics); Henry Winstanley (University of Oxford, Department of Statistics); Charlotte M. Deane (University of Oxford, Department of Statistics)

Abstract: We have created the first relative age estimation technique for protein folds. The ages presented show correlation with other protein age estimators and are used to investigate evolutionary pressure on fold topology and complexity. This shows for example very different age patterns of alpha/beta folds compared to small folds.

Poster A-40

Computational simulations suggest multiple routes for substrates and products in mammalian cytochrome P450s

Karin Schleinkofer (Department of Bioinformatics, Biocenter, University of Würzburg); Sudarko Sudarko (Department of Chemistry, Faculty of Mathematics and Natural Sciences, University of Jember); Peter J. Winn (European Molecular Biology Laboratory); Susanne K. Luedemann (European Molecular Biology Laboratory); Rebecca C. Wade (EML Research)

Abstract: By molecular dynamic simulation of a membrane-bound mammalian P450, the microsomal CYP2C5, substrate access and product egress routes are proposed that differ from those found in previous simulations of soluble bacterial P450s. This highlights the adaptability of the P450 fold to different substrates and to cellular localization.

Poster A-41

Finding motifs in RNA 3-D structures by complete enumeration of cycles of relations

Majid Behbahani (CIISE, Concordia University); Sbastien Lemieux (CIISE, Concordia University)

Abstract: We propose an algorithm that identifies motifs from an RNA 3-D structure by enumerating and comparing all cycles of a given length in the graph of relations (GOR). The GOR is an extension of the secondary structure including tertiary interactions. Using this algorithm several well known motifs were identified.

Poster A-42

Role of sequence and evolutionary information in DNA-binding sites in proteins

Shandar Ahmad (Department of Bioscience and Bioinformatics, Kyushu Institute of Technology); Akinori Sarai (Department of Bioscience and Bioinformatics, Kyushu Institute of Technology)

Abstract: We implemented a neural network based algorithm to utilize evolutionary information of amino acid sequences in terms of their position specific scoring matrices (PSSMs) for a better prediction of DNA-binding sites. An average of sensitivity and specificity using PSSMs is upto 8.7% better than the prediction with sequence information only.

Poster A-43

Predicting RNA secondary structure at temperatures other than 37 C

Zhi (John) Lu (University of Rochester); David Mathews (University of Rochester)

Abstract: In order to study RNA secondary structure formation in organisms such as thermophiles and in experiments performed at temperatures other than 37 C, nearest neighbor enthalpy parameters are derived from experimental results. Using enthalpy and free energy parameters for 37C, RNA secondary structure can be predicted at different temperatures.

Poster A-44

e-Protein: A Distributed Pipeline for Structure-based Proteome Annotation using GRID Technology

Shikta Das (London e-Science Centre, Department of Computing, Imperial College, London); Andrew McGough (London e-Science Centre, Department of Computing, Imperial College, London); Keiran Fleming (Structural Bioinformatics Unit, Faculty of Life Sciences, Imperial College, London); John Darlington (London e-Science Centre, Department of Computing, Imperial College, London); Michael

Sternberg (Structural Bioinformatics Unit, Faculty of Life Sciences, Imperial College, London)

Abstract: The e-Protein project aims to provide a fully automated distributed pipeline for large-scale structural and functional annotation of all major proteomes utilising GRID technologies. We are using ICENI - a grid middleware - allowing biologists to browse and monitor available services and compose workflow components within a graphical interface.

Poster A-45

Multiple Mapping Method: A new approach to improve sequence to structure alignments

Brajesh Rai (Department of Biochemistry, Albert Einstein College of Medicine); Andras Fiser (Department of Biochemistry and Seaver Foundation Center for Bioinformatics, Albert Einstein College of Medicine)

Abstract: A new approach, Multiple Mapping Method, has been developed to optimally combine fragments from alternative input alignments. On a benchmark dataset of 6500 template-target protein pairs, the alignments generated by this method consistently outperformed the average accuracy of input alignments.

Poster A-46

A study of the folding energy spectrum of RNAs

Jerome Waldispuhl (Boston College); Peter Clote (Boston College)

Abstract: An important aspect of the RNA folding process concerns the distribution and free energy of kinetic traps. We present a new algorithm which computes for a given RNA sequence, the Boltzmann partition function for all locally optimal secondary structures, and show new results which may help to characterize biological sequences.

Poster A-47

Identification of folding essential residues by looking at an extensive DB of the structure descriptors in Diamond STING

Paula Kuser (EMBRAPA/CNPTIA); Michel Yamagishi (EMBRAPA/CNPTIA); Luiz Borro (EMBRAPA/CNPTIA); Aduino Mancini (EMBRAPA/CNPTIA); Roberto Higa (EMBRAPA/CNPTIA); Goran Neshich (EMBRAPA/CNPTIA)

Abstract: Diamond STING suite of programs for comprehensive analysis of structure, function and stability is presented. We show here the in silico process of identification of the folding essential amino acids (previously determined by experiment) by means of range selecting for a set of the STING_DB parameters.

Poster A-48

Prediction of coaxial stacking configuration of helices in RNA multibranch loops.

Rahul Tyagi (University of Rochester Medical Center); David Mathews (University of Rochester Medical Center)

Abstract: A dynamic programming algorithm to predict the coaxial stacking configuration in RNA multibranch loops using free energy nearest neighbour parameters is presented. We show that coaxial stacking in crystal structures can be predicted with considerable success using thermodynamic parameters.

Poster A-49 (*There will also be an oral presentation of this poster.*)

All-Atom Modeling of RNA Conformational Changes

David Mathews (University of Rochester Medical Center); David Case (The Scripps Research Institute)

Abstract: The GG mismatches in the duplex rGCAGGCGUGC have been shown by NMR to be dynamic. Here, the minimum energy pathway for the conformational change is modeled with all-atom calculations using Nudged Elastic Band and the AMBER forcefield.

Poster A-50

Automated Protein Backbone Tracing in Electron Density Maps using Belief Propagation

Frank DiMaio (UW-Madison Computer Sciences Department); Jude Shavlik (UW-Madison Computer

Sciences Department); George Phillips (UW-Madison Biochemistry Department)

Abstract: One particularly time-consuming step in x-ray crystallography is interpretation of the electron density map. This paper describes an approach to automated backbone tracing in poor-quality density maps using belief propagation (BP). Several enhancements to BP are presented, making the algorithm feasible even for proteins several thousand residues in length.

Poster A-51

Structural Annotation Pipeline for Malaria proteins

Yolandi Joubert (University of Pretoria); Fourie Joubert (University of Pretoria)

Abstract: A pipeline for structural annotation of the *P.falciparum* proteins is being constructed. A series of established structure-related bioinformatics tools are included in the pipeline. Analyses are performed on a Linux cluster, with results being submitted to a PostgreSQL database. Once completed, it could be extended to other genomes.

Poster A-52

Function Inference Using Family-Specific Subgraph Fingerprints Mined from Protein Families

Deepak Bandyopadhyay (Department of Computer Science, University of North Carolina at Chapel Hill); Jun Huan (Department of Computer Science, University of North Carolina at Chapel Hill); Jinze Liu (Department of Computer Science, University of North Carolina at Chapel Hill); Jan Prins (Department of Computer Science, University of North Carolina at Chapel Hill); Jack Snoeyink (Department of Computer Science, University of North Carolina at Chapel Hill); Alexander Tropsha (Department of Medicinal Chemistry and Natural Products, School of Pharmacy, University of North Carolina at Chapel Hill); Wei Wang (Department of Computer Science, University of North Carolina at Chapel Hill)

Abstract: We propose a method for functional family inference by querying a new structure for occurrences of family-specific structural fingerprints mined from protein families using a graph representation. We compare against sequence, fold and other local structure based methods, and demonstrate applications to structural genomics targets and predicted structures.

Poster A-53

Evolutionary knowledge-based potentials for protein structure prediction

Alejandro Panjkovich (Pontificia Universidad Catolica); Andrej Sali (University of California, San Francisco); Marc Marti-Renom (University of California, San Francisco); Francisco Melo (Pontificia Universidad Catolica)

Abstract: A novel approach implementing evolutionary information into mean force potentials (MFPs) is presented and demonstrated to perform significantly better than current MFPs in fold assessment. The evolutionary potentials (EvPs) presented here are built in a fold-specific manner based on multiple sequence alignments and threading techniques.

Poster A-54

An Examination of Protein Stability Using Delaunay Tessellation that Includes Surface Hydration Effects

Gregory Reck (George Mason University); Iosif Vaisman (George Mason University)

Abstract: Several variations of a statistical potential function that include surface water effects are derived from Delaunay tessellation of a representative set of 1352 hydrated proteins. Each of the potential functions is correlated with previously reported experimental stability data from 366 single-point mutations of HIV reverse transcriptase.

Poster A-55

Diresidue neural network for the prediction of disulfide connectivity and ligand-bound cysteines

Fabrizio Ferre (Boston College); Peter Clote (Boston College)

Abstract: A novel diresidue neural network approach is used for the prediction of the cysteine disulfide connectivity and for the discrimination between ligand-bound, disulfide bond-involved and free cysteines. This method can be used to face problems for which protein sequences are more adequately modeled using diresidue, rather than monoresidue, position-specific scoring matrices.

Poster A-56

Comparative analysis of large protein structural families by TOPOFIT

Alex Abyzov (Biology Department, Northeastern University); Chesley Leslin (Biology Department, Northeastern University); Mounir Errami (Biology Department, Northeastern University); Valentin ILYIN (Biology Department, Northeastern University)

Abstract: Using novel TOPOFIT method we present a comparative analysis of several large protein families demonstrating: a clear identification of the common structural invariant, unambiguous and distinct clusters of proteins, strong correlation of active sites with the structural invariants, while reflecting the role of variable parts in specificity, recognition and flexibility

Poster A-57

Mining 3D-motifs using physical-chemical constraints: application to Cardiolipin binding sites

Dmitrii Polshakov (Ohio State University/Department of Chemistry/Department of Computer Science); Keith Marsolo (Ohio State University/Department of Computer Science); Srinivasan Parthasarathy (Ohio State University/Department of Computer Science)

Abstract: A new approach toward the discovery of biologically-meaningful structural motifs in proteins is presented. Using 3D-coordinates and a scaled set of physical-chemical properties, the approach is validated on several sets of functionally-related proteins. In addition, the first structural search on a subset of the membrane proteins containing Cardiolipin is performed.

Poster A-58

Computational Prediction of RNA-Binding Sites in Proteins Based on Amino Acid Sequence

Michael Terribilini (Iowa State University); Jae-Hyung Lee (Iowa State University); Changhui Yan (Iowa State University); Robert Jernigan (Iowa State University); Vasant Honavar (Iowa State University); Drena Dobbs (Iowa State University)

Abstract: We have developed a Naive Bayes classifier for predicting RNA-binding residues in proteins using only protein sequence as input. The classifier identifies interface residues with 86% accuracy, 0.35 correlation coefficient. To our knowledge, this approach provides the best available sequence- based prediction of protein-RNA interaction sites.

Poster A-59

Novel Approach to Multi-scale Modeling of Protein Structure, Folding, Dynamics and Function

Pratul Agarwal (Oak Ridge National Laboratory); Al Geist (Oak Ridge National Laboratory)

Abstract: An integrated view of protein structure, folding, dynamics and function is emerging where protein complexes are viewed as dynamical entities. We are developing novel theoretical and computational approaches to enable simulations of protein complexes on biologically relevant time-scales, and to investigate link between dynamics, folding and function (enzyme catalysis/biomolecular recognition).

Poster A-60

FILTREST3D: program for discrimination of protein structure models that match the restraints from experimental data

Marta Kaczor (International Institute for Molecular and Cell Biology); Michal Gajda (International Institute for Molecular and Cell Biology); Janusz Bujnicki (International Institute for Molecular and Cell Biology)

Abstract: We developed a method and a web server for discriminating among large number of protein structure models with "fuzzy" restraints derived from mutagenesis, chemical modification, and crosslinking experiments. Restraints include: distances between residues, amino acid burial and secondary structure. Tested on a set of ROSETTA decoys for restriction enzymes using restraints from mutagenesis and CD spectroscopy.

Poster A-61

Structural determinants of pKa shifts in RNA

Christopher Tang (Columbia University); Emil Alexov (Columbia University); Anna Marie Pyle (Yale University/HHMI); Barry Honig (Columbia University/HHMI)

Abstract: We describe the calculation of pKa shifts in RNA structures. We show that shifts in pKas are quantitatively accurate when compared to experiment and we describe the structural features of RNA that are responsible for changes in these ionization constants.

Poster A-62

Molecular Dynamics of SXR-SMRT Interactions Revealed the Preference of NR-interacting Domain ID2 over ID1

Ching (Nina) Wang (GSBS, UMDNJ-RWJMS); Chia-Wei Li (GSBS, UMDNJ-RWJMS); J. Don Chen (GSBS, UMDNJ-RWJMS); William Welsh (GSBS, UMDNJ-RWJMS)

Abstract: Steroid and xenobiotic receptor is a member of the orphan nuclear receptors that mediates mammalian xenobiotic response. SMRT can repress SXR-mediated transactivation by binding to cofactor site through its two interacting domains ID1 and ID2. Our molecular dynamics studies revealed essential interactions for the preference of ID2 over ID1.

Poster A-63

Domain-based Small Molecule Binding Site Annotation

Kevin Snyder (The Blueprint Initiative, 9th floor, 522 University Ave, Toronto ON; Samuel Lunenfeld Research Institute/University of Toronto, Toronto ON); Howard Feldman (The Blueprint Initiative, 9th floor, 522 University Ave, Toronto ON; Samuel Lunenfeld Research Institute/University of Toronto, Toronto ON); Brigitte Tuekam (The Blueprint Initiative, 9th floor, 522 University Ave, Toronto ON; Samuel Lunenfeld Research Institute/University of Toronto, Toronto ON); John Salama (The Blueprint Initiative, 9th floor, 522 University Ave, Toronto ON; Samuel Lunenfeld Research Institute/University of Toronto, Toronto ON); Michel Dumontier (The Blueprint Initiative, 9th floor, 522 University Ave, Toronto ON; Samuel Lunenfeld Research Institute/University of Toronto, Toronto ON); Christopher Hogue (The Blueprint Initiative, 9th floor, 522 University Ave, Toronto ON; Samuel Lunenfeld Research Institute/University of Toronto, Toronto ON)

Abstract: SMID-BLAST (<http://smid.blueprint.org/>) is a freely available, multi-purpose tool for the annotation and prediction of protein-small molecule interactions and binding sites. The tool uses NCBI's RPS-BLAST algorithm to identify domains in the query sequence and then looks these up in SMID, a database of domain-small molecule interactions generated from the PDB.

Poster A-64 (*There will also be an oral presentation of this poster.*)

A Novel Covariance Model Based RNA Motif Finding Algorithm

Zizhen Yao (University of Washington, Seattle); Walter L. Ruzzo (University of Washington, Seattle)

Abstract: CMfinder predicts RNA motifs in unaligned sequences. It is an expectation maximization algorithm using Covariance Models for motif description, carefully crafted heuristics for effective motif search, and a novel Bayesian framework for structure prediction combining folding energy and sequence covariation. It performs better than alternatives, and integrates directly with genome-scale homology search.

Poster A-65

Detecting Functional Sites in Protein Structures Using Dynamics Perturbation Analysis

Dengming Ming (Computer and Computational Science Division, Los Alamos National Laboratory); Michael E. Wall (Computer and Computational Science Division and Bioscience Division, Los Alamos National Laboratory)

Abstract: Recently, we introduced a theoretical framework to quantify allosteric effects in proteins. We have developed an algorithm which makes use of this framework to predict functional sites in protein structures. Here we present this algorithm and results of its performance in predicting ligand-binding sites for 298 protein/ligand structures.

Poster A-66

NMRQ: A Web Server for the Validation, Comparison and Analysis of Protein Structures Solved by NMR

Gary Van Domselaar (Depts. of Computing Science & Biological Sciences, University of Alberta); Paul Stothard (Depts. of Computing Science & Biological Sciences, University of Alberta); Trent Bjorndahl (Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta); David Wishart (Depts. of Computing Science & Biological Sciences, University of Alberta)

Abstract: NMRQ is a web server for assessing and visualizing the quality of NMR-derived protein structures. NMRQ uses chemical shift assignments, NOE restraints, structure ensemble superposition, and structure geometry to score and rank models and structural features. Results are presented as publication-quality graphical and textual HTML reports.

Poster A-68

A novel method for comparing topological models of protein structures enhanced with ligand information

Mallika VEERAMALAI (Bioinformatics Research Centre, Dept of Computing Science, University of Glasgow); David GILBERT (Bioinformatics Research Centre, Dept of Computing Science, University of Glasgow)

Abstract: Protein structure comparison methods plays a vital role in understanding structural and functional relationships between protein, also essential for estimating the evolutionary distance between proteins and protein families. Here, we present a novel protein structure comparison method based on 'TOPS Strings+ models' which are topological models enhanced with ligand interaction information.

Poster A-69

PROLOOP-C

Balaji Jayaraman (University of Missouri Kansas City School of Computing and Engineering); Deendayal Dinakarpanthian (University of Missouri Kansas City School of Computing and Engineering)

Abstract: Most proteins have parts of their structure that adopt an unstructured or random coil conformation. In some cases, these coils play an important functional role. ProLoop-C is a database of sequence independent clustering of random coils based on structural similarity.

Poster A-70 (*There will also be an oral presentation of this poster.*)

ncRNA genefinding in *C. elegans*

Shawn Stricklin (Department of Genetics, Washington University at St. Louis); Valerie Reinke (Department of Genetics, Yale University School of Medicine); Viktor Stolc (Center for Nanotechnology, NASA Ames Research Center); Sean Eddy (Howard Hughes Medical Institute, Washington University at St. Louis)

Abstract: We describe a computationally-directed screen for noncoding (ncRNA) genes based upon comparison of the genomes of *C. elegans*, *C. briggsae*, *C. remanei*, and more distantly-related nematodes. The ncRNA candidates identified by a multipronged computational approach are assayed using custom tiling microarrays targeting loci in all three nematode genomes.

Poster A-71

Exploiting Sequence and Structure Homologs to Identify Protein-Protein Binding Sites

Jo-Lan Chung (Department of Chemistry and Biochemistry, San Diego Supercomputer Center, University of California, San Diego); Wei Wang (Department of Chemistry and Biochemistry, University of California, San Diego); Philip Bourne (Department of Pharmacology, San Diego Supercomputer Center, University of California, San Diego)

Abstract: Structurally conserved residues, determined by the multiple structure alignments, were combined with other residue properties to predict protein-protein binding sites. The prediction results improved significantly and supported the hypothesis that in many cases protein interfaces require some residues to provide rigidity to minimize the entropic cost upon complex formation.

Poster A-72

Analysis and Prediction of Protein Ubiquitination Sites

Predrag Radivojac (Indiana University); Lilia Iakoucheva (The Rockefeller University)

Abstract: We describe the development of the ubiquitination sites predictor from a protein sequence. Our results indicate prevalence of E, D and phosphorylated residues in close proximity to Ub sites. The data provide evidence that Ub sites are preferentially located within disordered or flexibly ordered protein regions.

Poster A-73

A structure-based algorithm for the recognition of antifreeze proteins

Andrew Doxey (University of Waterloo); Mahmoud Yaish (University of Waterloo); Marilyn Griffith (University of Waterloo); Brendan McConkey (University of Waterloo)

Abstract: We present a simple and efficient structure-based algorithm capable of recognizing antifreeze proteins (AFPs) and their putative ice-binding surfaces. The algorithm discriminates AFPs from other structures in the protein data bank with high accuracy. We have applied the algorithm to identify a novel plant AFP.

Poster A-74

Predicting Protein Active Sites Using Protein Motion

Vinhthuy Phan (The University of Memphis); Sunder Tatta (The University of Memphis); Yongmei Wang (The University of Memphis)

Abstract: We describe a new method of predicting protein active sites based on motion. Information about active sites help biologists understand protein-protein interaction. Using the Elastic Network model to estimate global motion of proteins, we predict most deformed regions and hypothesize that these regions co-localize with active sites.

Poster A-75

Modelling Cotranslational Protein Folding

Fabien P.E. Huard (Department of Statistics, Macquarie University); Charlotte M. Deane (Department of Statistics, University of Oxford); G.R. Wood (Department of Statistics, Macquarie University)

Abstract: Cotranslational protein folding is acknowledged to occur. Simplified models of proteins (HP models) are used to explore the effect of key factors (such as surmountable energy barrier) on the difference between the native state of a cotranslationally folded protein and that of a protein folded from a fully extended state.

Poster A-76 (*There will also be an oral presentation of this poster.*)

Isostericity Matrices: Tools for Analyzing Recurrent Motifs and Structurally Aligning Homologous RNAs

Neocles Leontis (Bowling Green State University); Eric Westhof (Institut de Biologie Moleculaire et Cellulaire); Zirbel Craig (Bowling Green State University); Ali Mokdad (Bowling Green State University);

Jesse Stombaugh (Bowling Green State University); Michael Sarver (Bowling Green State University)

Abstract: Isostericity Matrices (IM) for non-Watson Crick basepairs are important tools for deriving sequence signatures of recurrent RNA motifs, scoring and refining RNA sequence alignments, and evaluating motif conservation across phylogeny. Progress in automating procedures for productive, iterative RNA structural sequence alignment based on IM will be described.

Poster A-77

Conditional Random Fields for RNA Structural Alignment

Kengo Sato (Keio University, Department of Biosciences and Informatics); Yasubumi Sakakibara (Keio University, Department of Biosciences and Informatics)

Abstract: We propose a novel approach for estimating the parameters including the substitution scores of base pairs and the state transition scores for RNA structural alignment with Conditional Random Fields, which can discriminate between correct alignments and incorrect ones most likely.

Poster A-78

The Effects of Quadratic (Two-Body) vs. Linear (Simplified Two-Body) Scoring Functions in Core Structure Threading

Natasha L. Sefcovic (NCBI / NLM / NIH and Biology Department, Johns Hopkins University); Aron Marchler-Bauer (NCBI / NLM / NIH); Anna R. Panchenko (NCBI / NLM / NIH); Stephen H. Bryant (NCBI / NLM / NIH)

Abstract: We directly compared a dynamic programming (DP) threading program to a Monte Carlo (MC) threading program to study the effects of quadratic vs. linear scoring functions. The MC program performed slightly better as measured by ROC, but surprisingly, not because of the scoring functions.

Poster A-79

Visualizing Bacterial tRNA Identity Determinants and Antideterminants Using Function Logos and Inverse Function Logos

Eva Freyhult (The Linnaeus Centre for Bioinformatics, Uppsala University); Vincent Moulton (School of Computing Sciences, University of East Anglia); David Ardell (The Linnaeus Centre for Bioinformatics, Uppsala University)

Abstract: Two extensions to sequence logo graphs, function logos and inverse logos, are introduced. These are useful for finding features that distinguish a subclass of sequences from a general sequence family and underrepresented sequence features or functions, respectively. We apply function and inverse function logos to structurally aligned bacterial tDNAs.

Poster A-80

Two applications of Delaunay contact matrices to the analysis of protein structures

Todd Taylor (George Mason University); Iosif Vaisman (George Mason University)

Abstract: We detail two applications of Delaunay contact matrices to the analysis of protein structures. First, a variation of the Ising-lattice domain definition method of W. Taylor is described. Second, we use MDS to find the dimensionality of Delaunay contact graphs and thereby define three characteristic length scales in proteins.

Poster A-81

Localization of protein binding sites within families of homologous proteins

Dmitry Korkin (University of California, San Francisco); Fred Davis (University of California, San Francisco); Andrej Sali (University of California, San Francisco)

Abstract: We analyze whether binding sites of homologous proteins are localized, ie whether they share similar relative positions on protein surfaces, irrespective of the identities of their binding partners. The analysis shows that 71% of the 1,884 SCOP domain families have binding sites with localization values greater than expected by chance.

Poster A-82

Unification of discrete and continuous effects on protein interfaces: an extension of the concept of hydrophobic effect and its application

Martin Jambon (The Burnham Institute); Christophe Geourjon (PBIL-IBCP); Francois Delfaud (MEDIT SA)

Abstract: We developed a computationally efficient system to represent proteins and identify functionally equivalent sites in structures that may not share any sequence or fold similarity, even locally. This involves microsites made of discrete and continuous components. This system is part of SuMo, online at <http://sumo-pbil.ibcp.fr>

Poster A-83

Portable virtual reality system using haptic device and naked eye 3D display, for molecular modeling.

Isao Okada (Tokyo Medical Dental University); Hiroshi Mizushima (Tokyo Medical Dental University); Takayuki Ohnishi (Tokyo Medical Dental University); Hiroshi Nagata (Tokyo Medical Dental University); Hiroshi Tanaka (Tokyo Medical Dental University)

Abstract: We have been developing Virtual Reality System for molecular modeling using haptic device. This time we have developed a easy-to-carry system for demonstration at the conference or at other labs using portable PC. We also integrated naked eye 3D display system for better reality.

Poster A-84

Characterization of Protein Structure Using Geometry and Topology

Bala Krishnamoorthy (Washington State University, Pullman, WA); J. Scott Provan (University of North Carolina at Chapel Hill, NC); Alexander Tropsha (University of North Carolina at Chapel Hill, NC)

Abstract: The alpha complex filtration of a protein represented by its alpha carbon atoms is analyzed. The topology of the neighborhood of a strand of residues is characterized by the largest connected components and holes in its filtration. A "motif" for 3D structure is characterized by the number of persistent components and holes, and their relative sizes.

Poster A-85

Sequence-Dependent Conformational Energy of DNA Derived from Molecular Dynamics Simulations: Towards the Understanding of Indirect Readout in Protein-DNA Recognition

Marcos J. Arauzo-Bravo (Kyushu Institute of Technology); Shandar Ahmad (Kyushu Institute of Technology); Satoshi Fujii (Kyushu University); Shigeori Takenaka (Kyushu University); Hidetoshi Kono (Neutron Research Center and Center for Promotion of Computational Science and Engineering, Japan Atomic Energy Research Institute); Nobuhiro Go (Neutron Research Center and Center for Promotion of Computational Science and Engineering, Japan Atomic Energy Research Institute); Akinori Sarai (Kyushu Institute of Technology)

Abstract: To estimate the sequence dependence of DNA conformation, we performed molecular dynamics of DNA including all possible tetranucleotide steps. From the MD trajectories we derived the equilibrium conformations and harmonic force field potentials. The force fields were applied to estimate the sequence specificity of free DNA and protein-DNA complexes

Poster A-87

IUPUI: Intrinsic Unstructured Protein Unsupervised-supervised Identifier - A Software Tool

Jack Yang (Indiana University School of Medicine, IUPUI); Mary Yang (Purdue University School of Electrical and Computer Engineering); Predrag Radivojac (Indiana University School of Medicine, IUPUI); Marc Cortese (Indiana University School of Medicine, IUPUI); Vladimir Uversky (Indiana University School of Medicine, IUPUI); Keith Dunker (Indiana University School of Medicine, IUPUI)

Abstract: Regions of proteins that have no definite tertiary structure are known as Intrinsically

Unstructured Protein (IUP) regions. We developed a software tool to aid in identifying such regions, called Intrinsically Unstructured Protein Unsupervised-supervised Identifier (IUPUI). We demonstrated the effectiveness of the IUPUI predictor, and compared favorably to existing approaches.

Poster A-88

The Victor/FRST Function for Model Quality Estimation

Silvio Tosatto (Dept. of Biology & CRIBI Biotech Centre, University of Padova)

Abstract: Scoring functions are widely used in the final step of model selection in protein structure prediction. A novel combination of pairwise, solvation and torsion angle potentials contain largely orthogonal information. Combining these features with a linear weighting function, a robust energy function for discrimination of native-like structures was constructed.

Poster A-89

Understanding the Origin of "High Energy" of ATP: Ab initio Studies of the Tri- and Diphosphate Fragments of Adenosine Triphosphate

Priti Hansia (Molecular Biophysics Unit, Indian Institute of Science); Nandini Guruprasad (Molecular Biophysics Unit, Indian Institute of Science); Saraswathi Vishveshwara (Molecular Biophysics Unit, Indian Institute of Science)

Abstract: Methyl triphosphate and methyl diphosphate in their different protonation states have been investigated at high levels of quantum mechanical calculations. The optimized geometries, molecular orbitals contributing to the high energy of ATP and dependence of vibrational frequencies on the number of phosphate groups and the charged states have been reported.

2 Ontologies and NLP

Poster B-1

Extraction of Transcript Diversity from Scientific Literature

Parantu Shah (EMBL); Peer Bork (EMBL)

Abstract: We developed an information extraction method specifically for extracting information about alternative transcript and associate information from scientific literature

Poster B-2

PTKB: Protein Translocation Knowledge-Base

Zhiyong Lu (University of Colorado School of Medicine); Philip Ogren (University of Colorado School of Medicine); Andrew Dolbey (University of Colorado School of Medicine); Larry Hunter (University of Colorado School of Medicine)

Abstract: Protein translocation, by which proteins are inserted into or across membranes, is essential to all living organisms. We propose to use NLP techniques to automatically transform GeneRIFs mentioning the intracellular transport into a formal knowledge representation that captures what is being transported, from where, to where and by what mechanisms.

Poster B-3

ORIGIN - An educational Ontology about the Central Dogma of Biology

Nuno T Alves (Portugal Telecom); Vitor Fonseca (IST/INESC-ID); Arsnio M Fialho (IST/BSRG); Ana T Freitas (IST/INESC-ID); H Sofia Pinto (IST/INESC-ID)

Abstract: This work presents an ontology about the Central Dogma of Molecular Biology processes for prokaryotic organisms, which will be connected to an inference engine to allow question answering. The main concepts represented include a definition of processes, activities, the roles and relations of the entities in the processes.

Poster B-4

MAO: Multiple Alignment Ontology

Julie Thompson (Institut de Genetique et de Biologie Moleculaire et Cellulaire); Patrice Koehl (UC Davis); Stephen Holbrook (Lawrence Berkeley National Laboratory); Kazutaka Katoh (Institut for Chemical Research, Kyoto University); Eric Westhof (Institut de Biologie Moleculaire et Cellulaire); Dino Moras (Institut de Genetique et de Biologie Moleculaire et Cellulaire); Olivier Poch (Institut de Genetique et de Biologie Moleculaire et Cellulaire)

Abstract: MAO is an ontology for data retrieval and exchange for DNA/RNA, protein sequence and structure alignment methods. MAO concepts cover the main features of multiple alignments and attributes are defined for residue conservation, structural location and function. MAO is available via the OBO web site (<http://obo.sourceforge.net/>).

Poster B-5 (*There will also be an oral presentation of this poster.*)

Where do we GO next? Refining the content of the Gene Ontology

Midori Harris (The Gene Ontology Consortium, EMBL-EBI)

Abstract: The Gene Ontology (GO) Consortium is committed to the continued refinement of its ontologies in response to the needs of database curators and many other users. The GO update procedure ensures that ontology changes are useful, accurate, logically consistent, and well documented.

Poster B-6

Mining Data from Mouse Mutagenesis Projects using Ontologies

Simon Greenaway (MRC Mammalian Genetics Unit); Georgios Gkoutos (MRC Mammalian Genetics Unit); Ann-Marie Mallon (MRC Mammalian Genetics Unit); John Hancock (MRC Mammalian Genetics Unit)

Abstract: We have applied our recently developed ontological schema for the description of mouse phenotypes to phenotype data from a major mouse mutagenesis project carried out at Harwell, U.K. and used data mining techniques identify internal correlations in the data. Results of the analysis will be presented.

Poster B-7 (*There will also be an oral presentation of this poster.*)

CGHGate: Array-CGH, Case Reports, Phenotypes and Biomedical Literature for Human Genome Annotation

Steven Van Vooren (Katholieke Universiteit Leuven, Department of Electrical Engineering, SISTA/BIOI); Nicole Maas (Center for Human Genetics, Universitaire Ziekenhuizen Gasthuisberg); Joris Vermeesch (Center for Human Genetics, Universitaire Ziekenhuizen Gasthuisberg); Yves Moreau (Katholieke Universiteit Leuven, Department of Electrical Engineering, SISTA/BIOI); Bart De Moor (Katholieke Universiteit Leuven, Department of Electrical Engineering, SISTA/BIOI)

Abstract: As Microarray-CGH is introduced into the clinical practice for the identification of submicroscopic genomic aberrations, tools to handle related data become essential for clinical geneticists. CGHGate is a web application that combines a constitutional cytogenetics database and tools for search, visualisation, genome annotation and data- and text-mining.

Poster B-8 (*There will also be an oral presentation of this poster.*)

Text-mining challenges for protein family database annotation.

Anna Divoli (The University of Manchester); Teresa Attwood (The University of Manchester)

Abstract: Different biological databases have different requirements for, and standards of, annotation. This work concerns the development of text-mining software to assist the curators particularly of protein family databases, using manually-crafted annotations to improve the design of a new decision-support tool, BioIE. We report here the results of this study.

Poster B-9

Combined data-mining of literature, gene/protein databases, and gene expression data

identify uncharacterized cancer-specific targets

Pavel Pospisil (Department of Radiology, Harvard Medical School, Harvard University); Lakshmanan Iyer (Bauer Center for Genomics Research, Harvard University); Amin Kassis (Department of Radiology, Harvard Medical School, Harvard University)

Abstract: We present a systematic data-mining study to identify cancer-gene targets. It covers literature, annotated sequence, and structure databases, as well as gene expression data sets. The results have allowed us to distinguish targets for the entrapment of radiolabeled compounds in the extracellular spaces of solid tumors.

Poster B-10 (*There will also be an oral presentation of this poster.*)

Ontological Visualization of Protein-Protein Interactions Using the Gene Ontologies

Harold Drabkin (Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, ME); Christopher Hollenbeck (Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY); David Hill (Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, ME); Mary Dolan (Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, ME); James Kadin (Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, ME); Judith Blake (Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, ME)

Abstract: Cellular processes require interaction of many proteins. Determining the collective network of such interactions will further understanding the role of individual proteins. The GO is used to provide functional annotation of proteins. We present a methodology for integrating and visualizing protein-protein interaction networks utilizing information encoded in GO annotations.

Poster B-11

Direct protein name recognition in full text, application to literature mining for receptor/G protein coupling interactions

Lei Shi (Weill Medical College of Cornell University); Fabien Campagne (Weill Medical College of Cornell University)

Abstract: The poster will present a computationally efficient method to extract protein names from full text without using a pre-existing protein name lexicon (such as collected from protein databases). We will also discuss its application to gathering information about the coupling of G Protein-Coupled Receptors to G proteins.

Poster B-12

Literature Based Functional Analysis of Microarray Data

Lai Wei (University of Tennessee Health Science Center); Kevin Heinrich (University of Tennessee); Lijing Xu (University of Tennessee Health Science Center); Michael Berry (University of Tennessee); Lawrence Pfeffer (University of Tennessee Health Science Center); Ramin Homayouni (University of Tennessee Health Science Center)

Abstract: We have developed an automated method which utilizes Latent Semantic Indexing (LSI) of titles and abstracts in MEDLINE citations to rank genes based on conceptual relationships to user defined keyword queries. Here, we demonstrate that this method provides a flexible tool for functional analysis of microarray expression data.

Poster B-13

Determining Domain-Specific Semantic Categories for Biological Named Entity Recognition System

Hyun-Sook Lee (Bioinformatics Research Team, Electronics and Telecommunication Research Institute, 161, Gajeong-dong, Daejeon, 305-350); Hyunchul Jang (Bioinformatics Research Team, Electronics and Telecommunication Research Institute, 161, Gajeong-dong, Daejeon, 305-350); Jasoo Lim (Bioinformatics Research Team, Electronics and Telecommunication Research Institute, 161, Gajeong-dong, Daejeon, 305-350); Soo-Jun Park (Bioinformatics Research Team, Electronics and Telecommunication Research

Institute, 161, Gajeong-dong, Daejeon, 305-350); Seon-Hee Park (Bioinformatics Research Team, Electronics and Telecommunication Research Institute, 161, Gajeong-dong, Daejeon, 305-350)

Abstract: To recognize named entities from bio-medical literature, appropriate semantic categories in a certain domain must be determined. This paper proposes a method of selecting domain-specific semantic categories automatically using UMLS without non-trivial tasks building domain knowledge. This method helps named entity recognizing system handle various domains effectively.

Poster B-14

A Corpus Tagging Tool and Rules for Biological Relation Events

Hyunchul Jang (Electronics and Telecommunications Research Institute); Hyun-Sook Lee (Electronics and Telecommunications Research Institute); Soo-Jun Park (Electronics and Telecommunications Research Institute); Seon-Hee Park (Electronics and Telecommunications Research Institute); Kyu-Chul Lee (Chungnam National University)

Abstract: We are creating a tagged corpus from MEDLINE abstracts to extract biological relationships. We are tagging named entities and their relation events. The categories of named entities and types of events cover most of UMLS semantic types. We made a tagging tool and defined tagging rules for this.

Poster B-15

Attack of the Clones: HL7 Clones vs. GenomicClones

Amnon Shabo (Shvo) (IBM Research Lab in Haifa)

Abstract: This poster presents HL7 standard specifications developed by the Clinical Genomics SIG. The core specification is the "Genotype" model which fuses bioinformatics markups into an HL7 schema, enabling the realization of the "Encapsulate & Bubble-up" paradigm and bridging the gap from genotypic to phenotypic clinical data.

Poster B-16

Structured Online Submission of Entity Relationships by Authors: Changing the Paradigm of Text Mining

Choon Kong Yap (Bioinformatics Institute); Sudhanshu Patwardhan (Bioinformatics Institute); Jagadish Hosagrahar Visvesvaraya (University of Michigan)

Abstract: Beyond Information Extraction, current literature mining tools fail to infer all relationships and derive accurate knowledge. Finer nuances of a paper can be captured and best represented only by the author. We propose structured online submissions of entity relationships by authors themselves, thus changing the paradigm of text mining.

Poster B-17

BIGRE: An Ontology Driven Bioinformatics Service Integration Environment

Olivier DUGAS (Universit Libre de Bruxelles); Joseph MAVOR (Universit Libre de Bruxelles); Pierre BUYLE (Facult Universitaire Notre-Dame de la Paix); Quentin DALLONS (Facult Universitaire Notre-Dame de la Paix); Amin MANTRACH (Universit Libre de Bruxelles); Utku SALIHOGLU (Universit Libre de Bruxelles); Hugues BERSINI (Universit Libre de Bruxelles); Vincent ENGLEBERT (Facult Universitaire Notre-Dame de la Paix); Marc COLET (Universit Libre de Bruxelles)

Abstract: BIGRE is a distributed generic service integration framework based on service semantics trying to solve the service interoperability problem. BIGRE uses ontologies describing service technical features and bioinformatics concepts. Client, Mediators and Wrapper are design-focused on easy accessability for end users and service integrators.

Poster B-18

Probe2GO: amplifying the GO annotation of Affymetrix probe sets

Enrique Muro (Ontario Genomics Innovation Centre, Ottawa Health Research Institute); Carolina Perez-Iratxeta (Ontario Genomics Innovation Centre, Ottawa Health Research Institute); Miguel Andrade

(Ontario Genomics Innovation Centre, Ottawa Health Research Institute)

Abstract: We present and evaluate a strategy for amplifying the GO annotations of entries from a sequence database. We applied it to the probes of the Affymetrix gene expression microarrays. The amplified annotations and the evidence supporting the annotation are accessible from a web server.

Poster B-19

Discovering Biomedical Domain-specific Action Vocabularies for Targeted Literature Mining

Merine Thomas (Dept of Computer and Information Science, IUPUI); Mathew Palakal (Dept of Computer and Information Science, IUPUI); Sudhanshu Patwardhan (Bioinformatics Institute, A*STAR); Muralidharan Kannan (Dept of Computer and Information Science, IUPUI)

Abstract: Each domain in biomedicine has a set of verbs or actions that get used more frequently and in a unique pattern in that particular domain compared to other domains. Discovering such sub-vocabularies has potential impact on setting rules for literature mining and the study gives confidence in the hypothesis.

Poster B-20 (*There will also be an oral presentation of this poster.*)

Human-Mouse Anatomical Ontology Mapping: Terminological and Structural Support

Sarah Luger (University of Edinburgh, Edinburgh, Scotland); Stuart Aitken (University of Edinburgh, Edinburgh, Scotland); Bonnie Webber (University of Edinburgh, Edinburgh, Scotland)

Abstract: Exploiting discoveries in mouse at a systems level requires the structuring mouse anatomical information in line with human anatomical information. We have found that we need to analyze both terminology and structure in order to support the alignment of anatomical ontologies between species, and propose automated methods for alignment.

Poster B-21

Extracting Genetic Pathways from Text and Grounding at the Spatio-Temporal Level

Gail Sinclair (University of Edinburgh); Bonnie Webber (University of Edinburgh); Duncan Davidson (Human Genetics Unit, Medical Research Council)

Abstract: In developmental biology, it is critical to link knowledge concerning genetic pathways with processes going on at cellular and tissue level. We are exploring methods of detecting and extracting information about such links from free text, by way of the description of events and their relations in space and time.

Poster B-22

Relating discrete annotation schemes in the functional space through literature analysis

Monica Chagoyen (Centro Nacional de Biotecnología - CSIC); Carlos Oscar S. Sorzano (Escuela Politécnica Superior, Universidad San Pablo-CEU); Pedro Carmona-Saez (Centro Nacional de Biotecnología - CSIC); Jose M. Carazo (Centro Nacional de Biotecnología - CSIC); Alberto Pascual-Montano (Dpto. Arquitectura de Computadores. Universidad Complutense de Madrid)

Abstract: We propose a methodology to create functional similarity measurements from data annotations using conceptual featural representations obtained from the analysis of relevant literature. The literature contains our current state of knowledge regarding gene function. Therefore, it is a good source of data from which to establish functional associations.

Poster B-23

A Novel Sentence Clustering Approach for Functional Annotation of Gene Expression Clusters

Jeyakumar Natarajan (Bioinformatics Research Group, University of Ulster); Eric G. Bremer (Brain Tumor Research Program, Children's Memorial Hospital, and Feinberg School of Medicine, Northwestern University); Catherine DeSesa (SPSS, Inc, Chicago); Catherine J. Hack (Bioinformatics Research Group, University of Ulster); Werner Dubitzky (Bioinformatics Research Group, University of Ulster)

Abstract: Information on gene function was extracted from fulltext using natural language processing and

sentence clustering and then used to interpret gene clusters identified in microarray data. Initial results have shown that the method effectively extracts information from fulltext, furthermore this information could not be identified through analysis of abstracts alone.

Poster B-24

Nearest Neighbor Categorization for CASP Function Prediction

Karin Verspoor (Los Alamos National Laboratory); Judith Cohn (Los Alamos National Laboratory); Susan Mniszewski (Los Alamos National Laboratory); Cliff Joslyn (Los Alamos National Laboratory)

Abstract: We present methods for protein function prediction, represented by Gene Ontology (GO) annotations. We identify neighbors of input sequences, collect GO nodes associated with these neighbors in Swiss-Prot, and categorize GO nodes utilizing Gene Ontology Categorizer technology. The resulting nodes are interpreted as the function of the original sequence.

Poster B-25

Machete: Carving Out Paths to Knowledge in Bioscience Literature

Shannon Bradshaw (The University of Iowa); Marc Light (The University of Iowa)

Abstract: Biologists face the daunting task of organizing volumes of scientific information. After retrieval of relevant articles, useful passages must be located. Compounding the problem is that the same information is extracted time and again by many individuals. Machete is a system targeted at these problems in managing bioscience knowledge.

Poster B-26

Transforming Full-Text Literature to Formalized Facts

Qing Dong (Stanford University); Rob Nash (Stanford University); Nicholas Stover (Stanford University); Christopher Lane (Stanford University); Shuai Weng (Stanford University); Rama Balakrishnan (Stanford University); Karen Christie (Stanford University); Maria Costanzo (Stanford University); Kara Dolinski (Princeton University); Stacia Engel (Stanford University); Dianna Fisk (Stanford University); Jodi Hirschman (Stanford University); Eurie Hong (Stanford University); Cynthia Krieger (Stanford University); Rose Oughtred (Princeton University); Marek Skrzypek (Stanford University); Chandra Theesfeld (Stanford University); Gail Binkley (Stanford University); Stuart Miyasato (Stanford University); Anand Sethuraman (Stanford University); Mayank Thanawala (Stanford University); Rey Andrada (Stanford University); David Botstein (Princeton University); J. Michael Cherry (Stanford University)

Abstract: To extract formalized facts from scientific literature, SGD builds an automated pipeline to collect full-text documents. Most of the documents archived have been reviewed by scientific curators and can serve as a training set for text-mining algorithms. We implemented Textpresso, a vocabulary-based information retrieval and extraction system developed at Wormbase.

Poster B-27

BLIMP: Biomedical Literature Mining Publications Forum; A Web-Based Resource

Hagit Shatkay (School of Computing, Queen's University, Ontario); Limin Zheng (School of Computing, Queen's University, Ontario)

Abstract: BLIMP is an online resource for compiling and sharing a complete bibliography on biomedical text mining. Bridging among the diverse research communities and publication venues, it holds hundreds of entries, features a search engine tailored for its scope, and supports submission of new items. (See <http://blimp.cs.queensu.ca>)

Poster B-28

Mining for Novel TNF Ligands using Unison, an Open Source Database for Target Discovery

Reece Hart (Genentech, Inc.)

Abstract: We describe Unison, our Open Source protein sequence and structure mining tool, and illustrate its applicability to mining for novel tumor necrosis factor ligands. Unison integrates protein

threading, HMM and PSSM alignment, localization, transmembrane, signal sequence, GPI, and other predictions to enable complex, holistic mining queries and facilitate therapeutic target discovery.

Poster B-29 (*There will also be an oral presentation of this poster.*)

Hubs of Knowledge: using the functional link structure in Biozon to mine for biologically significant entities

Paul Shafer (Cornell University); Timothy Isganitis (Cornell University); Golan Yona (Cornell University)

Abstract: We describe a system that builds upon the complex infrastructure of Biozon and applies methods equivalent to Google's PageRank to rank documents that match queries. We explore different models and study the spectral properties of their data graphs. A working ranking system of biological entities available at biozon.org

Poster B-30

GeneTegrate: a platform for integrating biology

Yanay Ofra (of Biochemistry and Molecular Biophysics, Columbia University); Guy Yachdav (of Biochemistry and Molecular Biophysics, Columbia University); Yechiam Yemini (Columbia University Center for Computational Biology and Bioinformatics (C2B2)); Sarah Gilman (Department of Computer Science, Columbia University); Burkhard Rost (Department of Biochemistry and Molecular Biophysics, Columbia University); Mark Treshock (Department of Computer Science, Columbia University)

Abstract: GeneTegrate provides a unifying semantic modeling layer to enrich, simplify and accelerate the analyses of distributed heterogeneous biological data. The system integrates many different levels of biological knowledge, from a single atom to a genetic network. More importantly, GeneTegrate displays the integrated data graphically, making the cognitive assimilation straightforward.

Poster B-31

MineOmics: Development of a Text Mining Tool that Provides Gene Information in Specific Disease and Biological Context

Matthew Tiller (Centers for Disease Control and Prevention); Eric Aslakson (Centers for Disease Control and Prevention); Suzanne Vernon (Centers for Disease Control and Prevention)

Abstract: Researchers are overwhelmed by voluminous high-throughput omic data. We introduce a pluggable text mining tool, called MineOmics. This tool utilizes statistical techniques and support vector machines to glean relevant information from electronic text repositories.

Poster B-32

Event Ontology: Biological ontology for annotating biological pathways and sub-pathways

Tatsuya KUSHIDA (Institute for Bioinformatics Research and Development, Japan Science and Technology Agency); Satoko YAMAMOTO (Institute for Bioinformatics Research and Development, Japan Science and Technology Agency); Takao ASANUMA (Institute for Bioinformatics Research and Development, Japan Science and Technology Agency); Emi HATTORI (Information and Mathematical Science Laboratory, Inc., Tokyo, JAPAN); Yuki YAMAGATA (Institute for Bioinformatics Research and Development, Japan Science and Technology Agency); Toshihisa TAKAGI (Graduate School of Frontier Sciences, University of Tokyo); Ken Ichiro FUKUDA (Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo JAPAN)

Abstract: Event Ontology is an ontology that organizes the terms of pathway objects such as sub-pathways, biological processes and experimental environments appearing in the biological pathways (e.g., signal transductions, disease pathways, metabolic pathways, etc.) The terms in the Event Ontology were manually extracted from scientific articles and text books.

Poster B-33

Maximum Entropy Modeling of Recognizing Biomedical Named Entities on the Literatures

Jaesoo Lim (Bioinformatics Research Team, Electronics and Telecommunications Research Institute, 161,

Gajeong-dong, Daejeon, 305-350); Hyun-Sook Lee (Bioinformatics Research Team, Electronics and Telecommunications Research Institute, 161, Gajeong-dong, Daejeon, 305-350); Hyunchul Jang (Bioinformatics Research Team, Electronics and Telecommunications Research Institute, 161, Gajeong-dong, Daejeon, 305-350); Soo-Jun Park (Bioinformatics Research Team, Electronics and Telecommunications Research Institute, 161, Gajeong-dong, Daejeon, 305-350); Seon-Hee Park (Bioinformatics Research Team, Electronics and Telecommunications Research Institute, 161, Gajeong-dong, Daejeon, 305-350)

Abstract: We suggest a Maximum Entropy modeling based on rich contextual features with closed vocabulary. We evaluated our resulting system on the GENIA corpus with the same way to the Bio-Entity Recognition Task at BioNLP/NLPBA 2004. The system exhibited a recall and a precision of 0.6332 and 0.6014, respectively.

Poster B-34 (*There will also be an oral presentation of this poster.*)

Literature Data Mining and Protein Ontology Development at the Protein Information Resource (PIR)

Zhang-Zhi Hu (Protein Information Resource, Georgetown University Medical Center); Inderjeet Mani (Department of Linguistics, Georgetown University); Hongfang Liu (Department of Information System, University of Maryland); Vijay Shanker (Department of Computer and Information Science, University of Delaware); Vincent Hermoso (Protein Information Resource, Georgetown University Medical Center); Anastasia Nikolskaya (Protein Information Resource, Georgetown University Medical Center); Darren Natale (Protein Information Resource, Georgetown University Medical Center); Cathy Wu (Protein Information Resource, Georgetown University Medical Center)

Abstract: A literature mining resource iProLINK is developed to provide data sources for research on literature-based curation and protein ontology development. A rule-based system RLIMS-P is benchmarked and used to extract phosphorylation information from PubMed abstracts, and a family classification-based protein ontology developed to complement other ontologies.

Poster B-35

Text Mining of MEDLINE Abstracts

Venu Dasigi (Southern Polytechnic State University); Sirisha Kanda (Southern Polytechnic State University); Sham Navathe (Georgia Institute Of Technology)

Abstract: A database is being constructed for supporting text mining on MEDLINE abstracts. This paper focuses on the usage of the data base. Creation of different views of the data for supporting different mining applications such as gene clustering, as well as practical issues relating to the size are described.

Poster B-36

Towards a comprehensive catalog of gene-disease and gene-drug relationships in cancer.

Christine M.E. Schueller (Biomax Informatics AG); Andreas Fritz (Biomax Informatics AG); Eduardo Torres Schumann (Biomax Informatics AG); Karsten Wenger (Biomax Informatics AG); Kaj Albermann (Biomax Informatics AG); George A. Komatsoulis (National Cancer Institute Center for Bioinformatics (NCICB)); Peter A. Covitz (National Cancer Institute Center for Bioinformatics (NCICB)); Lawrence W. Wright (National Cancer Institute Office of Communications); Frank Hartel (National Cancer Institute Center for Bioinformatics (NCICB))

Abstract: The National Cancer Institute (NCI) partnered with Biomax in order to expand the cancer gene section of the NCI Thesaurus to its full extent. Linguistic text analysis of Medline as well as thorough manual annotation using various ontologies was applied to populate a reference terminology with biologically meaningful content.

Poster B-37

Prediction of Protein Function Across Gene Ontology Terms

Roman Eisner (University of Alberta); Alona Fyshe (University of Alberta); Russell Greiner (University of

Alberta); Paul Lu (University of Alberta); Brandon Percy (University of Alberta); Brett Poulin (University of Alberta); Duane Szafron (University of Alberta); David Wishart (University of Alberta)

Abstract: We present a classification system which predicts the function of novel proteins. The predicted protein function(s) are terms within the Molecular Function aspect of Gene Ontology, which is a controlled vocabulary. We discuss our efforts to exploit the hierarchical structure of GO to increase our predictive accuracy and computational efficiency.

Poster B-38

Clustering of Pfam Protein Families in MeSH Space

Andreas Rechtsteiner (Los Alamos National Lab & Portland State University); Luis M Rocha (School of Informatics and Cognitive Science Program, Indiana University); Charlie E Strauss (Bioscience Division, Los Alamos National Lab)

Abstract: A large-scale, quantitative study of a literature mining algorithm for protein function prediction is presented. The test set is the Pfam protein sequence family classification. For 15200 proteins from 1611 Pfam families their family is predicted based on the MeSH terms of the literature associated with the proteins.

Poster B-39

WFLOW: A Browser Based Web Services Workflow Editor

James Long (University of Alaska Fairbanks); Tom Marr (University of Alaska Fairbanks)

Abstract: Workflows may be built using web services as building blocks. WFLOW is a browser-based workflow editor that uses Tigr Tree Menu, Graphviz, and gSOAP to build, display, and invoke web services workflows on our bioinformatics cluster. Future work will incorporate an ontology for our WSDL to constrain graph semantics.

Poster B-41

Mutation Miner

Christopher Baker (Concordia University); Rene' Witte (Universitaet Karlsruhe)

Abstract: Transfer of mutation specific raw-text annotations to protein structures requires an algorithm that can integrate natural language processing, database queries, sequence retrieval, sequence alignment and residue mapping. A multi component system is described for this purpose and we evaluate the use of text mining to drive protein structure visualization providing the protein engineer with enhanced access to the knowledge reported by multiple investigators.

Poster B-42

Quantitative Assessment for Relationship between Sequence Similarity and Function Similarity

Trupti Joshi (Digital Biology Laboratory, University of Missouri-Columbia, Columbia, MO); Dong Xu (Digital Biology Laboratory, University of Missouri-Columbia, Columbia, MO, USA)

Abstract: To quantify assignment errors in gene function prediction using comparative sequence analysis, we studied the relationship between sequence similarity and function similarity in terms of the three aspects of Gene Ontology (biological process, molecular function, and subcellular localization). Our study provides a benchmark to estimate the confidence in function assignment.

Poster B-43

Wnt Pathway Analysis with Automated Natural Language Processing

Carlos Santos (University of Michigan / Bioinformatics Program); Daniela Eggle (University of Michigan / Bioinformatics Program); David States (University of Michigan / Bioinformatics Program)

Abstract: We present a natural language processing pipeline to extract and analyze biomolecular interaction assertions from biomedical text. Focusing on the Wnt signaling pathway, the pipeline expands the existing canonical reference pathway with interaction assertions relating to that pathway, as well as

renders various organism-specific variations upon the canonical pathway.

Poster B-44

Improved Order Theoretical Techniques for GO Functional Annotation

Cliff Joslyn (LANL); Susan Mniszewski (LANL); Karin Verspoor (LANL); Judith Cohn (LANL)

Abstract: We present new order theoretical advances for the POSOC categorization algorithm applied to functional annotation: a pseudo-distance measure based on discrete Markov processes; an interval-valued rank measure in terms of vertical level in the GO; and order theoretical measures of horizontal distance based on so-called “fence” measures.

Poster B-45

Ontology-based pattern identification - a novel algorithm for gene function prediction

Yingyao Zhou (Genomics Institute of the Novartis Research Foundation); Jason Young (The Scripps Research Institute); Andrey Santrosyan (Genomics Institute of the Novartis Research Foundation); Kaisheng Chen (Genomics Institute of the Novartis Research Foundation); Frank Yan (Genomics Institute of the Novartis Research Foundation); Elizabeth Winzeler (Genomics Institute of the Novartis Research Foundation)

Abstract: Ontology-based Pattern Identification (OPI) is a novel data-mining algorithm that predicts gene function based on expression data and gene ontology. Instead of relying on a universal threshold of expression similarity, OPI automatically determines the optimal analysis settings that yield gene lists with highest statistical significance for function prediction.

Poster B-46

Semantic Model of NCI Thesaurus: Representing Genes and Alleles

Sherri de Coronado (National Cancer Institute Center for Bioinformatics); Gilberto Fragoso (National Cancer Institute Center for Bioinformatics); Francis Hartel (National Cancer Institute Center for Bioinformatics); Dan Lyman (IMC); Ranjana Srivastava (IMC)

Abstract: We present a vocabulary model of gene alleles developed for the NCI Thesaurus to satisfy various user needs for genes, alleles, diseases, drugs and related concepts including semantic relationships among gene classes, wild type and allelic variants, fusion genes, and oncogenes with related domains such as diseases, pathways and processes.

Poster B-47

Extraction and analysis of protein functional links from MEDLINE abstracts

Nikolai Daraselia (Ariadne Genomics, Inc); Sergei Egorov (Ariadne Genomics, Inc); Anton Yuryev (Ariadne Genomics, Inc); Andrey Yazhuk (Ariadne Genomics, Inc)

Abstract: We describe MedScan, a completely automated information extraction system, based on full sentence parsing. MedScan is tailored towards protein function information extraction, and was used to extract about 500,000 proteins functional links from the 2004 release of MEDLINE. A simple statistical analysis of the extracted data is presented.

Poster B-48

Extensions of the Gene Ontology in the Mouse Genome Informatics system

David P Hill (The Jackson Laboratory); Harold J Drabkin (The Jackson Laboratory); Mary E Dolan (The Jackson Laboratory); Li Ni (The Jackson Laboratory); Alexander D Diehl (The Jackson Laboratory); Christopher Hollenbeck (Rensselaer Polytechnic Institute); Joel E Richardson (The Jackson Laboratory); James A Kadin (The Jackson Laboratory); Judith A Blake (The Jackson Laboratory)

Abstract: The Mouse Genome Informatics (MGI) resource provides extensive information about laboratory mouse biology. The Gene Ontology (GO) is incorporated into MGI and used for functional annotations. GO advances within the MGI resource include user-friendly auto-text summaries, GO data analysis and visualization tools, extended annotation sets, and fully-integrated queries.

Poster B-49

Crossing of Subdiscipline Boundaries in the Biomedical Literature Explosion

Andrew Dolbey (Center for Computational Pharmacology, UCHSC); Lawrence Hunter (Center for Computational Pharmacology, UCHSC)

Abstract: In this poster, we show a case of subdiscipline boundary crossings in biomedical literature. The Medline citations for a single gene were collected. The spread of these citations across journals is tabulated, and then a graph of their distribution across subspecializations is demonstrated.

Poster B-50

An autonomous web service sequence analysis agent

Ayton Meintjes (Bioinformatics Unit, University of Pretoria); Fourie Joubert (Bioinformatics Unit, University of Pretoria)

Abstract: Describes the development of an automated software agent which accepts a sequence of interest and then queries various data sources for similar/related sequences. Metadata relating to these sequences are then retrieved and a subset of the results most likely to be of interest then further investigated.

Poster B-51

The Challenge of Phenotype Data: developing methods to access mouse phenotypes and human associations

Janan Eppig (The Jackson Laboratory); Howard Dene (The Jackson Laboratory); Susan Bello (The Jackson Laboratory); Megan Cassell (The Jackson Laboratory); Donna Burkart (The Jackson Laboratory); Ira Lu (The Jackson Laboratory); Linda Washburn (The Jackson Laboratory); Monika Tomczuk (The Jackson Laboratory); Anna Anagnostopoulos (The Jackson Laboratory); Cynthia Smith (The Jackson Laboratory)

Abstract: The mouse is the premier organism used as a model to study human biology and disease. The Mouse Genome Informatics (MGI) program is developing the Mammalian Phenotype Ontology for annotating mouse phenotypic data and providing integrated access to these data in human readable and computationally tractable formats.

3 Pathways, Networks and Proteomics

Poster C-1

Role of c-jun N-terminal MAP Kinase in rF1 induced activation of murine peritoneal macrophages

Rajesh Sharma (School of Biotechnology, Banaras Hindu University); Ajit Sodhi (School of Biotechnology, Banaras Hindu University); H. V. Batra (Division of Microbiology, DRDE)

Abstract: Fraction 1 of *Yersinia pestis* activated JNK MAP kinase. SP600125 inhibited the JNK phosphorylation. Where as, the rF1-induced JNK activity was correlated to inhibition of NO caused by SP600125 in the rF1-treated macrophages. Taken together, data suggests the involvement of JNK pathway in rF1 induced activation of macrophages.

Poster C-2

Bioinformatics tools to encode and integrate microscopy time-lapse sequences for drug discovery: Lineage analysis the basis for novel cell-based assays

Imtiaz Khan (Cardiff University); Lee Campbell (Cardiff University); Paul Smith (Cardiff University); Rachel J Errington (Cardiff University)

Abstract: Exploiting the potential for pharmacological modulation of tumour is a key goal for drug discovery. We describe here novel bioinformatics tools for encoding cell behaviour derived from time-lapse microscopy. The strategy is to inform mathematical models capable of predicting in silico drug signatures

for use in screening and therapeutics.

Poster C-3

A spatial resolution model for actin polymerization and fusion of lysosome phagosome.

Juilee Thakar (Department of Bioinformatics, University of Wrzburg); Mark Khnel (European Molecular Biology Laboratory); Gareth Griffiths (European Molecular Biology Laboratory); Thomas Dandekar (Department of Bioinformatics, University of Wrzburg)

Abstract: Some pathogens, example Mycobacterium Tuberculosis inhibit fusion of phagosome and lysosome leading to their survival in the phagosome. We developed a model to study formation of actin filaments on the phagosome membrane and in turn leading to fusion of phagosome with the lysosome to understand the critical steps in the process.

Poster C-4

Quantifying the relevance of different mediators in the human immune cell network

Paolo Tieri (Dept. Exp. Pathology Universit di Bologna); Silvana Valensin (C.I.G. Universit di Bologna); Vito Latora (Dept. Physics Universit di Catania); Gastone Castellani (C.I.G. Universit di Bologna); Daniel Remondini (C.I.G. Universit di Bologna); Massimo Marchiori (W3C MIT Lab for Computer Science); Claudio Franceschi (C.I.G. Universit di Bologna)

Abstract: Immune cells communicate through secreted mediator proteins. We present a method for quantifying the relevance of these mediators in an immune network where cells are nodes and mediators are the connecting links. Our results reveal that few mediators play a prominent role in the interactions among the immune cell types.

Poster C-5

Fusion of Multiple Decision Models in Proteomic Biomarker Discovery.

Asha Thomas (University of Louisville - Department of Computer Engineering and Computer Science); Georgia Tourassi (University of Louisville - Department of Computer Engineering and Computer Science); Adel Elmaghraby (University of Louisville - Department of Computer Engineering and Computer Science); Nigel G Cooper (University of Louisville - Anatomical Sciences and Neurobiology); Sumanth D Prabhu (University of Louisville - Medicine); Saeed A Jortani (University of Louisville - Pathology and Laboratory Medicine); Roland Valdes Jr (University of Louisville - Pathology and Laboratory Medicine)

Abstract: We investigated the feasibility of combining linear and nonlinear decision models to improve the discriminatory performance for protein mass spectrometry data of heart failure patients. The results obtained show that the fusion of multiple decision models is a promising approach in proteomic data analysis.

Poster C-7

Comprehensive Network Analysis of Glaucoma Pathophysiology

Yuri Nikolsky (GeneGo Inc.); Tatiana Nikolskaya (GeneGo, Inc.); Eugene Kirillov (GeneGo, Inc.); Eugene Rakhmatulin (GeneGo, Inc.); Svetlana Sorokina (GeneGo, Inc.); Tatiana Serebrijskaya (GeneGo, Inc.); Sean Ekins (GeneGo, Inc.); Andrej Bugrim (GeneGo, Inc.); Dmitri Novikov (University of Illinois); Valery Shestopalov (University of Miami); Robert Haselkorn; Dmitry Ivanov; Vadim Brodianskir; Olga A. Agapova; M. Rosario Hernandez

Abstract: We developed a general approach for assembly, prioritization and analysis of biological networks implicated in complex diseases using heterogeneous experimental datasets and known human protein interactions. Specifically, we studied the networks affected in optic nerve head astrocytes in primary open angle glaucoma based on microarray gene expression and genetic data.

Poster C-8

Conditional network analysis: exploring network dynamics and identifying key modulator genes from gene expression data

Kai Wang (Joint Centers for Systems Biology, Columbia University); Nilanjana Banerjee (Joint Centers for Systems Biology, Columbia University); Adam Margolin (Joint Centers for Systems Biology, Columbia University); Ilya Nemenman (Joint Centers for Systems Biology, Columbia University); Katia Basso (Institute of Cancer Genetics, Columbia University); Riccardo Dalla-Favera (Institute of Cancer Genetics, Columbia University); Andrea Califano (Joint Centers for Systems Biology, Columbia University)

Abstract: We develop a systematic approach for identifying key modulators of transcriptional interactions. By reverse-engineering thousands of cellular networks conditioned on the expression of candidate modulator genes, we identify putative modulators which cause statistically significant changes in network topology. These are indeed enriched in GO categories involved in cellular regulations.

Poster C-9

GASP: GC/MS Analysis Software Package

Paulo Augusto Suano Nuin (McMaster University, Dept of Biology); Elizabeth Weretilnyk (McMaster University, Dept of Biology); Peter Summers (McMaster University, Dept of Biology); David Guevara (McMaster University, Dept of Biology); Brian Golding (McMaster University, Dept of Biology)

Abstract: The GC/MS Analysis Software Package (GASP) allows for the comparison of data between different GC/MS experiments, and makes possible a comparative analysis of all chromatographically separated compounds between different GC/MS runs.

Poster C-10

Reconstruction of Genetic Regulatory Networks Using the Network Inference Testbed Software Environment

Ronald Taylor (Pacific Northwest National Laboratory (US Dept of Energy)); William Cannon (Pacific Northwest National Laboratory (US Dept of Energy))

Abstract: 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.

Poster C-11

PATIKA: An informatics infrastructure for cellular networks

Emek Demir (Bilkent University); Asli Ayaz (Bilkent University); Ozgun Babur (Bilkent University); Ahmet Cetintas (Bilkent University); Ugur Dogrusoz (Bilkent University); Emine Zeynep Erson (Bilkent University); Erhan Giral (Bilkent University); Cagri Aksay (Bilkent University); Fatma Arik (Bilkent University); Esra Ataer (Bilkent University); E. Belviranli; R. Colak; G. Cozen; A. Dilek; E. Kaya; H. Yildirim

Abstract: The PATIKA Project aims for an informatics infrastructure to cope with the inherently complex cellular pathway data and provides software tools with sophisticated visualization technology around a central database using an extensive ontology and data integration mechanisms. It also features advanced database querying, microarray data analysis, and automatic layout components.

Poster C-12

Discovery of biological networks from diverse functional genomic data

Chad Myers (Princeton University); Drew Robson (Princeton University); Adam Wible (Princeton University); Chandra Theesfeld (Saccharomyces Genome Database); Kara Dolinski (Princeton University); Olga Troyanskaya (Princeton University)

Abstract: We have developed a general system for discovery of biological pathways from diverse functional genomic data. Our methodology employs a Bayesian network to integrate 9 different types of evidence for protein relationships from over 950 publications and a graph search algorithm designed for recovering functionally coherent groups of proteins.

Poster C-13

A Sequence-Based Characterization Of Human Proteins Localized Within Cell Compartments

George Acquaaah-Mensah (Massachusetts College of Pharmacy and Health Sciences); Sonia Leach (University of Colorado School of Medicine); Cary Miller (University of Colorado School of Medicine)

Abstract: Exploratory Data Analysis was used to characterize amino acid sequences of human proteins in sub-cellular localizations. Based on hydrophobicity, polarity, polarizability, normalized van der Waals volume and charge, descriptions of amino acid composition, transitions and distribution for proteins localized in the nucleus, cytosol, plasma membrane and mitochondrion are provided.

Poster C-14

Combining Alignment and N-grams in G-Protein Coupling Specificity Prediction

Betty Cheng (Language Technologies Institute, School of Computer Science, Carnegie Mellon University); Jaime Carbonell (Language Technologies Institute, School of Computer Science, Carnegie Mellon University); Judith Klein-Seetharaman (Department of Pharmacology, University of Pittsburgh Medical School)

Abstract: Understanding the signalling mechanism of G-protein coupled receptors requires knowledge of the G-proteins a given receptor can couple with. By combining n-grams and alignment information and using the whole receptor instead of focusing on the intracellular regions as in previous studies, our coupling specificity prediction method outperforms the current state-of-the-art.

Poster C-15 (*There will also be an oral presentation of this poster.*)

Multiple Knockouts Analysis of Genetic Robustness in the Yeast Metabolic Network

David Deutscher (Tel Aviv University); Isaac Meilijson (Tel Aviv University); Eytan Ruppin (Tel Aviv University)

Abstract: Genetic robustness, a constant phenotype in face of genetic perturbations, is widespread in biological systems. By analyzing results of multiple concurrent knockouts to the metabolic genes of *S.cerevisiae*, we provide the first large-scale study of metabolic network robustness, portraying its architecture and shedding new light on its evolution.

Poster C-16

Conversion of CellML into SBML

Maria Schilstra (Biocomputation Research Group, STRI, University of Hertfordshire); Joanne Matthew (Biocomputation Research Group, STRI, University of Hertfordshire); Michael Hucka (Control and Dynamical Systems, Caltech); Andrew Finney (Biocomputation Research Group, STRI, University of Hertfordshire)

Abstract: SBML and CellML are XML-based standard languages for describing biochemical models, with comparable scope and representation. We have created a CellML to SBML transformation tool (in XSLT) that is capable of converting 95% of the models in the CellML model repository into valid SBML without loss of information.

Poster C-17

Method for quantitation of MS-peaks from 18O/16O labeled phospho-peptides

Claus A. Andersen (Siena Biotech SpA, Discovery Research); Stefano Gotta (Siena Biotech SpA, Discovery Research); Roberto Raggiaschi (Siena Biotech SpA, Discovery Research); Andreas Kremer (Siena Biotech SpA, Discovery Research); Letizia Magnoni (Siena Biotech SpA, Discovery Research); Georg C. Terstappen (Siena Biotech SpA, Discovery Research)

Abstract: The effect of amyloid- on neuronal cells was quantified from 18O/16O labeled phospho-peptides analyzed by MS. The peptide abundance was estimated from three overlapping isotope contours using multivariate linear regression. The spectra matched the theoretical reconstruction nicely ($\pm 10\%$ error, 90% of spectra) with a high degree of robustness.

Poster C-18 (*There will also be an oral presentation of this poster.*)

Dynamic complex formation during the yeast cell cycle

Ulrik de Lichtenberg (Center for Biological Sequence Analysis, Technical University of Denmark); Lars Juhl Jensen (European Molecular Biology Laboratory (EMBL), Heidelberg); Soren Brunak (Center for Biological Sequence Analysis, Technical University of Denmark); Peer Bork (European Molecular Biology Laboratory (EMBL), Heidelberg)

Abstract: To analyze the dynamics of protein complexes during the yeast cell cycle, we integrated data on protein interactions and gene expression. We find that all complexes contain both static and dynamic subunits, and propose a mechanism in which the dynamic subunits control the timing of complex assembly and function.

Poster C-19

Construction of theoretical spectra for peptide tandem mass spectra identification through database search

Tema Fridman (UT/ORNL); Vladimir Protopopescu (ORNL); Greg Hurst (ORNL); Andrei Borziak (ORNL); Andrey Gorin (ORNL)

Abstract: We derive the dependence of the probability of false identification on the number of peaks in the theoretical spectra and on the types of ions that the peaks represent. It is shown that inclusion of neutral loss ions into the theoretical pattern sharply raises the false identification rate.

Poster C-20

Comparative analysis of computationally predicted bacterial protein subcellular localization: perspectives on network complexity and the need for more diverse training data

Jennifer Gardy (Dept. of Molecular Biology and Biochemistry, Simon Fraser University); Fiona Brinkman (Dept. of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, B.C., Canada)

Abstract: Subcellular localization is generally conserved regardless of genome size, suggesting that evolution involves acquisition of functional networks spanning multiple localizations. Some variation in localization is observed, however, and appears to correlate with "non-model" organisms and poorly characterized localization sites, highlighting the need for a phylogenetically and compartmentally diverse dataset.

Poster C-21

User-guided Extraction of Biological Networks from Scientific Literature

Aditya Vailaya (Agilent Technologies); Allan Kuchinsky (Agilent Technologies); Annette Adler (Agilent Technologies); Michael Creech (Agilent contractor)

Abstract: We present a meta-search tool for automatically querying multiple text-based search engines in order to aid biologists faced with the daunting task of manually searching and extracting associations among genes/proteins of interest. Computationally extracted associations are grouped into a network that is viewed and manipulated in Cytoscape.

Poster C-22

Probabilistic paths in protein interaction networks

Hailiang Huang (Department of Biomedical Engineering, Johns Hopkins University); Lan Zhang (Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School); Frederick Roth (Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School); Joel Bader (Department of Biomedical Engineering, Johns Hopkins University)

Abstract: Understanding how proteins are organized into complexes and pathways is increasingly based on high-throughput experiments, which is unreliable, however, due to high false-positives and false-negatives. Algorithms are needed to predict protein complexes from the error prone data. We have developed and compared protein complex inference algorithms using pre-calculated confidence scores.

Poster C-23

Systems Approach to High Throughput Data Analysis

Aditya Vailaya (Agilent Laboratories); Allan Kuchinsky (Agilent Laboratories); Robert Kincaid (Agilent Laboratories); Raymond Tabibiazar (Stanford University); Roger Wagner (Stanford University); Jennifer King (Stanford University); Rossella Ferrara (Stanford University); Euan Ashley (Stanford University); Thomas Quertermous (Stanford University)

Abstract: We present statistical methods for identifying discriminatory biological processes via multiple (pathways, literature associations, and GO) analysis of differentially expressed genes. For complex diseases involving multiple processes, we construct literature-based de novo networks and identify "nexus" genes as potential regulatory therapeutic targets. Results are presented on three microarray datasets.

Poster C-24

Intrinsic Bayesian estimates of the eukaryotic protein interaction network complexity

Hailiang Huang (Department of Biomedical Engineering, Johns Hopkins University); Myung Lee (Department of Biomedical Engineering, Johns Hopkins University); Andy Cheng (Department of Biomedical Engineering, Johns Hopkins University); Joel Bader (Department of Biomedical Engineering, Johns Hopkins University)

Abstract: Despite progress in protein interaction screens, the total number of protein-protein interactions remains uncertain. We present a method for an intrinsic estimate of the number of total interactions based on samples observed. Intrinsic estimates for the completion of screens for *S. cerevisia*, *C. elegans*, and *D. melanogaster* have been obtained.

Poster C-25

Predicting pathways from graph structure containing gene expression similarity and protein-protein interaction

Ho-Youl Jung (ETRI); Ji Eun Kim (ETRI); Seon-Hee Park (ETRI)

Abstract: Our system predicts the biological pathway using mRNA expression pattern and protein-protein interaction data. We can make the graph structure whose vertices (edges) are composed of the genes (the similarity). In this similarity graph, we can predict the sub-pathway using the shortest path algorithm. We can also predict the sub-pathway using the shortest path algorithm in interaction graph. We apply the graph matching algorithm to sub-pathways, and make the unified final pathway.

Poster C-26

Mathematical modeling of molecular genetic network controlling cholesterol homeostasis in vertebrate cells

Alexander Ratushny (Institute of Cytology and Genetics SB RAS); Vitaly Likhoshvai (Institute of Cytology and Genetics SB RAS)

Abstract: We present a new version of a mathematical model of molecular genetic network controlling cholesterol homeostasis in vertebrate cells with more detailed described regulatory mechanisms. There were analyzed different kinds of genetic and metabolic disorders of cholesterol biosynthesis and cholesterol uptake into the cell and probable consequences of such events.

Poster C-27

Knowledge-Based Framework for Hypothesis Formation in Biochemical Networks

Nam Tran (Arizona State University); Chitta Baral (Arizona State University); Vinay Nagaraj (Arizona State University); Lokesh Joshi (Arizona State University)

Abstract: A framework for hypothesis formation is presented, which supports: seamless integration of hypothesis formation with elaboration tolerant knowledge representation and non-monotonic reasoning; use of various resources of biological data as well as human expertise to intelligently generate hypotheses; and ranking hypotheses and for designing experiments to verify hypotheses.

Poster C-28 *(There will also be an oral presentation of this poster.)*

Genome-wide transcription regulatory circuits controlling cellular malignant

Yuval Tabach (The Department of Complex System, Weizmann Institute of Science); Michael Milyavsky (The Department of Molecular Cell Biology Weizmann Institute of Science); Or Zuk (The Department of Complex System, Weizmann Institute of Science); Assif Yitzhaki (The Department of Complex System, Weizmann Institute of Science); Paz Polak (The Department of Complex System, Weizmann Institute of Science); Eytan Domany (The Department of Complex System, Weizmann Institute of Science); Varda Rotter (The Department of Molecular Cell Biology Weizmann Institute of Science); Yitzhak Pilpel (The Molecular Genetics Weizmann Institute of Science)

Abstract: Here we analyzed a 600-days in-vitro malignant transformation process. Following genome-wide transcription-profiling and promoter analysis we focused on a cell-cycle related cluster that is prominent in a diversity of cancers. Our work links three levels in a complex regulatory network, namely, gene expression, promoter architecture and activity of tumor suppressors.

Poster C-29

A polyketide database and analysis system for microbial genomes.

Hongseok Tae (Dept. of Computer Engineering, Chungnam National University); Hyeweon Nam (Information Technology Institute, SmallSoft Co., Ltd.); Jaekyung Song (Sun moon University); Kiejung Park (Information Technology Institute, SmallSoft Co., Ltd.)

Abstract: In this paper, we describe a system named ASMPKS(Analysis System of Modular Polyketide Synthesis) for the overall computational management of modular polyketide synthesis including PKS database construction, new polyketide assembly, visualization of a polyketide structure, and PKS prediction against genome sequences. ASMPKS operates on web interface to construct the database and to analyze polyketide.

Poster C-30

Commensurate Distances and Motifs from Genetic Congruence and Protein Interaction Networks in Yeast

Ping Ye (Johns Hopkins University); Joel Bader (Johns Hopkins University)

Abstract: We have demonstrated that the genetic congruence network inferred from direct genetic interactions largely overlaps the protein interaction network, with corresponding distances and transitive motifs, while the genetic interaction network does not. This finding suggests that very strong genetic congruence can be used to predict novel protein interactions.

Poster C-31

Microbial Pathway Mapping using Diverse Information

Fenglou Mao (University of Georgia); Zhengchang Su (University of Georgia); Victor Olman (University of Georgia); Ying Xu (University of Georgia)

Abstract: A new algorithm, P-MAP, for pathway mapping from a template genome to a target genome is presented. P-MAP maps pathways through application of sequence similarity information and genomic structure information such as operons and regulons in both the template and target genomes. A P-MAP web server has been built for public service. The evaluation of P-MAP shows that the program outperformed the existing methods by a large margin in both pathway mapping and orthologs prediction.

Poster C-32

DMSP - Database for Modeling Signaling Pathways. A new way of combining biological and mathematical modeling information concerning cell signaling pathways.

Mahesh Visvanathan (Institute for Biomedical Signal Processing and Imaging,UMIT); Marc Breit (Institute for Biomedical Signal Processing and Imaging,UMIT); Bernhard Pfeifer (Institute for Biomedical Signal Processing and Imaging,UMIT); Robert Modre-Osprian (Institute for Biomedical Signal Processing

and Imaging,UMIT); Bernhard Tilg (Institute for Biomedical Signal Processing and Imaging,UMIT)

Abstract: Understanding signaling pathways will have a number of potential applications for medications and drug design. The challenge here is to derive a way to quantitatively understand signaling pathways from a qualitative level. To address this issue we developed DMSP -Database for Modeling Signaling Pathways storing related biological and modeling information.

Poster C-33

Intraflagellar proteins of Leishmania spp., as assessed through in silico analysis, unveil a flagellar remodeling network

Joao J.S. Gouveia (Ncleo de Genmica e Bioinformtica, Faculdade de Veterinria, Universidade Estadual do Ceara (UECE)); Nilo B. Diniz (Ncleo de Genmica e Bioinformtica, Faculdade de Veterinria, Universidade Estadual do Ceara (UECE)); Elton J.R. Vasconcelos (Ncleo de Genmica e Bioinformtica, Faculdade de Veterinria, Universidade Estadual do Ceara (UECE)); Ana C.L. Pacheco (Ncleo de Genmica e Bioinformtica, Faculdade de Veterinria, Universidade Estadual do Ceara (UECE)); Michely C. Diniz (Ncleo de Genmica e Bioinformtica, Faculdade de Veterinria, Universidade Estadual do Ceara (UECE)); Allan R.S. Maia (Ncleo de Genmica e Bioinformtica, Faculdade de Veterinria, Universidade Estadual do Ceara (UECE)); Daniel A. Viana (Ncleo de Genmica e Bioinformtica, Faculdade de Veterinria, Universidade Estadual do Ceara (UECE)); Adriana R. Tome (Ncleo de Genmica e Bioinformtica, Faculdade de Veterinria, Universidade Estadual do Ceara (UECE)); Diana M. Oliveira (Ncleo de Genmica e Bioinformtica, Faculdade de Veterinria, Universidade Estadual do Ceara (UECE))

Abstract: We have performed in silico analyses of several Leishmania flagellar proteins (profilin, kinesin, katanin, coronin and the putative IFT20 and IFT88, among others). Results with the predicted actin-related and intraflagellar transport sequences lead to the assumption of a flagellar remodeling network that might link the pathogen locomotion/movement to virulence.

Poster C-34

M. tuberculosis Functional Network Analysis by Global Subcellular Protein Profiling

Kwasi Mawuenyega (Bioscience Division, Los Alamos National Laboratory); Christian Forst (Bioscience Division, Los Alamos National Laboratory); Karen Dobos (Mycobacteria Research Laboratories, Colorado State Univ.); John Belisle (Mycobacteria Research Laboratories, Colorado State Univ.); Jin Chen (Bioscience Division, Los Alamos National Laboratory); Morton Bradbury (Bioscience Division, Los Alamos National Laboratory and UC Davis); Andrew Bradbury (Bioscience Division, Los Alamos National Laboratory); Xian Chen (Bioscience Division, Los Alamos National Laboratory)

Abstract: We provide a systematic analysis of Mycobacterium tuberculosis by directly profiling its gene products, combining highthroughput proteomics and computational systems biology approaches. From predicted response networks for fatty acid degradation and lipid biosynthesis pathways we identified proteins and their subcellular locations providing novel insights into the compartmentalization of these pathways.

Poster C-35

Prediction of human protein interactions from interolog model

Tao-Wei Huang (Department of Computer Science and Information Engineering, National Taiwan University); Cheng-Yan Kao (Department of Computer Science and Information Engineering, National Taiwan University)

Abstract: We propose a relative conservation score by finding maximal quasi-clique in protein interaction networks. In validation, we reveal confidence exist between the predicted interactions and the accuracy of the prediction. Comparisons among existing methods are also indicated our method associated with better performance in prediction of human protein interactions.

Poster C-36

A study of functional protein interactions in signaling and degradation

Cheryl Wolting (Hospital for Sick Children, University of Toronto); Christopher Hogue (Blueprint Initiative Mount Sinai Hospital, University of Toronto); Jane McGlade (Hospital for Sick Children, University of Toronto)

Abstract: 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 a bioinformatics tool that provides visual summaries of proteomic results for efficient data analysis.

Poster C-37

SigPath - An information management system for quantitative modeling of cell signaling pathways and networks

Eliza Chan (Weill Cornell Medical College); Manuel Martin (Weill Cornell Medical College); Ravi Iyengar (Mount Sinai School of Medicine); Harel Weinstein (Weill Cornell Medical College); Fabien Campagne (Weill Cornell Medical College)

Abstract: We are developing the SigPath information management system to support quantitative studies of signaling pathways. This poster will present the types of data that the system helps store, view and integrate with other data sources, and an overview of current capabilities. SigPath is an open-source web-based system.

Poster C-39

SBML Model-Based Simulation with MathSBML

Bruce E. Shapiro (California Institute of Technology); Michael Hucka (California Institute of Technology); Andrew Finney (University of Hertfordshire)

Abstract: MathSBML is an open-source, freely-downloadable Mathematica package that supports examination, creation and modification of SBML models; conversion to DAEs, stoichiometry matrices, and mass-balance equations; and full simulation of models including events. It includes an API for model modification, and is fully extensible to models of any size.

Poster C-41

Network constrained clustering for gene microarray data

Dongxiao Zhu (Bioinformatics Program, University of Michigan); Alfred Hero (Depts of EECS, Biomedical Engineering and Statistics, University of Michigan); Hong Cheng (Depts of Ophthalmology, Visual Sciences and Human genetics, University of Michigan); Masayuki Akimoto (Depts of Ophthalmology, Visual Sciences and Human genetics, University of Michigan); Ritu Khanna (Depts of Ophthalmology, Visual Sciences and Human genetics, University of Michigan); Anand Swaroop (Depts of Ophthalmology, Visual Sciences and Human genetics, University of Michigan)

Abstract: We propose a new network constrained clustering approach that is able to group both similarly co-expressed genes and transitively co-expressed genes into tight clusters of interest. We compare the new clustering approach to the traditional approach on a yeast galactose metabolism dataset and a retinal gene expression dataset.

Poster C-42

Application of modular neural network in predicting protein functions

Doosung Hwang (Dankook University); Jae-Young Jung (Bioinformatics Research Team, Electronic and Telecommunication Research Institute)

Abstract: The predictive model of protein function utilizes a protein-protein interaction map based on the concept of guilt-by-association. This model can not predict the functions of proteins that do not have interactions to proteins with functions. This study considers the given problem as a K-class classification task in consideration of data coding, learning interference, class imbalance problems and proposes a predictive approach using a modular neural network. The proposed approach uses interaction data and protein related attributes as well. The experimental results with Yeast protein data show that the

proposed approach is comparable to the methodologies in KDD Cup and MIPS data.

Poster C-43 (*There will also be an oral presentation of this poster.*)

Detection of Horizontal Gene Transfer in Whole Metabolic Pathways

Tsai-Tien Tseng (Center for Biophysics & Computational Biology, University of Illinois at Urbana-Champaign); Kristen Aquino (Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana-Champaign); Lei Liu (Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana-Champaign)

Abstract: We here present analyses and partial re-annotation of ten photosynthetic bacterial genomes for studying potential horizontally transferred genes in various metabolic pathways. Pathways associated with carbon fixation pathways are targets of our study. Furthermore, our report investigates multi-gene pathway horizontally transferred in the process of evolution.

Poster C-44

A Composite Statistical Framework for Peptide Identification from Tandem Mass Spectrometry Data

Kristin Jarman (Pacific Northwest National Lab); Bobbie-Jo Webb-Robertson (Pacific Northwest National Lab); Douglas Baxter (Pacific Northwest National Lab); William Cannon (Pacific Northwest National Lab); Christopher S. Oehmen Oehmen (Pacific Northwest National Lab); Kenneth Jarman (Pacific Northwest National Lab); Alejandro Heredia-Langner (Pacific Northwest National Lab); Kenneth Auberry (Pacific Northwest National Lab); Gordon Anderson (Pacific Northwest National Lab)

Abstract: We present a flexible, multifaceted statistical framework for identification of peptides from MS/MS data that expands recent work and contributes new developments in several key areas. The method is significantly more sensitive than industry standard analyses.

Poster C-45

Pathway Simulation Tool - open software project for system biologists

Azat Badretdinov (Ariadne Genomics Inc.); Ilya Mazo (Ariadne Genomics Inc.)

Abstract: Pathway Simulation Tool reads kinetic models of pathway in SBML format and applies variety of ODE solvers to produce time course of concentrations. The novelty of the approach is that it uses analytical derivatives to calculate the Jacobean in order to perform efficient calculations. The tool is distributed under GPL.

Poster C-46

Computational Approach to Enhance the Prediction Accuracy of Protein-Protein Interactions

Jae-Young Jung (ETRI); Jae-Hun Choi (ETRI); Jong-Min Park (ETRI); Seon-Hee Park (ETRI)

Abstract: In assigning protein functions to proteins without known functions, the use of protein-protein interaction data is considered as a more reliable method in proteomics compared to gene expression correlation or phenotype data. The proposed approach is based on expression correlation and interaction generality to make prediction results more reliable.

Poster C-47

Automatic biomarker and compound identification using mass spectral data

Juergen Cox (Genedata); Marc Flesch (Genedata); Peter Haberl (Genedata); Ruediger Heidenblut (Genedata); Melanie Markmann (Genedata); Jim Samuelsson (Genedata)

Abstract: We present a set of novel methods for mass spectrometry data analysis which improve the performance of biomarker detection followed by compound identification. The general nature of these methods make them applicable to proteomics and metabolomics research. Furthermore, the methods are ideally well-suited for a streamlined and automated workflow.

Poster C-48

Conceptual Abstractions of a Protein-Protein Interaction Network

Jae-Hun Choi (Bioinformatics Research Team/Electronic and Telecommunication Research Institute(ETRI)); Jong-Min Park (Bioinformatics Research Team/Electronic and Telecommunication Research Institute(ETRI)); Jae-Young Jung (Bioinformatics Research Team/Electronic and Telecommunication Research Institute(ETRI)); Seon-Hee Park (Bioinformatics Research Team/Electronic and Telecommunication Research Institute(ETRI))

Abstract: we design and implement a concept-based method for abstracting a PIN into a composite network, which is defined as a set of relationship among composites or proteins. A composite is a vertex collapsing a sub-network, which is a sub-set of protein and relationship included in PIN.

Poster C-49

A Domain Knowledge Based Approach for Managing Bio-Object Interaction Networks

Jong-Min Park (Bioinformatics Research Team, Electronic and Telecommunication Research Institute(ETRI)); Jae-Hun Choi (Bioinformatics Research Team, Electronic and Telecommunication Research Institute(ETRI)); Jae-Young Jung (Bioinformatics Research Team, Electronic and Telecommunication Research Institute(ETRI)); Seon-Hee Park (Bioinformatics Research Team, Electronic and Telecommunication Research Institute(ETRI))

Abstract: Numerous bio-objects in a cell and complicated interactions among them can be expressed as an interaction network. Recently we propose system for managing interaction networks efficiently and systematically that can express various types of bio-objects including complex and interactions with domain knowledge such as GO, UMLS, Swiss-Prot etc.

Poster C-50

Sequence graphs for Phosphorylation motifs

Ravikumar K.E. (AU-KBC Research centre, MIT campus, Chennai); Chandrakumar A (AU-KBC Research centre, MIT campus, Chennai); Raghuram D (Chennai); Meenakshi Narayanaswamy (AU-KBC Research centre, MIT campus, Chennai)

Abstract: In this poster we describe BIONET, a tool, which incorporates various graph algorithms that, are helpful in studying biological networks. Here we describe the problem of constructing sequence graphs for phosphorylation motifs using BIONET. This study might help in clustering the phosphorylation motifs based on the kinases that phosphorylate them.

Poster C-51

Simultaneous exploration of existing knowledge and novel experimental data

Robert Hoffmann (CNB, National Center of Biotechnology); Alfonso Valencia (CNB, National Center of Biotechnology)

Abstract: Novel insights generally crystallize around existing knowledge. Analyses of microarrays or protein interaction screens for example, often implicate literature investigation on different genes or proteins. In practice this involves jumping back and forth between the experimental data and free text searches in PubMed. We show that the direct superimposition of experimental data on the biomedical literature makes a simultaneous exploration of novel and existing knowledge possible.

Poster C-52

Database driven approach for automatic construction of dynamic models of cell-wide metabolic pathways

Kazuharu Arakawa (Institute for Advanced Biosciences, Keio University); Yukino Ogawa (Institute for Advanced Biosciences, Keio University); Yoichi Nakayama (Institute for Advanced Biosciences, Keio University); Masaru Tomita (Institute for Advanced Biosciences, Keio University)

Abstract: The Genome-based E-cell Modeling System (GEM System) realizes a fully automatic conversion of genome sequence data into a quantitative in silico cell-wide metabolic pathway model. Manually curated

database of kinetic in silico models maintained for this purpose facilitates the dynamic modeling process.

Poster C-53

Protein function prediction and classification using uncertainty

James Bradford (School of Biochemistry and Microbiology, University of Leeds); Chris Needham (School of Computing, University of Leeds); Andy Bulpitt (School of Computing, University of Leeds); David Westhead (School of Biochemistry and Microbiology, University of Leeds)

Abstract: We are investigating the use of Bayesian networks to predict protein-protein interfaces and the functional effects of single nucleotide polymorphisms. We also aim to predict protein function by implementing the Gene Ontology (GO) as a Bayesian network to handle uncertain data and relate functional categories.

Poster C-54

PolyP - A Flexible Proteomics Workflow System

Matthew Sullivan (University College Dublin); Andreas De Stefani (University College Dublin); Gerard Cagney (University College Dublin)

Abstract: The Polytypic Proteomics System (PolyP) is a process and data management platform for biological analysis. The architecture is flexible and allows the creation of platform independent software pipelines and automated information management system for biological data. We demonstrate the use of PolyP to implement a proteomics workflow pipeline

Poster C-55

ArrayXPath: mapping and visualizing microarray gene-expression data with biomedical ontologies and integrated biological pathway resources using Scalable Vector Graphics

Hee-Joon Chung (SNUBI); Chan Hee Park (SNUBI); Mi-Ryung Han (SNUBI); Jihun Kim (SNUBI); Su Yeon Lee (SNUBI); Ju Han Kim (SNUBI)

Abstract: ArrayXPath (<http://www.snubi.org/software/ArrayXPath/>) is a web-based service for mapping and visualizing microarray gene-expression data with integrated biological pathway resources using Scalable Vector Graphics (SVG). Deciphering the crosstalk among pathways and integrating biomedical ontologies and knowledge bases may help biological interpretation of microarray data.

Poster C-56

Multivariate analysis of 2D DIGE data

Christian Andersson Sthlberg (GE Healthcare, Discovery Systems, Amersham Biosciences AB Uppsala); Josef Buelles (GE Healthcare, Discovery Systems, Amersham Biosciences AB Uppsala); Stephen David (GE Healthcare, Discovery Systems, Amersham Biosciences AB Uppsala); Stephanie Bourin (GE Healthcare, Discovery Systems, Amersham Biosciences AB Uppsala)

Abstract: The Ettan DIGE system has significantly increased the accuracy of quantitation achieved with 2D electrophoresis. This study demonstrates human ovarian tumor classification using DIGE data by applying multivariate methods to find putative diagnostic markers and to classify unknown samples. Multivariate analysis in Proteomics is a vital step towards personalized healthcare.

Poster C-58

AGML Central: An Infrastructure for the Analysis and Dissemination of Proteomic Data

Romesh Stanislaus (Medical University of South Carolina); Chuming Chen (Medical University of South Carolina); Jonas S Almeida (Medical University of South Carolina)

Abstract: AGML Central infrastructure is a web-based open source public infrastructure developed for the analysis and dissemination of 2-D gel electrophoresis data. AGML Central is based on the AGML data representation and provides the proteomic researcher a central location for analyzing and storing data.

Poster C-59

Extending ELM resource functionality using Grid technology

Jan Christian Bryne (Department of Informatics, University of Bergen); Christine Gemund (European Molecular Biology Laboratory, Structural and Computational Biology Unit, Heidelberg); David M.A. Martin (Post-Genomics and Molecular Interactions Centre, School of Life Sciences, University of Dundee); Rein Aasland (Department of Molecular Biology, University of Bergen, Norway); Toby J. Gibson (European Molecular Biology Laboratory, Structural and Computational Biology Unit, Heidelberg); P1 Puntervoll (Computational Biology Unit, BCCS, University of Bergen)

Abstract: ELM (Eukaryotic Linear Motifs) predicts functional sites in eukaryotic protein sequences and limits overprediction by using biological context information. By using new types of context information we expect that prediction of functional sites by ELM can be considerably improved. The additions will be implemented using Grid technology.

Poster C-60

Analysis of the in vivo effects of human erythropoietin receptor agonists on murine bone marrow populations using different clustering algorithms

Jin Lu (Departments of Molecular Discovery Technologies, Centocor, Inc (a wholly owned subsidiary of Johnson & Johnson)); Renold Capocasale (Toxicology and Investigative Pharmacology, Centocor, Inc (a wholly owned subsidiary of Johnson & Johnson)); Peter Bugelski (Toxicology and Investigative Pharmacology, Centocor, Inc (a wholly owned subsidiary of Johnson & Johnson))

Abstract: Erythropoietin mediates production of blood cells. The effects of different EPO receptor agonists are less well understood. We utilized cytometry and statistical modeling and clustering techniques to study the effects of EPO-R agonists on murine bone marrow cells in vivo. Our findings demonstrate the power of this mathematical approach.

Poster C-61

Creation of a biological-relevant simulation environment for incorporation of multi-dimensional data for prediction of patient outcomes

Victor Weigman (University of North Carolina at Chapel Hill; Department of Biology); Julie Leonard (BD Technologies); Charles Schmitt (BD Technologies); Jason Herschkowitz (University of North Carolina at Chapel Hill; Department of Genetics); Zhiyuan Hu (University of North Carolina at Chapel Hill; Lineberger Comprehensive Cancer Center); Xiaping He (University of North Carolina at Chapel Hill; Lineberger Comprehensive Cancer Center); Gamze Karaca (University of North Carolina at Chapel Hill; Lineberger Comprehensive Cancer Center); Perry Haaland (BD Technologies); Charles Perou (University of North Carolina at Chapel Hill; Department of Genetics)

Abstract: In the current schema of medical informatics, heterogeneous data have been generated through many high throughput methods. However, most of this data is analyzed separately, without consideration of the biochemical networks that govern disease. We introduce a simulation environment that incorporates heterogeneous data across these networks to predict patient outcome.

Poster C-63

Modular Decomposition of Protein Interaction Networks

Feng Luo (Department of Pathology, U.T. Southwestern Medical Center at Dallas); Yunfeng Yang (Environmental Sciences Division, Oak Ridge National Laboratory); jizhong zhou (Environmental Sciences Division, Oak Ridge National Laboratory); Richard Scheuermann (Department of Pathology, U.T. Southwestern Medical Center at Dallas)

Abstract: We present an agglomerative algorithm utilizing a new formal definition of module to decompose yeast protein interaction network into modules. The modules obtained have been found to be biological meaningful. Further, the higher-level interactions between modules allow a system-level understanding of the relationship between different biological processes.

Poster C-64 (*There will also be an oral presentation of this poster.*)

A Probabilistic Functional Gene Network of Yeast - Version 2.0

In Suk Lee (Center for Systems and Synthetic Biology, University of Texas, Austin); Edward Marcotte (Center for Systems and Synthetic Biology, University of Texas, Austin)

Abstract: We propose a conceptual framework for reconstructing cellular systems via integrating heterogeneous functional genomics data: Probabilistic FUNctional GENE NETwork (Pfungene). Within this framework, we reconstructed a functional gene network for *S. cerevisiae* with 5,064 genes linked by 57,000 probabilistic linkages comparable in accuracy to small-scale protein interaction assays.

Poster C-65

An Eigenface based approach to classify cell-signaling profiles

Nikesh Kotecha (Stanford University); Jonathan Irish (Stanford University); Garry Nolan (Stanford University)

Abstract: Analyses of alterations in cell signaling profiles, generated by intracellular flow cytometry, provide critical insights into human diseases. However, this approach is manual, subjective and difficult to scale. We propose a scalable eigenface based approach to classify cell signaling profiles and identify cells responsible for adverse clinical outcome.

Poster C-66

Identifying the functional constraints that shape the quantitative design of a metabolic circuit

Armando Salvador (Chemistry Department / Faculty of Science and Technology of the University of Coimbra); Michael Savageau (Department of Biomedical Engineering / The University of California at Davis)

Abstract: A principled reengineering of biochemical systems requires an appreciation of the functional constraints that shape the values of biochemical parameters. Applying principles and techniques from Biochemical Systems Theory we identify the functional constraints that have shaped the quantitative design of the coupled redox cycles of NADPH and glutathione in human erythrocytes.

Poster C-67

Rapid Identification of Biological Pathway Regulation Using Gene Expression Data and a Novel Algorithm Based on Mahalanobis Distance.

Kory Johnson (Gene Logic, Inc.); Lawrence Mertz (Gene Logic Inc.)

Abstract: To enhance the detection of pathway-centric gene expression responses to perturbations in biological state, we developed a multivariate based algorithm, termed Pathway Prioritizer, which ranks biological pathways with respect to the significance of gene expression changes for pathway members. We applied this algorithm to a human ovarian cancer study.

Poster C-68

Biological Modularity in Yeast Protein-Protein Interaction Network

Jingchun Chen (Integrated Biomedical Sciences Graduate Program, The Ohio State University, Columbus, OH); Zoe Gokhale (Biophysics Program, The Ohio State University, Columbus, OH); Daolong Wang (Washington University, St. Louis, MO); Fa Zhang (Biomedical Informatics, The Ohio State University, Columbus, OH); Junshui Ma (Ohio Supercomputing Center, Columbus, OH); Bo Yuan (Biomedical Informatics, Pharmacology, The Ohio State University, Columbus, OH)

Abstract: By analyzing a reliable yeast protein-protein interaction dataset, we found that 1) Clique is the predominant frequent subgraph in the proteome network; 2) Yeast phenotype modularity exists at the level of network motif, protein complex and biological process; and 3) Modularity helps us understanding the complex genotype-phenotype relationship.

Poster C-69 (*There will also be an oral presentation of this poster.*)

New machine learning approaches for classification of mass spectrometry database search results

Peter Ulintz (Bioinformatics Program, University of Michigan); Ji Zhu (Department of Statistics, University of Michigan); P.C. Andrews (Department of Biological Chemistry, University of Michigan); Steve Qin (Department of Biostatistics, University of Michigan)

Abstract: This research demonstrates the effectiveness of current machine learning approaches in classifying the results of mass spectrometry database search algorithms such as SEQUEST as either 'correct' or 'incorrect' with a well-defined measure of confidence.

Poster C-70

Nebulon-NetView: a web tool for visualization and analysis of genomic networks of functional interactions

Edgar Diaz-Peredo (Centre for Genomic Sciences , UNAM); Fabiola Snchez-Solano (Centre for Genomic Sciences , UNAM); Sarath Chandra Janga (Centre for Genomic Sciences , UNAM); Julio Collado-vides (Centre for Genomic Sciences , UNAM); Gabriel Moreno-Hagelsieb (Wilfrid Laurier University)

Abstract: We present Nebulon-NetView, a Java-Enterprise-Application based on J2EE-Patterns and the graphLayout API of Sun-Microsystems. The application integrates several kinds of genomic and functional data like information about genes, their GOs, COGs, confidence in the linkage obtained from operon predictions etc to allow the user to explore and functionally analyse complex datasets. The dynamic graphs generated show a genomic network of functional associations inferred from gene fusions, distance-based operon predictions across genomes for a query gene in a chosen genome. The nodes represent genes and the edges represent the inferred functional interactions. It can be accessed from <http://tikal.cifn.unam.mx/nebulon>.

Poster C-71

Learning Classifiers for Assigning Protein Sequences to Gene Ontology Functional Families

Carson Andorf (Iowa State University); Adrian Silvescu (Iowa State University); Vasant Honavar (Iowa State University); Drena Dobbs (Iowa State University)

Abstract: We explore several machine learning approaches to data-driven construction of classifiers for assigning protein sequences to appropriate Gene Ontology (GO) function families using a class conditional probabilistic representation of amino acid sequences. Our methods are able to accurately predict GO labels and determine regions known to be associated with function.

Poster C-72 (*There will also be an oral presentation of this poster.*)

Conserved protein domain interactions from structure data: A modeling resource for novel protein interactions

Ben Shoemaker (NCBI/NLM/NIH); Stephen Bryant (NCBI/NLM/NIH)

Abstract: Conserved protein-protein interactions were surveyed from the structure database. A conserved interaction is inferred when different members of interacting domain families dock in the same way, such that the structural complexes superimpose. These conserved interaction modes, fewer than previously reported, generate a library of docking templates for molecular modeling.

Poster C-74

Mathematical modeling of residual variation for differential 2d gel electrophoresis proteomics

Jonas Almeida (Dept Biometry Bioinformatics and Epidemiology, Medical University of South Carolina); Romesh Stanislaus (Dept Biometry Bioinformatics and Epidemiology, Medical University of South Carolina); Ed Krug (Dept Cell Biology & Anatomy, Medical University of South Carolina); John Arthur (Div. Nephrology, Dept. of Medicine, Medical University of South Carolina)

Abstract: Some high throughput methods, notably in proteomics, have not benefited from the same degree of modeling of variance that transcriptomics presently enjoys. We have accordingly targeted 2D gel electrophoresis to model biological and methodological variability and found that multimodal density distributions are to be expected that require novel approach.

Poster C-75

Learning Causal Relationship between Genes with Feedback Loops from Steady-State Observations

Xin Zhang (Arizona State University); Chitta Baral (Arizona State University); Seungchan Kim (Arizona State University and Translational Genomics Research Institute)

Abstract: IC algorithm can only infer the causal relationship of the directed acyclic graph. However, in the biological system the molecular components may have feedback loops. We propose an algorithm based on IC algorithm that leaves one gene out at a time to infer the cyclic causal relationships among genes.

Poster C-76

Proofreading Reactome: A case study in pathway knowledgebase verification

Stephen Racunas (Huck Institutes of Life Sciences); Nigam Shah (Huck Institutes of Life Sciences); Nina Fedoroff (Huck Institutes of Life Sciences)

Abstract: We develop methods for "proofreading" pathway knowledgebases and demonstrate them on the Reactome knowledgebase. We make explicit the formal language implicit in Reactome and specify a logic under which to test properties such as consistency and directness. We assay Reactome's expressiveness and potential for supporting inference tools like HyBrow.

Poster C-77

Combining Bayesian Networks and Decision Trees to Predict Drosophila melanogaster Protein-Protein Interactions

Jingkai Yu (Department of Computer Science, Wayne State University); Farshad Fotouhi (Department of Computer Science, Wayne State University); Russell Finley, Jr. (Wayne State University School of Medicine)

Abstract: We set out to predict Drosophila protein-protein interactions using existing experimental data combined with Gene Ontology annotations of proteins. We show that GO annotations can be a useful predictor and prediction performance can be improved by combining results from both decision trees and Bayesian networks in a simple way.

Poster C-78

Recognize Interaction Domains with Support Vector Machines

Ya Zhang (Penn State University)

Abstract: The study of interaction domains is an essential part of functional genomics. However, automatically determining interaction domains is challenging and only a few well-studied interaction domains are identified. We employ Support Vector Machines to identify interaction domains in proteins. Domain sequences, structures, lengths, frequency, and function diversity are used as input for the classifier.

Poster C-80

Java-based software framework for comparison of quantitative proteomics and DNA microarray data

Sergii Ivakhno (Taras Shevchenko Kyiv National University); Alexander Kornelyuk (Institute of Molecular Biology and Genetics)

Abstract: We have developed Java-based software framework for analysis and comparison of mass-spectrometry-based quantitative proteomics and DNA microarray data. Not only reliability of data from separate experiments can be verified, but potentially new pathways of gene expression can be explored by observing correlation profiles between mRNA/protein abundances.

Poster C-81

Nebulon: a system for the inference of functional relationships of gene products from the rearrangement of predicted operons

Sarath Chandra Janga (Centre for Genomic Sciences, UNAM); Julio Collado-Vides (Centre for Genomic

Sciences , UNAM); Gabriel Moreno-Hagelsieb (Wilfrid Laurier University)

Abstract: We present nebulon, a system to build networks of functional relationships of gene products based on their organization into operons in any available genome. Our system can use different kinds of thresholds to accept a functional relationship, either related to the prediction of operons, or to the number of non-redundant genomes that support the associations. The method shows high reliability benchmarked against knowledge-bases of functional interactions. We also illustrate the use of Nebulon in finding new members of regulons, and of other functional groups of genes.

Poster C-82

Phylogeny of Tumor Progression

Sven Bilke (National Cancer Institute); Qingrong Chen (National Cancer Institute); Craig Whiteford (National Cancer Institute); Javed Khan (National Cancer Institute)

Abstract: The distribution of chromosomal locations of genomic imbalances correlates with the stage of tumors. Progression of the disease leaves a characteristic signature. We describe a method to map cytogenetic data to tumor progression models derived from characteristic imbalance pattern and infer a model for the progression of neuroblastoma.

Poster C-83

Co-evolutionary Analysis of Metabolism

Natalia Maltsev (Argonne National Laboratory); Galina Ovchinnikova (Argonne National Laboratory); Elizabeth Glass (Argonne National Laboratory); Alex Rodriguez (Argonne National Laboratory); Tanuja Bompada (Argonne National Laboratory); Mark D'Souza (Argonne National Laboratory); Dinanath Sulakhe (Argonne National Laboratory); Yi Zhang (Argonne National Laboratory)

Abstract: We will describe an approach for analyzing co-evolutionary changes of metabolic pathways, genomic organization, and enzymes characteristic of taxonomic and phenotypic groups of organisms. Comparative analysis of central metabolism in four Cyanobacteria will be presented as an example of our approach.

Poster C-84

Networks of coexpressed genes based on multiple microarray experiments

Wieslawa Mentzen (Iowa State University); Nick Ransom (Iowa State University); Dianne Cook (Iowa State University); Basil Nikolau (Iowa State University); Eve Wurtele (Iowa State University)

Abstract: We use combined transcriptomics data from 1000 Arabidopsis Affymetrix chips to establish pairs of genes with correlated expression profiles across all experiments. Based on the example of an interaction network revealed by inspecting correlations among genes from acetyl-CoA biotin pathways, we show that such correlations are in fact meaningful and informative.

Poster C-85

Assigning confidence scores to protein function predictions based on protein network topology

Guozhen Liu (Center for Molecular Medicine and Genetics, Wayne State University School of Medicine); Stephen Guest (Center for Molecular Medicine and Genetics, Wayne State University School of Medicine); Russ Finley (Center for Molecular Medicine and Genetics, Wayne State University School of Medicine)

Abstract: A new algorithm to predict protein functions and assign confidence scores based on the topology of nodes as well as edges around the protein in protein-protein interaction networks.

Poster C-86

The conservation and evolutionary modularity of metabolic and regulatory pathways

Jose M. Peregrin-Alvarez (Hospital for Sick Children); John Parkinson (Hospital for Sick Children)

Abstract: The availability of partial genome datasets greatly complements the existence of full genome sequences. Here, we integrate these data in order to explore the conservation of metabolic and regulatory

pathways. We also present our findings on the evolutionary modularity of these pathways based on similarity between phylogenetic profiles.

Poster C-87

Modelling genetic network modules through CTRNN

David Camacho-Trujillo (iBIOS/German Cancer Research Center); Axel Szabowski (Signal Transduction & Growth Control/German Cancer Research Center); Roland Eils (iBIOS/German Cancer Research Center)

Abstract: The AP-1-module was modelled by continuous time recurrent neuronal network (CTRNN). Our modelling approach could reproduce the dynamics of this system, also it gives some insights on its functioning through a matrix of weights between genes. Moreover, it is able to predict certain behaviour of AP-1 module.

Poster C-88

Cell++ - simulating biochemical pathways within a spatial context

Matthew Yip (Hospital for Sick Children); Chris Sanford (Hospital for Sick Children); Carl White (Hospital for Sick Children); John Parkinson (Hospital for Sick Children)

Abstract: Cell++ is a novel 3D 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.

Poster C-89

GenMAPP 2.0 for Pathway Analysis of Genomic Data: Current and Future Development

Alexander Pico (Gladstone Institute of Cardiovascular Disease, UCSF); Nathan Salomonis (Gladstone Institute of Cardiovascular Disease, UCSF); Kristina Hanspers (Gladstone Institute of Cardiovascular Disease, UCSF); Karen Vranizan (University of California, Berkeley, Functional Genomics Lab); Alexander Zambon (Gladstone Institute of Cardiovascular Disease, UCSF); Steven Lawlor (Gladstone Institute of Cardiovascular Disease, UCSF); Scott Doniger (Washington University School of Medicine); Kam Dahlquist (Vassar College, Department of Biology); Bruce Conklin (Gladstone Institute of Cardiovascular Disease, UCSF); Lynn Ferrante; Jeff Lawlor

Abstract: GenMAPP is a free program designed for viewing and analyzing genomic data in the context of biological pathways. The current release of GenMAPP 2.0 integrates the analysis of pathway data with curated or user-defined databases of gene annotations. BioPAX-compliant interactions will be implemented in the future development.

Poster C-90

Breaking the Power Law: Improved Model Selection Suggests that Many Biological Networks Have Multiple Hidden Edge Types

Debra S. Goldberg (Harvard Medical School); Giovanni Franklin (Harvard Medical School); Frederick P. Roth (Harvard Medical School)

Abstract: For 55 biological networks (six different data types), we systematically evaluated degree distribution models, including models allowing two underlying link types, e.g., true and false positive edges. We find many networks previously reported to be scale-free or follow another simple model show a dramatically better fit to more complex models.

Poster C-91

BISIP: Biased iterative searching for the identification of peptides in noisy Mass Spectrometric data

Sunil Wagh (University of Missouri Kansas City School of Computing and Engineering); Deendayal Dinakarpanthian (University of Missouri Kansas City School of Computing and Engineering); William Morgan (University of Missouri Kansas City School of Biological Sciences)

Abstract: Despite impressive advances in the sensitivity of Mass Spectrometric instruments, the signal to

noise ratio of some samples can be too low to allow confident identification of the protein in the sample. We propose an iterative search algorithm that increases the probability of successful identification when the spectrum is noisy.

Poster C-92 (*There will also be an oral presentation of this poster.*)

An EM algorithm for unambiguous assignment of genes to biochemical pathways

Liviu Popescu (Cornell University); Golan Yona (Cornell University)

Abstract: We developed an EM algorithm to compute assignment probabilities of genes to cellular pathways using database annotations, statistical models of enzyme families and expression data. This solves the ambiguity introduced by current systems which classify all genes from a specific enzyme family to all the pathways containing the corresponding reaction.

Poster C-93

Robust Subcellular Location Pattern Recognition among Different Cell Types

Xiang Chen (Carnegie Mellon University); Robert Murphy (Carnegie Mellon University)

Abstract: We have previously designed numerical features to quantitatively describe protein subcellular location patterns in high resolution fluorescence microscopy images. In the current study, we extended our approaches to a heterogeneous dataset and the results suggest that under proper feature transformations, similar patterns from different sources were grouped together.

Poster C-94

Discovering k-minimal shortest physical interaction pathways in biological networks

Meeyoung Park (University of Missouri - Kansas City); Jubin Sanghvi (University of Missouri - Kansas City); Deendayal Dinakarpandian (University of Missouri - Kansas City)

Abstract: We propose a novel algorithmic approach for the discovery of k-minimal shortest physical interaction pathways in biological interaction networks, taking biological constraints into consideration. We have successfully evaluated our algorithm on information from the Database of Interacting Proteins (DIP) and the Biomolecular Interaction Network Database (BIND).

Poster C-95

Classification of Secreted Proteins Using a Single Domain

Carlos P. Sosa (IBM and University of Minnesota); Eric W. Klee (University of Minnesota); Stephen C. Ekker (University of Minnesota); Lynda B. M. Ellis (University of Minnesota)

Abstract: Secreted proteins (the secretome) make up approximately 10-20% of the vertebrate proteome, control cell-cell interactions, and are major targets for drug discovery. We developed csP, a technique that uses protein domain classification instead of signal sequence identification to predict secreted proteins.

Poster C-96

Analyzing duplex iTRAQ experiments to detect differentially expressed proteins

Catherine Grasso (University of Michigan); George Michailidis (University of Michigan); Phil Andrews (University of Michigan)

Abstract: We develop a statistical framework for detecting differentially expressed proteins in duplex experiments using iTRAQ reagents quantified through MS/MS spectrometry. The methodology incorporates the information obtained from the MS/MS fragmentation process, and is illustrated on data from human lung adenocarcinoma A549 cell-lines subjected to TGF-Beta treatment.

Poster C-97

Prediction of regulatory binding sites with high accuracy through comparative genomics

Zhengchang Su (Bioinformatics Institute, Department of Biochemistry and Molecular Biology, University of Georgia, and 2Computational Biology Institute, Oak Ridge National Laboratories); victor Olman (Bioinformatics Institute, Department of Biochemistry and Molecular Biology, University of Georgia);

Fenglou Mao (Bioinformatics Institute, Department of Biochemistry and Molecular Biology, University of Georgia); Ying Xu (Bioinformatics Institute, Department of Biochemistry and Molecular Biology, University of Georgia, and 2Computational Biology Institute, Oak Ridge National Laboratories)

Abstract: We have improved the conventional phylogenetic footprinting procedure by looking for multiple signature sequences in the promoter regions as well as similar information for the orthologues in other genomes. We also have developed a probabilistic model to fish out the true binding sites with high accuracy from the noisy background.

Poster C-98

Iterative Weighting of Phylogenetic Profiles Increases Classification Accuracy

Roger Craig (University of Delaware); Li Liao (University of Delaware)

Abstract: Phylogenetic profiles of proteins, encoding the presence and absence of proteins in genomes, have been utilized to predict functionally-linked proteins. We present an iterative weighting scheme for extending phylogenetic profiles to incorporate evolutionary relations represented in phylogenetic tree. Consequently, the accuracy of a support vector machine classifier was increased significantly.

Poster C-99

An information integration system for automated reconstruction and dynamical modeling of gene regulatory networks.

Michael Baitaluk (San Diego Supercomputer Center, University of California San Diego); Xufei Qian (San Diego Supercomputer Center, University of California San Diego); Shubhada Godbole (Molecular Computing, Keck Graduate Institute); Vijay Chickarmane (Molecular Computing, Keck Graduate Institute); Amarnath Gupta (San Diego Supercomputer Center, University of California San Diego); Animesh Ray (Molecular Computing, Keck Graduate Institute)

Abstract: We report the development of Biological Networks Server that provides querying services (querying language plus querying engine) and dynamical modeling framework over PathSys - database system, integrating over 14 curated and public data sources for the *Saccharomyces cerevisiae*, containing molecular and genetic interactions, localization, and microarray data available through published literature.

Poster C-100

Multi-algorithm prediction of reverse phase chromatography elution time and its use in the validation of shotgun mass spectrometry protein identification

Steve Russell (University of Colorado Health Sciences Center); Katheryn Resing (University of Colorado-Boulder); Larry Hunter (University of Colorado Health Sciences Center)

Abstract: Integrated multilinear regression and neural network results reduced the peptide identification error rate from 4% to 2% by capturing unexpected and useful details in LC-MS/MS data. Amino acid hydrophobicity was the basis for deriving coefficients and weights for Reverse Phase chromatography, using spectrometry data from complex biological samples.

Poster C-101

PROTEOME ISOFORM ANALYZER (PIA) - analyzing proteomics data for protein isoforms generated by alternative splicing

Natasha Levenkova (Wistar); Hsin Yao Tang (Wistar); Nadeem Ali-Khan (Wistar); Won-A Joo (Wistar); David Speicher (Wistar); John Rux (Wistar)

Abstract: We describe a tool - Proteome Isoform Analyzer (PIA) - for identifying alternative splice variants from tryptic peptides generated by proteomic experiments. We show how PIA can be used for global proteomic analyses of human tumor-associated proteins to explore the possibility that alternatively spliced variants might serve as potential cancer biomarkers.

Poster C-102 (*There will also be an oral presentation of this poster.*)

BioPAX - Biological Pathway Data Exchange Format

Michael Cary (Memorial Sloan-Kettering Cancer Center); BioPAX Workgroup N/A (BioPAX.org); Gary Bader (Memorial Sloan-Kettering Cancer Center); Chris Sander (Memorial Sloan-Kettering Cancer Center); Erik Brauner (Memorial Sloan-Kettering Cancer Center); Robert Goldberg (Memorial Sloan-Kettering Cancer Center); Chris Hogue (Memorial Sloan-Kettering Cancer Center); Peter Karp (Memorial Sloan-Kettering Cancer Center); Joanne Luciano (Memorial Sloan-Kettering Cancer Center); Debbie Marks (Memorial Sloan-Kettering Cancer Center); Natalia Maltsev (Memorial Sloan-Kettering Cancer Center); Eric Neumann (Memorial Sloan-Kettering Cancer Center); Suzanne Paley (Memorial Sloan-Kettering Cancer Center); John Pick (Memorial Sloan-Kettering Cancer Center); Aviv Regev (Memorial Sloan-Kettering Cancer Center); Andrey Rzhetsky (Memorial Sloan-Kettering Cancer Center); Chris Sander (Memorial Sloan-Kettering Cancer Center); Vincent Schachter (Memorial Sloan-Kettering Cancer Center); Imran Shah (Memorial Sloan-Kettering Cancer Center); Jeremy Zucker (Memorial Sloan-Kettering Cancer Center); Mirit Aladjem (Memorial Sloan-Kettering Cancer Center); Gary D. Bader (Memorial Sloan-Kettering Cancer Center); Michael P. Cary (Memorial Sloan-Kettering Cancer Center); Kam Dahlquist (Memorial Sloan-Kettering Cancer Center); Emek Demir (Memorial Sloan-Kettering Cancer Center); Peter D'Eustachio (Memorial Sloan-Kettering Cancer Center); Ken Fukuda (Memorial Sloan-Kettering Cancer Center); Frank Gibbons (Memorial Sloan-Kettering Cancer Center); Marc Gillespie (Memorial Sloan-Kettering Cancer Center); Chris Hogue (Memorial Sloan-Kettering Cancer Center); Michael Hucka (Memorial Sloan-Kettering Cancer Center); Geeta Joshi-Tope (Memorial Sloan-Kettering Cancer Center); David Kane (Memorial Sloan-Kettering Cancer Center); Peter Karp (Memorial Sloan-Kettering Cancer Center); Christian Lemer (Memorial Sloan-Kettering Cancer Center); Joanne Luciano (Memorial Sloan-Kettering Cancer Center); Natalia Maltsev (Memorial Sloan-Kettering Cancer Center); Eric Neumann (Memorial Sloan-Kettering Cancer Center); Suzanne Paley (Memorial Sloan-Kettering Cancer Center); Elgar Pichler (Memorial Sloan-Kettering Cancer Center); Jonathan Rees (Memorial Sloan-Kettering Cancer Center); Andrey Rzhetsky (Memorial Sloan-Kettering Cancer Center); Vincent Schachter (Memorial Sloan-Kettering Cancer Center); Andrea Splendiani (Memorial Sloan-Kettering Cancer Center); Mustafa Syed (Memorial Sloan-Kettering Cancer Center); Edgar Wingender (Memorial Sloan-Kettering Cancer Center); Guanming Wu (Memorial Sloan-Kettering Cancer Center); Jeremy Zucker (Memorial Sloan-Kettering Cancer Center)

Abstract: BioPAX (biopax.org) is a data exchange format for biological pathways developed by pathway databases, such as BioCyc, WIT, KEGG, aMAZE, Reactome, BIND and others. Level 1 supports metabolic pathways. Level 2 adds support for molecular interactions. Level 3 will add additional support for signal transduction and genetic regulatory networks.

Poster C-103

Fly-DPI: Database of Protein Interactomes for *Drosophila melanogaster*

Chung-Yen Lin (National Health Research Institutes, Taiwan); Chia-Ling Chen (National Health Research Institutes, Taiwan); Fan-Kai Lin (National Health Research Institutes, Taiwan); Chi-Shiang Cho (National Health Research Institutes, Taiwan); Chia-Ming Chang (National Health Research Institutes, Taiwan); Pao-Yang Chen (National Health Research Institutes, Taiwan); Chieh-Hua Lin (National Health Research Institutes, Taiwan); Shu-Hwa Chen (Academia Sinica); Chen-Zen Lo (National Health Research Institutes, Taiwan); Chao A. Hsiung (National Health Research Institutes, Taiwan)

Abstract: In the era of proteomics, protein networks play a major role in expanding human knowledge towards biology. We choose *Drosophila melanogaster* to build "fly-DPI", which is aiming at providing succinct protein network maps from disorderly protein interactions, by means of integrating statistic model and biologically important information.

Poster C-104

Using error-correcting output coding for predicting multi-class subcellular localization

Mark Doderer (University of Texas at San Antonio); Stephen Kwek (University of Texas at San Antonio); John Salinas (University of Texas at San Antonio); Kihoon Yoon (University of Texas at San Antonio)

Abstract: A number of techniques that exist for predicting protein destination are limited by predicting a

small set of possible locations because performance of prediction degrades by trying to expand the learning model. We propose using error-correcting output coding for producing scalable learning models to better fit real world scenarios.

Poster C-105

MassPective: a graphical tool for peptide and post-translational modification identification from tandem mass spectra

SeungJin Na (Dept. of Mechanical and Information Engineering, University of Seoul); Santae Kim (Dept. of Computer Science, Korea Military Academy); Heejin Park (College of Information and Communications, Hanyang University); Eunok Paek (Dept. of Mechanical and Information Engineering, University of Seoul)

Abstract: Identifying post-translational modifications (PTM) in a peptide is essential to understand biological functions of a protein, but has remained a computationally difficult problem. While MODi provides a list of candidate PTM interpretations, MassPective facilitates manual inspection of MODi interpretation and allows manual sequencing augmented to the MODi interpretation as well.

Poster C-106

Diagnosing Prion Disease Using Proteomic Mass Spec Data

Sean McIlwain (Department of Computer Sciences, Department of Biostatistics and Medical Informatics, University of Wisconsin - Madison); Allen Herbst (Department of Animal Health and Biomedical Sciences, University of Wisconsin - Madison); Joshua Schmidt (Department of Pharmacy, University of Wisconsin - Madison); David Page (Department of Computer Sciences, Department of Biostatistics and Medical Informatics, University of Wisconsin - Madison); Lingjun Li (Department of Pharmacy, University of Wisconsin - Madison); Judd Aiken (Department of Animal Health and Biomedical Sciences, University of Wisconsin - Madison)

Abstract: Transmissible Spongiform Encephalopathies, called Prion diseases, are difficult to diagnose pre-mortem. We explore a method capable of detecting Prion disease in live animals by analyzing the proteomic content of cerebrospinal fluid measured through mass spectrometry. We present our data analysis results for a number of classification algorithms.

Poster C-107

Structure Refinement in Pathway Modeling

Irene Ong (University of Wisconsin-Madison); Daniel McFarlin (University of Wisconsin-Madison); David Page (University of Wisconsin-Madison)

Abstract: We explore an approach that has the capability of refining deterministic models by making use of extensive prior knowledge and various data types. This will enable the refinement of players or species in the reactions and potentially reveal the underlying mechanism for a particular pathway.

Poster C-108

INOH pathway database

Ken Ichiro Fukuda (CBRC, AIST); Takao Asanuma (BIRD, JST); Satoko Yamamoto (BIRD, JST); Tatsuya Kushida (BIRD, JST); Emi Hattori (Information and Mathematical Science Laboratory, Inc.); Yuki Yamagata (BIRD, JST); Toshihisa Takagi (Graduate School of Frontier Sciences, University of Tokyo)

Abstract: INOH is a human curated signal transduction pathway database of model organisms including human, mouse, rat and others. Every pathway component in INOH is annotated by a biological ontology and these ontological knowledge can be utilized during a pathway query. The system are freely available at <http://www.inoh.org>.

Poster C-109

A Recursive Maximum Contrast Tree Approach to Learning Protein Functional Classes

Mary Yang (Purdue University); Jack Yang (Indiana University School of Medicine)

Abstract: The problem of classifying proteins of unknown function into functional classes is complicated

by the fact that certain proteins participate in multiple biological pathways, and thus belong to more than one functional class. We developed a new Multiply Labeled Instance Classifier (MLIC) that can handle this type of data.

Poster C-110

Analysis for the movement of an enzyme.

Noriko Hiroi (JST/ERATO-SORST Kitano Symbiotic Systems Project); Akira Funahashi (JST/ERATO-SORST Kitano Symbiotic Systems Project); Hiroaki Kitano (JST/ERATO-SORST Kitano Symbiotic Systems Project)

Abstract: To study how to express the biochemical reactions in the condensed environment like inner cellular spaces, we analyzed the EcoRV diffusion process on DNA, which is believed as a highly correlated process, by numerical models and biochemical experiments. Our analyses suggested the most major diffusion process of enzyme is 'hopping'.

Poster C-111

Identification of GPI modification signal peptides and prediction of their cleavage sites.

Yu Zhang (Center for Biological Sequence Analysis, BioCentrum, Technical University of Denmark); Thomas Skjelt Jensen (Center for Biological Sequence Analysis, BioCentrum, Technical University of Denmark); Ulrik de Lichtenberg (Center for Biological Sequence Analysis, BioCentrum, Technical University of Denmark); Soren Brunak (Center for Biological Sequence Analysis, BioCentrum, Technical University of Denmark)

Abstract: We have developed a new sequence-based prediction tool of Glycosylphosphatidylinositol (GPI)-anchored proteins based on neural networks trained on experimentally verified data extracted from Swiss-Prot. The method predicts GPI-anchored proteins with high accuracy and performs better than all other prediction schemes currently available.

Poster C-112

Prediction of regulatory network among gene clusters

Hye Young Kim (Hanyang University); Jin Il Han (Hanyang University); Yong Sung Lee (Hanyang University); Young Seek Lee (Hanyang University); Jin Hyuk Kim (Hanyang University)

Abstract: We have implemented the algorithm, "refining and enriching clusters" to obtain overlapping clusters divided into adjustable numbers of clusters. This algorithm was applied to 10K cDNA microarray data obtained from mouse embryonic stem cells after induction of neural differentiation. We obtained 184 overlapping clusters and found edges among gene clusters.

Poster C-113

Suggestion of method for reconstruction of genetic regulatory network with genetic programming

Min Jung Kim (Hanyang University); Bo Kyung Kim (Hanyang University); Yong Sung Lee (Hanyang University); Young Seek Lee (Hanyang University); Jin Hyuk Kim (Hanyang University)

Abstract: We tried to reconstruct genetic regulatory network with high-throughput gene expression data acquired from differentiation of mouse embryonic stem cells to neurons. From a gene module based on correlation strength between reference genes and clusters, we deduced a genetic regulatory network by iterating evolutionary generation with genetic programming.

Poster C-114

Inference of transcriptional factor binding site with genetic algorithm (MEGA: Motif Elicitation with Genetic Algorithm)

Jin Hyuk Kim (Hanyang University); Min Jung Kim (Hanyang University); Hye Young Kim (Hanyang University); Yong Sung Lee (Hanyang University); Young Seek Lee (Hanyang University)

Abstract: Motif Elicitation with Genetic Algorithm (MEGA) was devised for the purpose of discovering

motifs, revising randomly generated motifs with genetic algorithm. MEGA successfully executes to find motifs in 5kbp upstream sequences of several tens of genes belonging to a cluster, with high reproducibility and small consumption of computation resources.

Poster C-115

aMAZE: a multi-layer data model for biological processes.

Christian LEMER (Service de Conformation des Macromolecules Biologiques et de Bioinformatique - ULB); Fabian COUCHE (Service de Conformation des Macromolecules Biologiques et de Bioinformatique - ULB); Simon DE KEYZER (Service de Conformation des Macromolecules Biologiques et de Bioinformatique - ULB); Yves DEVILLE (Department of Computing Science and Engineering - UCL); Frederic FAYS (Service de Conformation des Macromolecules Biologiques et de Bioinformatique - ULB); Olivier HUBAUT (Service de Conformation des Macromolecules Biologiques et de Bioinformatique - ULB); Olivier SAND (Service de Conformation des Macromolecules Biologiques et de Bioinformatique - ULB); Jacques VAN HELDEN (Service de Conformation des Macromolecules Biologiques et de Bioinformatique - ULB); Shoshana WODAK (Centre for Computational Biology - The Hospital for Sick Children - Toronto); Jean RICHELLE (Service de Conformation des Macromolecules Biologiques et de Bioinformatique - ULB)

Abstract: The aMAZE data model features three main layers: The biochemical layer, describes elementary building blocks - metabolites, polypeptides, genes, reactions and expressions. The systemic layer describes the networks of gene regulation, metabolic and signaling events as an integrated graph.. Finally, the functional layer describes pathways of functionally related biological events.

Poster C-117

Selection of Input Node from Probabilistic Boolean Network

Ha Seong Kim (Seoul National University); Sung-Gon Yi (Seoul National University); Taesung Park (Seoul National University)

Abstract: Probabilistic Boolean network(PBN) is a simple network model with binary status. For time sequence microarray data, PBN requires heavy computing times to select k input nodes which influences a specific node at time t+1. We propose a statistical procedure to identify input nodes with much less computing times.

4 Sequence Analysis, Phylogeny and Evolution

Poster D-1

WinPop 2.5: software for representing Population Genetics phenomena

Paulo Augusto Suano Nuin (Dept of Biology, McMaster University)

Abstract: WinPop is a user-friendly software for population genetics courses and basic research, providing a visual tool that allows simulation and representation of population genetics phenomena. WinPop contains six different modules that represent and simulate population genetics models: panmixia, drift, assortative matings, selection, gene flow, and mutation.

Poster D-2

Using Permuted Variable Length Markov Model (PVLMM) to characterize motifs in syntenic regions of chloroplast genomes

Beatrice Kilel (George Mason University)

Abstract: The conserved nature in chloroplast genomes has been found to be useful in establishing gene function and possibly inferring genome evolution. Some of the important genes in these genomes are involved in several processes. Ribosomal proteins rpl14 and rps8 are used to provide a simple interpretation of the prevailing nucleic acid binding sites and PVLMM is used to identify small DNA motifs.

Poster D-3 (*There will also be an oral presentation of this poster.*)

Combining Phylogenetic Data and Network Topology to Identify Regulatory Motifs

Ting Wang (Washington University); Gary Stormo (Washington University)

Abstract: Predicting regulatory motifs from whole genome sequences remains a challenging problem due to the statistical limitations of conventional motif finding algorithms. We introduce a new algorithm "PhyloNet" that combines phylogenetic information and network topology to define conserved motifs and co-regulated promoters and to build a regulatory network *ab initio*.

Poster D-4 (*There will also be an oral presentation of this poster.*)

Identification of novel human splice variants using "genomic fossils"

Ronen Shemesh (Compugen Ltd); Amit Novik (Compugen Ltd); Sarit Edelheit (Compugen Ltd); Rotem Sorek (Tel Aviv University and Compugen)

Abstract: We present a novel method to discover new splice variants in the human genome using processed pseudogenes. We detected hundreds of new transcript variants so far unidentified. An experimental verification of a subset of these variants indicates that most of them are still active transcripts in the human transcriptome.

Poster D-5

Mapping Recombination Hot-Spots onto Gaussian Markov Random Fields

Vladimir N. Minin (Department of Biomathematics, University of California, Los Angeles); Karin S. Dorman (Departments of Statistics and Genetics, Cell & Development Biology and the Program in Bioinformatics and Computational Biology, Iowa State University); Marc A. Suchard (Department of Biomathematics, University of California, Los Angeles)

Abstract: We propose to harvest and combine results of individual recombination analyses to elucidate patterns of recombination occurrences along the HIV genome. To address the sparseness of available information on recombination break-point locations we recruit Gaussian Markov random field priors in a Bayesian hierarchical approach.

Poster D-6

Improving the Sensitivity of Multiple-Sequence Alignments by Incorporating Prior Knowledge

Sumedha Gunewardena (University of Toronto, Charles H Best Institute)

Abstract: We present efficient modifications to the well-established progressive multiple sequence alignment algorithm for biological sequences. These modifications are designed to allow the user to easily incorporate prior knowledge about the sequences and so greatly improve the sensitivity of the resulting alignments.

Poster D-7

Comparison of Current BLAST Software on Nucleotide Sequences

I. Elizabeth Cha (Department of Computer Engineering and Computer Science, University of Louisville, Kentucky); Eric C. Rouchka (Department of Computer Engineering and Computer Science, University of Louisville, Kentucky)

Abstract: The computational power needed for searching large biological databases has increased dramatically. Three implementations of BLAST are studied for their efficiency on nucleotide sequences. The performance is evaluated using databases and query sequences constructed from human genomic and EST sequences. Our results suggest each performs optimally under differing conditions.

Poster D-8

DNA sequence organism determination by sequence composition analysis

Joo Paulo Piazza (Laboratório de Bioinformática / Instituto de Computação / Unicamp); Joo Carlos Setubal (Virginia Bioinformatics Institute / Virginia Tech)

Abstract: We present a new methodology for computational ascertainment of organismal origin of DNA

sequences. It is based on intrinsic information extraction, combination, and classification. When applied to a set of EST sequences, it detected several new contaminations, previously not detected by similarity-based methods.

Poster D-9

Using Partial Least Square Regression to Classify Protein Family with Weak Sequence Similarities

Stephen Opiyo (Department of Agronomy and Horticulture, University of Nebraska, Lincoln); Etsuko Moriyama (School of Biological Sciences and Plant Science Initiative, University of Nebraska, Lincoln)

Abstract: For a better protein classification, how many samples should be included in a training dataset? We examined the effect of training samples on various protein classifiers. We found that the size of training datasets affected hidden Markov models and PSI-BLAST but had little effect on partial least square regression methods.

Poster D-10

Improved EST and genomic sequence alignment

Miao Zhang (Washington University in Saint Louis); Warren Gish (Washington University in Saint Louis)

Abstract: A software tool named Exalin has been developed to align spliced transcript sequences to genomic sequence accurately, particularly in error prone situations. Exalin integrates improved splice site models with dynamic programming. To reduce memory requirements and increase speed, Exalin can be guided by an input file produced by WU-BLASTN.

Poster D-11

A Statistical Method of Identifying Protein Motifs by Combining Amino Acid Sequences and Secondary Structures

Nak-Kyeong Kim (Purdue University); Jun Xie (Purdue University)

Abstract: We developed a new statistical method that models protein motifs by both amino acid sequences and the predicted secondary structures. A parameter SOV (Structural Overlap) is introduced to measure the similarity between secondary structures. Test data sets from BALiBASE show a great improvement in identifying protein motifs.

Poster D-12 (*There will also be an oral presentation of this poster.*)

A comparative genomics approach to distinguishing exon creation and loss events shows increased exon creation associated with alternative splicing

Alexander Alekseyenko (1. Department of Biomathematics, David Geffen School of Medicine, University of California Los Angeles); Christopher Lee (2. Department of Chemistry & Biochemistry, University of California Los Angeles)

Abstract: We have analyzed phylogenies of alternatively spliced exons in four mammalian genomes using outgroups to show that most conservation divergences are due to exon creation events. These data suggest an important role for alternative splicing in accelerating transcriptome evolution and in recruiting new sources of sequence diversity into human proteome.

Poster D-14

Sequence analysis of MNUDC-like protein and SQS among flagellar gene families of Leishmania spp.: new putative virulence factors in leishmaniasis

Diana M. Oliveira (Nucleo de Genomica e Bioinformatica, Universidade Estadual do Ceara); Joao J.S. Gouveia (Nucleo de Genomica e Bioinformatica, Universidade Estadual do Ceara); Elton J.R. Vasconcelos (Nucleo de Genomica e Bioinformatica, Universidade Estadual do Ceara); Michely C. Diniz (Nucleo de Genomica e Bioinformatica, Universidade Estadual do Ceara); Ana C.L. Pacheco (Nucleo de Genomica e Bioinformatica, Universidade Estadual do Ceara); Daniel A. Viana (Nucleo de Genomica e Bioinformatica, Universidade Estadual do Ceara); Allan R.S. Maia (Nucleo de Genomica e Bioinformatica, Universidade

Estadual do Ceara); Nilo B. Diniz (Nucleo de Genomica e Bioinformatica, Universidade Estadual do Ceara); Thiago D. Ferreira (Nucleo de Genomica e Bioinformatica, Universidade Estadual do Ceara); Marianna C. Albuquerque (Nucleo de Genomica e Bioinformatica, Universidade Estadual do Ceara)

Abstract: Through the application of combined bioinformatics tools, we were able to distinguish some flagellar gene families in *Leishmania* species and, among them, the MNUDC-like protein (involved in nuclear development and genomic fusion) and SQS (an enzyme of sterol biosynthesis). Results of this work will be available at <http://nugen.lcc.uece.br/LPGportao>.

Poster D-15

Comparative Analysis of Alignment-based and Alignment-free Protein Classifiers

Pooja Strope (School of Biological Sciences, University of Nebraska - Lincoln); Etsuko Moriyama (School of Biological Sciences, University of Nebraska - Lincoln)

Abstract: We present the performance analysis of various protein classification methods. Our results show that use of support vector machines with amino acid frequencies is promising for protein classification even when sequence similarities are too low to generate reliable multiple alignments or when only short subsequences are available

Poster D-16

Evaluating the Association of Mitochondrial SNP Haplotypes with Disease Phenotypes using a Novel in silico Tool E-MIDAS

Anshu Bhardwaj (Centre for Cellular and Molecular Biology); Shrish Tiwari (Centre for Cellular and Molecular Biology)

Abstract: Based on in silico analysis of human mtDNA we propose the use of specific SNP markers, identified by their DNA sequence context, in association studies. We believe the selection of the subset of markers, suggested by our study, will improve the reproducibility of association studies and reduce the load of genotyping.

Poster D-17

Global Super Paramagnetic Clustering of Protein Sequences

Igor Tetko (Institute for Bioinformatics, MIPS, GSF); Axel Facius (Institute for Bioinformatics, MIPS, GSF); Andreas Ruepp (Institute for Bioinformatics, MIPS, GSF); Dimitrij Surmeli (Institute for Bioinformatics, MIPS, GSF); Werner Mewes (Institute for Bioinformatics, MIPS, GSF)

Abstract: The global SPC (gSPC) algorithm is introduced as an extension to Super Paramagnetic Clustering. The algorithms cluster input data based on a method that is analogous to the treatment of an inhomogeneous ferromagnet in physics. In a benchmark study the gSPC provides improved performance compared to TRIBE-MCL algorithm.

Poster D-18

A Computational Model for Identification of High Occupancy Myc Binding Sites and Prediction of Associated Target Genes

Yili Chen (University of Michigan); Thomas Blackwell (University of Michigan); Angel Lee (University of Michigan); David States (University of Michigan)

Abstract: c-myc is an important transcriptional factor. We build a computational model to predict the likelihood of genomic DNA methylation. This model is combined with cross species analysis to predict c-myc recognition sites likely to exhibit high occupancy binding in chromatin immunoprecipitation studies and to identify associated c-myc target genes.

Poster D-19

Discovery of highly polymorphic genes in tomato cultivars

Angela Baldo (USDA-ARS Plant Genetic Resources Unit); Larry Robertson (USDA-ARS Plant Genetic Resources Unit); Joanne Labate (USDA-ARS Plant Genetic Resources Unit)

Abstract: Cultivated tomatoes are genetically extremely similar. We predicted SNPs from public expressed tomato sequences. Resequencing regions from 53 Unigenes uncovered an unexpected wealth of polymorphism (62 SNPs and 12 indels in 21 Unigenes). This included nonsynonymous nucleotide and nonconservative amino acid changes. We hypothesize these regions represent wild species introgressions.

Poster D-20

Translation from the Structural Language of Proteins into the Language of Nucleotide Sequences as a Possible Mechanism of the Deterministic Evolution of Species

Alexey Melkikh (Ural State Technical University, Molecular Physics Chair, Yekaterinburg)

Abstract: Deterministic model of the evolution has been proposed. The structural language of proteins (where information about species evolution is encoded) is translated into the language of nucleotide sequences during the evolution. The structure of genes is controlled such that the transition to a nearest ecological niche takes a minimum time.

Poster D-21

FESTIVA, a web system for EST analysis and visualization

Yuan-Yuan Li (Shanghai Center for Bioinformation Technology (SCBIT)); Hao Xu (Shanghai Center for Bioinformation Technology (SCBIT)); Hao Tan (Shanghai Center for Bioinformation Technology (SCBIT)); Fu-Dong Yu (Shanghai Center for Bioinformation Technology (SCBIT)); Hong Yu (Shanghai Center for Bioinformation Technology (SCBIT)); Wei-Zhong He (Shanghai Center for Bioinformation Technology (SCBIT)); Yi-Xue Li (Shanghai Center for Bioinformation Technology (SCBIT)); Lei Liu (The W. M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana-Champaign)

Abstract: FESTIVA (Facility for EST Information Visualization and Analysis) was created to meet the requirements of EST analysis and management. It provides a web interface for users to submit a job, query and retrieve the results. FESTIVA has unique features, and has been used for S.Japanicum and B.Napus projects.

Poster D-22

G-InforBIO: An integrated suite for comparative genome analysis

Naoto Tanaka (Institute for Bioinformatics Research and Development (BIRD), Japan Science and Technology Corporation (JST)); Takashi Abe (Center for Information Biology and DNA Data Bank of Japan, National Institute of genetics); Satoru Miyazaki (Faculty of Pharmaceutical Science, Tokyo University of Science); Hideaki Sugawara (Center for Information Biology and DNA Data Bank of Japan, National Institute of genetics)

Abstract: We have developed a suite (G-InforBIO) for management and analysis of genomic data published in the International Nucleotide Sequence Database (DDBJ/EMBL/GenBank) and also local data in research laboratories. Major functions of G-InforBIO are database management, data retrieval, sequence data analysis and the visualization for comparative genomics.

Poster D-23

Molecular and Biochemical Characterization of a Novel Extremophile, Exiguobacterium sp. and its Application in Bioremediation

Anil Kumar (Institute of Genomics and Integrative Biology); Rita Kumar (Institute of Genomics and Integrative Biology)

Abstract: Isolation and characterization of extremophiles have received a great attention because of their potential biotechnological applications. Present work deals with the characterization of Exiguobacterium sp. and its applications. Partial 16S rDNA sequence shows a similarity of 99.4% to Exiguobacterium aurantiacum Z8 but physiological results are different which show its novelty.

Poster D-24

Cryptic Amino Acid Sequences in Model Organism Proteomes

Michelle Simon (MRC Mammalian Genetics Unit); John Hancock (MRC Mammalian Genetics Unit)

Abstract: We extend the characterisation of cryptic amino acid repeats from yeast to a further eight major model organisms. The significant associations of some common repeats with transcription factors and protein kinases seen in yeast are also seen in other species, but we also detect novel associations and differences between species.

Poster D-25

An iterative search algorithm for identifying promoter regions in bacterial genomes

Karthikeyan Sivaraman (Anna University); Gautam Pennathur (Anna University)

Abstract: We describe the results of an iterative search algorithm for predicting promoters. It uses a trained position specific probability matrix to search for high scoring regions in a DNA sequence. It searches dependent signals first and then searches the core signal in the result allowing degeneracy.

Poster D-26

EMBOSS; European Molecular Biology Open Software Suite

Peter Rice (European Bioinformatics Institute); Alan Bleasby (Rosalind Franklin Centre for Genomics Research); Jon Ison (Rosalind Franklin Centre for Genomics Research); Tim Carver (Wellcome Trust Sanger Institute); Lisa Mullan (European Bioinformatics Institute)

Abstract: EMBOSS is an open source project for sequence analysis, protein structure, and other bioinformatics areas. EMBOSS programs can be used for inhouse development and integration of legacy packages. There are over 50 known interfaces to EMBOSS, including the JEMBOSS GUI developed by the EMBOSS team, and Taverna workflows.

Poster D-27

Markov chain Monte Carlo and the repeat history of a sequence

Melanie Huntley (McMaster University, Dept. of Biology); Brian Golding (McMaster University, Dept. of Biology)

Abstract: Repetitive sequences within eukaryotic proteins are abundant, yet little is known about their function and history. We have developed a Markov chain Monte Carlo method to detect such repeats and to infer the evolutionary history of these sequences within proteins.

Poster D-28

An evolutionary hybrid approach to phylogenetic reconstruction

Ryosuke Watanabe (ITESM, Campus Estado de Mexico); Edgar Vallejo (ITESM, Campus Estado Mexico); Enrique Morett (Instituto de Biotecnología, UNAM)

Abstract: We present a method for phylogenetic reconstruction based on genetic algorithms. We formulate a fitness function as a linear combination of UPMGA and MP scores. We use simulated phylogenies and actual HIV sequences to evaluate the proposed method. Experimental results indicate that our model is capable of producing accurate phylogenies.

Poster D-29

Large-scale detection of chromosomal translocation events in the human genome

Qiang Xu (Department of Bioinformatics, Genentech, Inc., 1 DNA Way, South San Francisco, California); Christopher J. Grimaldi (Department of Molecular Biology, Genentech, Inc., 1 DNA Way, South San Francisco, California); Thomas D. Wu (Department of Bioinformatics, Genentech, Inc., 1 DNA Way, South San Francisco, California)

Abstract: Using a genomic mapping and alignment program (GMAP), we identified 341 potential translocation events in the human genome. False-positive and false-negative rates were estimated using independent dataset and simulation test. Chimeric sequences showed a strong association with cancer samples. Some known and novel examples were illustrated.

Poster D-30

Are lateral transfers niche specific?

Weilong Hao (McMaster University); Brian Golding (McMaster University)

Abstract: The study of gene insertions/deletions is necessary to understand bacterial genome evolution. This study shows that more insertions/deletions take place at the tips of phylogeny. This implies that many of the lineage-specific insertions are lost during evolution and that perhaps many of the indels are niche specific.

Poster D-31

A genetic algorithm for bipartite sequence alignment with application to CAR/RXR binding sites

Chengpeng Bi (Children Mercy Hospital); Peter Rogan (Children Mercy Hospital)

Abstract: We developed a new genetic algorithm called Ga-bipad to perform multiple local sequence alignment for bipartite binding sites. The method involves evolving a population of bipartite alignments and gradually improving the fitness of the population as measured by an objective function.

Poster D-32

Evolutionary conserved allosteric network of ligand-gated ion channels revealed by sequence-based statistical method

Yonghui Chen (The University of Alabama at Birmingham); Kevin Reilly (The University of Alabama at Birmingham); Yongchang Chang (Division of Neurobiology, Barrow Neurological Institute, Phonex)

Abstract: Ligand-gated ion channels (LGICs) mediate fast synaptic transmission for communication between neurons. These channels are allosteric proteins, in which binding of a neurotransmitter to its binding site in the extracellular amino-terminal domain triggers structural changes in distant transmembrane domains to open a channel for ion flow. To gain insight in the structural basis of this long-range allosteric coupling, we analyzed multiple sequences of LGICs by adapting a previously developed statistical method as well as a clustering analysis and a 3-D structural model. Statistical coupling energy calculation along with clustering analysis revealed a highly coupled cluster. Mapping the positions in this cluster onto a 3-D structural model demonstrated that these highly coupled positions were mainly clustered in previously identified important functional domains for binding, coupling, and gating. Thus, our results revealed a genetically interconnected network, which potentially plays an important role in the allosteric activation of LGICs.

Poster D-33

Prediction for CH zinc finger protein DNA binding sites.

Min Young Park (Dept of Biochemistry, Han-yang University, Ansan, South Korea); Seong Jin Park (Dept of Biochemistry, Han-yang University, Ansan, South Korea); Chai jin Chai (Dept of Biochemistry, Han-yang University, Ansan, South Korea); Young Seek Lee (Dept of Biochemistry, Han-yang University, Ansan, South Korea)

Abstract: Zinc finger domain has ability to bind to specific DNA sequence with high affinity. Most of their recognition sites are not known yet. We try to predict the binding sites of the rest of the zinc finger containing proteins whose binding sequences are unknown.

Poster D-34

Searching for Regulatory Elements of Alternative Splicing Using Phylogenetic Footprinting

Osamu Maruyama (Faculty of Mathematics, Kyushu University); Daichi Shigemizu Shigemizu (Graduate School of Systems Life Sciences, Kyushu University)

Abstract: We propose a motif-finding method based on phylogenetic footprinting with positive and negative orthologous sequences, and apply it to the problem of finding regulatory elements of alternative splicing. The results show that the candidates of regulatory elements are located in the introns flanking the alternatively spliced exons.

Poster D-35

Relationships among Codon Usage Bias, Disease, and Tissue-wide Coexpression

Hyun Goo Woo (National Genome Information Center(NGIC), KRIBB); Sangsoo Kim (Dept of Bioinformatics, Soongsil Univ.); In-Sun Chu (National Genome Information Center(NGIC), KRIBB)

Abstract: We observed that the genes with highly biased codon usage are more likely to be disease-associated and more likely to encode secretory proteins in human genome. In addition, the rate of disease-genes and codon usage bias are associated with tissue-wide coexpression.

Poster D-36

How well do function classification systems correlate with evolutionarily conserved sequence and structure features of enzymes?

Ranyee Chiang (University of California, San Francisco); Andrej Sali (University of California, San Francisco); Patricia Babbitt (University of California, San Francisco)

Abstract: We assess the Enzyme Commission nomenclature system, Gene Ontology, and binding specificity for how well they capture evolutionarily conserved aspects of function that are associated with conserved protein features.

Poster D-37

Detection of horizontal gene transfer from human to *Schistosoma japonicum* indicates a mechanism for host-parasite interaction

Fu-Dong Yu (Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences); Yuan-Yuan Li (Shanghai Center for Bioinformation Technology); Hao Xu (Shanghai Center for Bioinformation Technology); Hong Yu (Shanghai Center for Bioinformation Technology); Hao Tan (Shanghai Center for Bioinformation Technology); Yi-Xue Li (Shanghai Center for Bioinformation Technology); Lei Liu (The W. M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana-Champaign)

Abstract: Comparing *S.japonicum* ESTs to the human genome with subsequent phylogenetic and codon-usage analysis, we discovered several *S.japonicum* genes that might be of human origin. Function annotations suggest that the parasite may exploit host pathways for development and maturation, and horizontal gene transfer may play a role in host-parasite interaction.

Poster D-38

Trend of amino acid composition of proteins of different taxa

Oxana Galzitskaya (Institute of Protein Research, Russian Academy of Sciences); Alexei Finkelstein (Institute of Protein Research, Russian Academy of Sciences); Natalya Bogatyreva (Institute of Protein Research, Russian Academy of Sciences)

Abstract: Our results demonstrate surprisingly small selection for amino acid composition of proteins of higher organisms and their viruses in comparison with the frequency following from a uniform usage of codons of the universal genetic code, while the lower organisms demonstrate an enhanced selection of amino acids used in these organisms.

Poster D-39

AFLP In Silico Mapping

Jan van Oeveren (Keygene); Antoine Janssen (Keygene); Johan Peleman (Keygene); Harold Verstegen (Keygene)

Abstract: We describe a new and fast procedure for mapping random AFLP fragments. Starting from a complete genome sequence we are able to predict AFLP fragments in silico and compare the results to experimental AFLP fingerprints. Matching fragments are identified and directly positioned on the physical map.

Poster D-40

pSLIP: SVM based Protein Subcellular Localization Prediction using Multiple Physicochemical Properties

Deepak Sarda (Bioinformatics Institute); Gek-Huey Chua (Bioinformatics Institute); Kuo-Bin Li (Bioinformatics Institute); Francis Tang (Bioinformatics Institute); Arun Krishnan (Bioinformatics Institute)

Abstract: We propose the pSLIP algorithm that uses multiple physicochemical properties of amino acids and support vector machines to solve the protein localization prediction problem. Cross-validation tests conducted on eukaryotic proteins belonging to six subcellular locations yielded a prediction accuracy of 93.1% - among the best obtained so far.

Poster D-41

A Novel Methodology for Detecting Low Homology Proteins from the same Family using Conserved Physicochemical Signals

Gek Huey Chua (Bioinformatics Institute); Kuo-Bin Li (Bioinformatics Institute); Deepak Sarda (Bioinformatics Institute); Francis Tang (Bioinformatics Institute); Arun Krishnan (Bioinformatics Institute)

Abstract: An HMM based prediction tool using different physicochemical properties for different regions of a domain is developed for proteins from same family. This method complements existing methods that fall short of identifying distant protein family members and those that only use single or multiple physicochemical properties for the whole protein.

Poster D-42

An automatic method to cluster orthologs of multiple genomes from best reciprocal BLAST hits

Sunshin Kim (Chungbuk National University); Chung-Sei Rhee (Chungbuk National University); Keun Ho Ryu (Chungbuk National University); Hae-Ryong Kwon (Chungbuk National University); Young-Chang Kim (Chungbuk National University); Jeongsu Oh (Chungbuk National University); Taekyuong Kim (Chungbuk National University); Wansup Cho (Chungbuk National University)

Abstract: It takes much time to construct Orthologous Groups(OG). We propose an automatic method to cluster more exact orthologous groups with less manual work. We measured the similarity with COGs and KO, which is about 90 percent and inclines to increase according to the growth of the score cut-off.

Poster D-43

Hierarchical Classification Methods for Biological Patterns

Ilana Granovsky (Technion - Israel Institute of Technology); Isak Gath (Technion - Israel Institute of Technology)

Abstract: In this study we investigate the most appropriate hierarchical clustering method for phylogenetic and forensic analysis based on mitochondrial DNA differences. A method employing pedigrees as a reference for evaluation of the clustering results is outlined and a new custom metric for calculating dissimilarities between the mtDNA sequences is proposed.

Poster D-44

ELEANALYZER: an automated method for analysis of retrotransposons in genomes and prediction of their insertion sites.

Kamal Rawal (Center for Computational Biology and Bioinformatics, School of Information Technology, Jawaharlal Nehru University); Abhijeet Bakre (School of Environmental Sciences, Jawaharlal Nehru University); Alok Bhattacharya (School of Life Sciences, Jawaharlal Nehru University); Sudha Bhattacharya (School of Environmental Sciences, Jawaharlal Nehru University); Ram Ramaswamy (School of Physical Sciences, Jawaharlal Nehru University)

Abstract: ELEANALYSER identifies unoccupied insertion sites, characterizes occupied insertion sites and finds their distribution. The algorithm incorporates Bayesian scoring and boosting. A webserver has been

developed for automated analysis of truncation hotspots. It also allows examination of the occurrence of insertion elements in the vicinity of genes.

Poster D-45

Evolutionary diversity among organelles involved in membrane traffic

Hajime Ohyanagi (CIB/DDBJ, National Institute of Genetics); Takashi Gojobori (CIB/DDBJ, National Institute of Genetics)

Abstract: Each eukaryotic proteins involved in membrane system must have been evolved with each bias of its own, so proteins which are localized in the same position in cells may show the same characteristics. We are conducting analyses of comparative genomics to give insight to the evolutionary studies of membrane system.

Poster D-46 (*There will also be an oral presentation of this poster.*)

A Gamma mixture model better accounts for among site rate heterogeneity

Itay Mayrose (Tel Aviv University); Nir Friedman (Hebrew University); Tal Pupko (Tel Aviv University)

Abstract: A novel evolutionary model is presented in which the variation of evolutionary rates across sequence sites is modeled using a mixture of gamma distributions. We show that the proposed model fits protein datasets significantly better than the most widely used Gamma model, and result in better estimation of evolutionary rates.

Poster D-47

Title: A novel calcium-binding protein in calcified endoskeleton: Global analysis of protein sequence

Md. Azizur Rahamn (University of the Ryukyus, Department of Marine & Environmental Sciences); Yeishin Isa (University of the Ryukyus, Department of Marine & Environmental Sciences); Tsuyoshi Uehara (University of the Ryukyus, Department of Marine & Environmental Sciences)

Abstract: A novel calcium-binding protein of sclerites in alcyonarian has been identified by ⁴⁵Ca autoradiography and N-termini were sequenced. Newly derived protein sequence was subjected to standard sequence analysis involving identification of similarities to other proteins in data bases, prediction of secondary structure, mapping of potential glycosylation and other motifs.

Poster D-49

FunCat functional assignment by Belief Propagation

Dimitrij Surmeli (Institute for Bioinformatics, GSF GmbH, Munich); Oliver Ratmann (Universita di Lecce); Igor Tetko (Institute for Bioinformatics, GSF GmbH, Munich); Hans-Werner Mewes (Germany)

Abstract: We propose a system utilizing Belief Propagation for automatic assignment of FunCat functional categories to sequences. It also uses a Bayesian network to integrate different genomic features. We show results on a benchmark set of four bacterial genomes.

Poster D-50

Automated annotation of a difficult genome - finding genes for a Sea Squirt

Jan-Hinnerk Vogel (Wellcome Trust Sanger Institute); Dan Andrews (The Wellcome Trust Sanger Institute); Laura Clarke (The Wellcome Trust Sanger Institute); Kevin Howe (The Wellcome Trust Sanger Institute); Vivek Iyer (The Wellcome Trust Sanger Institute); Felix Kokocinski (The Wellcome Trust Sanger Institute); Simon White (The Wellcome Trust Sanger Institute); Val Curwen (The Wellcome Trust Sanger Institute); Steve Searle (The Wellcome Trust Sanger Institute); Ewan Birney (The European Bioinformatics Institute); M. Caccamo; G. Cameron; Y. Chen; G. Coates; T. Cox; F. Cunningham; T. Cutts; T. Down; R. Durbin; XM Fernandez-Suarez; J. Gilbert; M. Hammond; J. Herrero; H. Hotz; K. Jekosch; A. Kahari; A. Kasprzyk; D. Keefe; D. London; I. Longden; P. Meidl; G. Proctor; M. Rae; D. Rios; M. Schuster; J. Severin; G. Slater; D. Smedley; J. Smith; W. Spooner; A. Stabenau; J. Stalker; S. Trevanion; A. Ureta-Vidal; C. Woodwark; T. Hubbard

Abstract: Due to its phylogenetic positioning, the genome of *Ciona intestinalis* represents a valuable resource for comparative genomics. However, this evolutionary position and a lack of *C. intestinalis*-specific sequence data make its automatic annotation challenging. We present Ensembl gene annotation of the latest *C. intestinalis* genome sequence assembly.

Poster D-51

EMSA: A Multiple Spliced Alignment Algorithm

Fang Rong Hsu (Department of Information Engineering and Computer Science, Feng Chia University); Wei-Chung Shia (Department of Bioinformatics, Taichung Healthcare and Management University)

Abstract: Traditional spliced alignment algorithms only consider how to align a single EST to genome. Suppose we have many ESTs at hand, it is also important to consider the alignment among all ESTs and the genomic sequence. We propose a new algorithm to align a set of ESTs and the genome.

Poster D-52

Simulation of protein families in the Twilight Zone

Cory L. Strobe (University of Nebraska – Lincoln); Etsuko N. Moriyama (University of Nebraska – Lincoln)

Abstract: We present a tool that introduces insertion and deletion (indel) events based on empirical evidence during sequence simulation. Our method also allows for the creation of multidomain sequence families by specifying different evolutionary rates to different domains, and outputs the "true" multiple alignment of the simulated sequences.

Poster D-53

Evolution of Circular Permutations in Multi-Domain Proteins

January Weiner 3rd (The Westfalian Wilhelms University of Mnster); Erich Bornberg-Bauer (The Westfalian Wilhelms University of Mnster)

Abstract: Circular permutations (CPs) are rearrangements of proteins reversing the order of fragments. We developed an algorithm detecting CPs and searched the PRODOM database. We find several group of CPs and demonstrate that there are mostly two mechanisms. The spread of CPs is surprising and influences our perception of protein evolution.

Poster D-54

Molecular characterization of population diversity in *Feronia limonia* L. (Swingle) using ISSR markers

Shailendra Vyas (Plant Biotechnology Laboratory, Department of Botany, Mohanlal Sukhadia University); Sunil Purohit (Plant Biotechnology Laboratory, Department of Botany, Mohanlal Sukhadia University)

Abstract: Genetic variation within three populations of *Feronia limonia* L. (Swingle) (Rutaceae) from Aravallis of south-east Rajasthan was studied using Inter-short sequence repeat (ISSR) markers. Results have shown that the diversity within *Feronia* is distributed across whole of its range and partitioned mainly within populations (80%).

Poster D-55

Identification and Characterization of Conserved Microsynteny with SynBrowse

Xiaokang Pan (Department of Genetics, Development and Cell Biology, Iowa State University)

Abstract: I have developed a system for identification and characterization of conserved microsynteny among plant genomes using SynBrowse, a synteny browser. With this system, I identified conserved microsyntenic regions among genomes of four plant model systems and performed an evolutionary analysis of these regions (length, genes, gene function annotations, etc.).

Poster D-56

Protein homology detection using sparse profile hidden Markov models

Pai-Hsi Huang (Dept. of Computer Science, Rutgers University, NJ); Vladimir Pavlovic (Dept. of

Computer Science, Rutgers University, NJ); Alexander Kister (Dept. of Mathematics, Rutgers University, NJ)

Abstract: Based on recent research studies, we hypothesize that the knowledge of a set of key residues and the between each neighboring pair allows one to classify a given protein into an appropriate group. We propose a class of models and a training algorithm in attempt to solve this problem.

Poster D-57

Lateral Gene Transfer in Mycobacterium avium subsp. paratuberculosis

Pradeep Reddy Marri (Department of Biology, McMaster University); John P. Bannantine (National Animal Disease Center -USDA-ARS); G. Brian Golding (Department of Biology, McMaster University)

Abstract: We analyzed the genome of Mycobacterium avium subsp. paratuberculosis for lateral gene transfers by first identifying unique genes and then phylogenetically classifying these genes. Presumed ancestral species for 146 out of a set of 804 unique genes were identified. Proteobacteria and soil dwelling actinobacteria contribute to most of the transfers, suggesting that M. a. paratuberculosis may have evolved by acquiring genes necessary for its survival in soil.

Poster D-58

Estimate of Genomewide Mutation Rate Difference between Human and Mouse

HoJoon Lee (Center for Evolutionary Functional Genomics, The Biodesign Institute, Arizona State University); Sankar Subramanian (Center for Evolutionary Functional Genomics, The Biodesign Institute, Arizona State University); Sudhir Kumar (Center for Evolutionary Functional Genomics, The Biodesign Institute, Arizona State University)

Abstract: Long-term mutation rates difference between human and mouse genomes are inferred using 30 largest ancestral interspersed repeat families. In contrast to previous reports of a two-fold difference, we find a 16% rate difference between humans and mice, which is similar to the 18% difference seen in their fourfold-degenerate sites.

Poster D-59

High Throughput Mutation Detection

Andrew Yates (Cancer Genome Project - The Wellcome Trust Sanger Institute); Jon Teague (Cancer Genome Project - The Wellcome Trust Sanger Institute); Ed Dicks (Cancer Genome Project - The Wellcome Trust Sanger Institute); Ken Edwards (Cancer Genome Project - The Wellcome Trust Sanger Institute); Keiran Raine (Cancer Genome Project - The Wellcome Trust Sanger Institute); Simon Forbes (Cancer Genome Project - The Wellcome Trust Sanger Institute); Adam Butler (Cancer Genome Project - The Wellcome Trust Sanger Institute); Andy Futreal (Cancer Genome Project - The Wellcome Trust Sanger Institute); Richard Wooster (Cancer Genome Project - The Wellcome Trust Sanger Institute); Mike Stratton (Cancer Genome Project - The Wellcome Trust Sanger Institute)

Abstract: The Cancer Genome Project is screening the coding Human Genome in a series of tumours to identify somatic mutations and identify novel cancer genes. In order to achieve this we have developed a highly automated system to find these mutations in DNA sequence traces.

Poster D-60

Cross-species Analysis in the Oryza Map Alignment Project (OMAP)

Bonnie Hurwitz (Cold Spring Harbor Laboratory); Dave Kudrna (Arizona Genomics Institute and Computational Laboratories, Department of Plant Sciences); Doreen Ware (Cold Spring Harbor Laboratory); Hye-Ran Kim (Arizona Genomics Institute and Computational Laboratories, Department of Plant Sciences); Yeisoo Yu (Arizona Genomics Institute and Computational Laboratories, Department of Plant Sciences); Kiran Rao (Arizona Genomics Institute and Computational Laboratories, Department of Plant Sciences); Will Nelson (Arizona Genomics Institute and Computational Laboratories, Department of Plant Sciences); Rod Wing (Arizona Genomics Institute and Computational Laboratories, Department of Plant Sciences); Lincoln Stein (Cold Spring Harbor Laboratory); Cari Soderlund (Arizona Genomics

Institute and Computational Laboratories, Department of Plant Sciences)

Abstract: Domestication of cultivated rice is manifested at both the molecular and phenotypic level. Our research seeks to build infrastructure to understand genome evolution, including genome arrangement and diversity between wild and domesticated rice, by comparing the cultivated sequenced diploid rice to 11 wild rice genomes, including both diploids and tetraploids.

Poster D-61 (*There will also be an oral presentation of this poster.*)

Evidence of a large-scale functional organization of mammalian chromosomes

Joel Graber (The Jackson Laboratory); Gary Churchill (The Jackson Laboratory); Keith DiPetrillo (The Jackson Laboratory); Benjamin King (The Jackson Laboratory); Petko Petkov (The Jackson Laboratory); Ken Paigen (The Jackson Laboratory)

Abstract: Studies of a large set of SNPs across sixty inbred mouse strains reveal a pattern of chromosome organization characterized by extensive contiguous domains of functionally related (evidenced by common annotation) elements. We also find associations between widely separated and interchromosomal domains that imply a scale-free network of interactions.

Poster D-62

Expert Gene Annotation Based on Oriented Acyclic Coloured Multigraphs

Sarah Djebali (Genome Organisation and Dynamics Laboratory); Franck Delaplace (Lami Genopole); Hugues Roest Crolius (Genome Organisation and Dynamics Laboratory)

Abstract: Automatic gene annotation is a basic requirement to identify the precise location of all the genes in a genome. We have developed Exogean, an expert gene annotation method based on acyclic oriented coloured multigraphs (AOCM). Here we describe the Exogean algorithm and its application on the annotation of human chromosome 22.

Poster D-63

New Global Protein-Nucleotide Alignment Tool

Boris Kiryutin (NCBI/NLM/NIH); Alexandre Souvorov (NCBI/NLM/NIH)

Abstract: Accurate prediction of splice sites is very important for gene prediction in eukaryota. To address the issue a new global protein-nucleotide alignment tool was created. Special effort is taken to score properly the spliced aminoacids and frameshifts. Examples of protein alignments on the rice and poplar genomes are presented.

Poster D-64

GenDB & The SEED: Two Genome Annotation Systems Integrated

Heiko Neuweger (BRF Cebitec Bielefeld University); Alexander Goesmann (BRF Cebitec Bielefeld University); Daniela Bartels (BRF Cebitec Bielefeld University); Sebastian Konietzny (BRF Cebitec Bielefeld University); Lutz Krause (BRF Cebitec Bielefeld University); Alexander Lenhardt (BRF Cebitec Bielefeld University); Burkhard Linke (BRF Cebitec Bielefeld University); Oliver Rupp (BRF Cebitec Bielefeld University); Veronika Vonstein (FIG); Ross Overbeek (FIG); Folker Meyer (BRF Cebitec Bielefeld University)

Abstract: The continuously growing number of sequenced genomes has increased the need for genome annotation systems that support the in depth analysis of genome features as well as comparative studies across multiple genomes. The integration of the two genome annotation systems GenDB and the SEED offers a multitude of complementary enhancements.

Poster D-65

Tangles Around MAPT

Jaime Duckworth (Laboratory of Neurogenetics, NIA, NIH); Amanda Myers (Laboratory of Neurogenetics, NIA, NIH); Alan Pittman (University College London); Rohan de Silva (University College London); Hon Chung Fung (Chang Gung University); John Hardy (Laboratory of Neurogenetics, NIA, NIH)

Abstract: With human genetics and sequence analysis, we found the evidence of large genomic rearrangement at chromosome 17q21.31 from Genbank clones annotated as shuffled sequence pieces to Human Assembly. Mechanisms of the rearrangement are proposed. Such computational studies are expected to shed lights to the aetiology of numerous neurodegenerative diseases.

Poster D-66

Phylogenetic Factorial Hidden Markov Models for Detecting Mosaic Structures in DNA Sequence Alignments

Dirk Husmeier (Biomathematics and Statistics Scotland)

Abstract: A phylogenetic tree is married to a factorial hidden Markov model to detect mosaic structures in DNA sequence alignments. The first hidden state represents tree topologies. The second hidden state represents different selective pressures. The model was found to successfully distinguish between recombination and rate variation.

Poster D-67

Propagation of a novel MITE in cyanobacteria revealed by comparative genomics

Toshiaki Katayama (University of Tokyo); Rei Narikawa (University of Tokyo); Shinobu Okamoto (Kyoto University); Jeff Elhai (Virginia Commonwealth University); Masahiko Ikeuchi (University of Tokyo); Minoru Kanehisa (Kyoto University)

Abstract: We found a novel MITE (AnaMITE1) in cyanobacteria and performed a comprehensive search for this transposable element, taking into account predicted secondary structure. We identified 283 candidates, and by comparing syntenic genome sequence regions of two *Anabaena* strains, found evidence regarding the mechanism of their propagation.

Poster D-68

The LAMA and CYRCA Protein Profile-to-Profile Alignment and Multiple Profile Alignment Methods

Milana Frenkel-Morgenstern (Weizmann Institute of Science); Shmuel Pietrokovski (Weizmann Institute of Science)

Abstract: We present updates and improvements to our profile-profile alignment programs. Input can be gapped (block) or ungapped profiles. Output alignments significance is approximated by E-values based on profile column and alignment score probabilities. OneBlock CYRCA, is a new automated procedure, identifying multiple block alignments from single block queries.

Poster D-69

Evolution and Structural Signature of the Cadherin Domain

Shoshana Posy (Columbia University); Barry Honig (Columbia University)

Abstract: Cadherins mediate cell-cell adhesion and regulate tissue development during morphogenesis. We identify a hydrophobic structural signature of the cadherin domain and show that the strand-swapped interface evolved from a conserved internal motif. We present a model of cadherin evolution from an ancestral domain to multidomain families.

Poster D-70

HHsearch – Protein Homology Detection by HMM–HMM Comparison

Johannes Sding (Max-Planck-Institute for Developmental Biology)

Abstract: Pairwise comparison of hidden Markov models (HMMs) has the potential for improved sensitivity and alignment quality over comparison of sequence profiles. HHsearch is based on a statistical approach to HMM-HMM comparison. We benchmarked HHsearch with PSIBLAST, HMMER, PROF.SIM, COMPASS, and PRC by pairwise comparison of 3691 SCOP domains.

Poster D-71

Evidence-N: Integrating Diverse Sources of Evidence for Automatic Bacteria Gene Identification

Yujing Zeng (University of Delaware); Javier Garcia-Frias (University of Delaware); Jean-Francois Tomb (DuPont company)

Abstract: Gene identification is an essential first step for large-scale analysis of genomes. This poster proposes a framework for automated bacterial gene finding that integrates multiple features using the Dempster-Shafer theory and network analysis. The integration model shows promising performance when applied to several completed genomes.

Poster D-72 (*There will also be an oral presentation of this poster.*)

Trees and forests: A genome-wide reconstruction of orthologous gene groups in fungi

Ilan Wapinski (Harvard University); Nir Friedman (Hebrew University); Aviv Regev (Harvard University); Avi Pfeffer (Harvard University)

Abstract: As more genome sequences become available, resolving orthologous relations between genes is becoming increasingly challenging and important. We present a reliable method for genome-wide reconstruction of ancestral relations between genes across multiple species. Tracing divergence and duplication events allows us to hypothesize about the evolutionary forces governing genetic adaptations.

Poster D-73

A Comparison Tool for Bacterial Taxonomic Distributions

Qiong Wang (Michigan State University); Benli Chai (Michigan State University); Ryan Farris (Michigan State University); Siddique Kulam (Michigan State University); Donna McGarrell (Michigan State University); George Garrity (Michigan State University); James Tiedje (Michigan State University); James Cole (Michigan State University)

Abstract: Microbial community comparison based on 16S rRNA gene sequence libraries has become commonplace in microbial ecology. However, most comparison methods fail to put differences in a taxonomic context. The RDP Sample Comparison Tool combines a rapid taxonomic classifier with a statistical test to flag taxa differing significantly between samples.

Poster D-74 (*There will also be an oral presentation of this poster.*)

Structure prediction and analysis of remote homology relations for proteins in complete genomes

Ruslan Sadreyev (Howard Hughes Medical Institute); Nick Grishin (Howard Hughes Medical Institute, University of Texas Southwestern Medical Center)

Abstract: For the proteins from complete genomes represented in the COG database, we combine definition of mobile sequence modules by ADDA, structural domain classification by SCOP, and remote sequence homology detection in order to provide a comprehensive database of structural domain predictions, and compare networks of sequence-based and structure-based protein relations.

Poster D-75 (*There will also be an oral presentation of this poster.*)

The Canonical Representation of Proteins: Application to Remote Homology Recognition

Chin-Jen Ku (Cornell University); Golan Yona (Cornell University)

Abstract: We present a protein representation where each protein is mapped to a high-dimensional feature space based on its association profile with respect to a library of representative proteins. Through an adequate processing of the feature vectors, this representation produces a significant improvement in the detection of remote homology.

Poster D-76 (*There will also be an oral presentation of this poster.*)

Domain-based Protein Hierarchy and Detection of Semantically Significant Domain Architectures

Chin-Jen Ku (Cornell University); Golan Yona (Cornell University)

Abstract: We present two hierarchical organizations of multi-domain proteins and two hierarchical relationships among domain families. We validate the relationships based on the biological (dis)similarity between the proteins associated to these domain families. We introduce the notion of semantically significant domain architecture and propose a method to detect it.

Poster D-77

Using Bait and Prey Sequences Obtained in the Y2H Experiment of *Drosophila melanogaster* for Further Genome Annotation

Shan Guan (Johns Hopkins University); Joel Bader (Johns Hopkins University)

Abstract: Based on bait and prey sequences obtained in the Y2H experiment of *Drosophila melanogaster* and BDGP database, (1) hundreds of potential novel genes were found and mapped on the chromosomes; (2) Indels and substitutions originated from individual diversity were analyzed; (3) Details of some predicted splicing sites were examined.

Poster D-78

Computationally exploring DNA recognition codes for C2H2 zinc-finger transcription factors

Jiajian Liu (Department of Genetics, Washington University School of Medicine); Gary Stormo (Department of Genetics, Washington University School of Medicine)

Abstract: Identifying DNA recognition codes has been one of the most challenging problems in both computational biology and molecular biology. We developed a new computational method for predicting DNA recognition codes of transcription factors of the C2H2 zinc-finger protein family based on the non-linear neural network model. Using cross validation tests, our binary predictions for various C2H2 zinc-finger proteins were verified with a sensitivity of 93% and a specificity of 88%. In addition, the preferred DNA sites predicted with network models for test zinc finger proteins were also confirmed by comparing them with experimentally determined specificities.

Poster D-79

Large scale analysis of exon boundary conservation in low sequence identity proteins using multiple structural alignments by TOPOFIT

chesley leslin (Department of Biology, Northeastern University); Alex Abyzov (Department of Biology, Northeastern University); Valentin Ilyin (Department of Biology, Northeastern University)

Abstract: Gene structure information from SEDB has been merged onto 11,310 multiple structure sequence alignments allowing for a thorough examination of exon boundary conservation, and shown the occurrence of a threshold where there is a significant decrease in exon boundary conservation, coupled with a dramatic increase in nonconserved exon boundaries.

Poster D-80

Molecular evolution of SET-domain protein families in eukaryotes

Chendhore Veerappan (University of Nebraska-Lincoln); Zoya Avramova (University of Nebraska-Lincoln); Etsuko Moriyama (University of Nebraska-Lincoln)

Abstract: Eukaryotic SET-domain protein families are involved in histone tail lysine methylation in gene activation/repression mechanisms. To elucidate evolutionary relationships and distribution, we conducted genomic searches for various eukaryotic genomes. Results showed species-specific clusters and lineage-specific gain and loss, indicating their functional specialization in different organismal lineages.

Poster D-81

Mauve Multiple Genome Alignment Enables Detection of Extensive Homologous Recombination in Enterobacteria

Aaron Darling (Dept. of Computer Science, University of Wisconsin-Madison); Bob Mau (Dept. of Animal Health and Biomedical Sciences, Univ. of Wisconsin-Madison); Jeremy Glasner (Dept. of Animal Health and Biomedical Sciences, Univ. of Wisconsin-Madison); Nicole Perna (Dept. of Animal Health and

Biomedical Sciences, Univ. of Wisconsin-Madison)

Abstract: Mauve constructs global alignments of multiple rearranged genomes, enabling analysis of phylogenetic signal in conserved regions of seven enterobacteria with a novel random-walk technique. The analysis provides support for numerous homologous recombination events during the evolution of enterobacteria. Interestingly, the analysis reveals a bias towards homologous recombination of biosynthetic genes.

Poster D-83

Comparison of the genomes of T7-like Viruses

Jianwen Fang (University of Kansas); Ryan Haas (University of Kansas); Ashok Bhagwat (Wayne State University)

Abstract: T7-like viruses are a family of double-stranded bacteriophages which are characterized by short, non-contractile tails. In this study we compare eight sequenced T7-like viral genomes to investigate the temporal and functional distribution of genes, the codon usage and composition patterns of these genomes.

Poster D-84

SNPNB: analysis of neighboring-nucleotide biases on SNPs and evaluation of the effective size of SNPs

Zhongming Zhao (Virginia Commonwealth University, USA and Kunming Institute of Zoology, China); Fengkai Zhang (Virginia Commonwealth University)

Abstract: SNPNB is a user-friendly and platform-independent application for analyzing SNP neighboring sequence context and nucleotide bias patterns, and subsequently evaluating the effective SNP size for the bias patterns observed from the whole (e.g., genome or chromosome) data. SNPNB and its full description are freely available at <http://bioinfo.vipbg.vcu.edu/SNPNB/>.

Poster D-85

Novel genome comparison analysis using visual analytics and high-performance computing resources

Heidi Sofia (Pacific Northwest National Laboratory); Grant Nakamura (Pacific Northwest National Laboratory); Joel Malard (Pacific Northwest National Laboratory)

Abstract: We have developed a novel method for gene neighbor prediction with dramatically improved accuracy and resolution based on evolutionary structures in the data detected using the Similarity Box Java visualization software. We are currently building a high-throughput version using advanced computing resources such as parallel BLAST and clustering codes.

Poster D-86 (*There will also be an oral presentation of this poster.*)

Pattern-based phylogenetic distance estimation and tree reconstruction

Michael Hoehl (Institute for Molecular Bioscience, The University of Queensland); Isidore Rigoutsos (Bioinformatics and Pattern Discovery Group, IBM Thomas J Watson Research Center); Mark Ragan (Institute for Molecular Bioscience, The University of Queensland)

Abstract: We have developed a novel alignment-free method for estimating phylogenetic distances, and inferred trees from them. We have evaluated our pattern-based approach against trees resulting from an alignment benchmark database, and compared its accuracy to that of a conventional approach employing a local or global alignment tool.

Poster D-87

A Novel Bioinformatics Approach for Microbial diversity of Environmental Samples on the basis of Self-Organizing Map (SOM).

Takashi Abe (Center for Information Biology and DNA Data Bank of Japan National Institute of Genetics); Toshimichi Ikemura (The Graduate University for Advanced Studies); Makoto Kinouchi (Yamagata Univ); Shigehiko Kanaya (Nara Institute of Science and Technology); Hideaki Sugawara

(Center for Information Biology and DNA Data Bank of Japan National Institute of Genetics)

Abstract: We developed Self-Organizing Map as a novel bioinformatics strategy to capture and visualize microbial diversity and relative abundance of microorganisms within an environmental sample. The classification could be done without orthologous sequence sets. Therefore, SOM was especially useful to analyze novel sequences from poorly characterized species for industrial applications.

Poster D-88

Evaluation of alternative methods for determining maximally informative tagSNPs

Jyoti Shah (University of Texas Southwestern Medical Center); Jennifer Cai (University of Texas Southwestern Medical Center); Richard Scheuermann (University of Texas Southwestern Medical Center)

Abstract: In this study, we evaluated most of the contemporary alternative methods and computational approaches to determine the best method to select tagSNPs for candidate-gene risk association studies. Preliminary analysis suggests that linkage disequilibrium methods to determine maximal sets of informative SNPs are more effective than alternative approaches.

Poster D-89

FOOTER: a quantitative comparative genomics method for efficient recognition of cis-regulatory elements

David Corcoran (University of Pittsburgh); Eleanor Feingold (University of Pittsburgh); Jessica Dominick (University of Pittsburgh); Marietta Wright (Childrens Hospital of Pittsburgh); Massimo Trucco (Childrens Hospital of Pittsburgh); Nick Giannoukakis (Childrens Hospital of Pittsburgh); Panayiotis Benos (University of Pittsburgh)

Abstract: We have developed a novel method, called Footer, for pattern identification that compares a pair of putative binding sites in two species and assigns two probability scores based on the relative position of the sites in the promoter and their agreement with a known model of binding preferences.

Poster D-90

BASys: A Web Server for Automated Bacterial Genome Annotation

Gary Van Domselaar (University of Alberta); Paul Stothard (University of Alberta); Savita Shrivastava (University of Alberta); Joseph Cruz (University of Alberta); An Chi Guo (University of Alberta); Xiaoli Dong (University of Alberta); Paul Lu (University of Alberta); Duane Szafron (University of Alberta); Russ Greiner (University of Alberta); David Wishart (University of Alberta)

Abstract: BASys (Bacterial Annotation System) is a web server designed to permit automated, in-depth annotations of bacterial genomes. Using only the DNA sequence as input BASys will provide more than 60 data fields about each gene along with clickable, browseable maps and files that can be readily downloaded.

Poster D-91

Strategies for identifying homologs of yeast autophagy proteins in Arabidopsis Thaliana

Chi-Fai Kwan (Bioinformatics Program, Eastern Michigan University); Marianne Laporte (Biology Department, Eastern Michigan University); Benjamin Keller (Computer Science, Eastern Michigan University)

Abstract: We are identifying in Arabidopsis thaliana putative homologs of proteins involved in autophagy in Saccharomyces cerevisiae (yeast). Some of the yeast proteins may not occur in plants, and the others contain common domains that complicate homology searches. We describe a strategy that helps reduce the hits in the latter case.

Poster D-92

In silico identification of nucleus-encoded mitochondrial proteins in protists

Yaoqing Shen (University of Montreal)

Abstract: In silico identification of nucleus-encoded mitochondrial proteins based on Jacobids ESTs data is made by both sequence alignment and available programs, which employ machine learning approaches.

Only about 26% 37% of the two results are consistent. The reason of the inconsistency is analyzed.

Poster D-93

A Statistical Geometry Approach to the Study of Functional Effects of Human Non-Synonymous SNPs

Maxim Barenboim (George Mason University); Curtis Jamison (George Mason University); Iosif Vaisman (George Mason University)

Abstract: Predicting the effect of non-synonymous SNPs on protein function is important for association studies. We present a novel statistical geometry approach to nsSNP classification based on Delaunay tessellation. The nsSNP impact on protein correlates with change in the protein four-body statistical potential caused by amino acid substitution.

Poster D-94

IFREE: A tool for the efficient tree-based search of biomolecular sequences based on multiple Position Frequency Matrices

Jubin Sanghvi (University of Missouri-Kansas City); Deendayal Dinakarpanidian (University of Missouri-Kansas City)

Abstract: We have developed IFREE (Indexed Forest of Representer Expression Extractor), a tool that dynamically creates an indexed search forest from a set of Position Frequency Matrices (PFMs). This helps in the rapid PFM-based screening of motifs in a given query sequence.

Poster D-95

Phylogenetic Motifs: A Robust Methodology for Accurate Protein Functional Site Prediction

David La (Department of Biological Sciences, California State Polytechnic University, Pomona); Dennis Livesay (Department of Chemistry, California State Polytechnic University, Pomona)

Abstract: Phylogenetic motifs are multiple sequence alignment fragments that conserve the overall familial phylogeny. Across a structurally and functionally heterogeneous dataset, phylogenetic motifs correspond to surface loops, active site clefts, and multimer interfaces. We present an algorithm to identify phylogenetic motifs and demonstrate its robustness in large-scale protein functional site prediction.

Poster D-96

A New Kernel Based on a Matrix of High-Scored Pairs of k-Peptides and Its Application in Protein Subcellular Localization Prediction

Zhengdeng Lei (Dept. Bioengineering/University of Illinois at Chicago); Yang Dai (Dept. Bioengineering/University of Illinois at Chicago)

Abstract: A new SVM kernel for protein sequences is defined in a space to which the k-peptide vectors are mapped based on a matrix formed by the BLOSUM scores associated with a pair of k-peptides. Computational results demonstrated its superior performance in prediction of protein sub-cellular localizations.

Poster D-97

Classification of Low Complexity Regions in Proteins

Nicolas Tilmans (University of Maryland); Steve Mount (University of Maryland)

Abstract: Low complexity regions within proteins have diverse and important biological functions. While a number of tools have been developed for identifying low complexity sequences, their functional classification lags behind that of more complex domains. We describe a database of low complexity sequence profiles for the classification of these domains.

Poster D-98

PFP: Automatic annotation of protein function by relative GO association in multiple functional contexts

Troy Hawkins (Purdue University); Daisuke Kihara (Purdue University)

Abstract: Our automated function prediction method, PFP, produces a scored list of GO function annotations in order of relative probability within each of the three GO function categories. Scores are assigned from the calculated association of all GO functions in unique functional contexts and provide comprehensive coverage of functional space.

Poster D-99

The impact of 6000 human-specific Alu insertions on the human genome

Ping Liang (Roswell Park Cancer Institute); Jianxin Wang (Roswell Park Cancer Institute)

Abstract: Comparative genome analysis reveals that Alu retrotransposition is three times more active in the human genome than in the chimpanzee genome. Among the 6000 human specific Alu insertions, we identified cases in which Alus participate in generating human specific genes or alternative splicing likely related to disease susceptibility.

Poster D-100

Identification of Genomic Islands: a key to the evolution of Neisseria

Catherine Putonti (University of Houston); Audrey Hart-Van Tassell (University of Texas Medical Branch); Petri Urvil (University of Texas Medical Branch); Meizhuo Zhang (University of Houston); Sergey Chumakov (University of Guadalajara); Bogdan Nowicki (University of Texas Medical Branch); Stella Nowicki (University of Texas Medical Branch); Yuriy Fofanov (University of Houston)

Abstract: Utilizing a novel graphical approach for rapid large scale comparison of genomic sequences, we have analyzed four Neisseria genomes. This study revealed a number of new potential pathogenicity islands. Furthermore, comparison of each genome to the remaining three Neisserias provides greater insight into the evolution of the species.

Poster D-101

phyloXML: Describing Annotated Phylogenetic Trees With XML

Christian Zmasek (Genomics Institute of the Novartis Research Foundation); Ethalinda Cannon (University of Minnesota)

Abstract: We present a definition for phyloXML, an XML document type to describe richly annotated phylogenetic trees; together with an updated version of the tree display and phylogenomic analysis tool ATV, which is able to read and write trees in phyloXML. Website: <http://www.phylogenomics.us>.

Poster D-102

An Event-Driven Generator for Evolutionary Networks

Monique Morin (Dept. of Computer Science, U. of New Mexico)

Abstract: Since trees are not appropriate models for some evolutionary histories, we developed a new simulator to support research in network evolution and reconstruction. Our simulator, NetGen, extends the traditional birth-death model to incorporate diploid hybrids formed using simulated sequence information as well as future extensions to lateral gene transfer.

Poster D-103

Substrate Specificity Of Acyl-Adenylate And Glycosyltransferase Superfamily: An In Silico Analysis

Pankaj Kamra (National Institute of Immunology); Rajesh Gokhale (National Institute of Immunology); Debasisa Mohanty (National Institute of Immunology)

Abstract: Numerous uncharacterized proteins belonging to acyl-adenylate and glycosyltransferase superfamily have been found in various genomes. Using information from three-dimensional structure, comparative analysis of the sequences of proteins with known substrates and structural modeling of the substrate in the active site have provided clues about structural basis for substrate specificity.

Poster D-104

A Unifying Elementary Operation to Calibrate Genomic Distance

Richard Friedberg (Columbia University, Department of Physics); Oliver Attie (Center for the Study of Gene Structure and Function, Hunter College); Sophia Yancopoulos (Institute for Medical Research at NS-LIJ Health System)

Abstract: We seek a biologically reasonable elementary operation for which genomic distance has a simple form. Our DCJ operation produces familiar operations in different contexts, e.g. for linear chromosomes: reversals, translocations, fissions, fusions and temporary circular chromosomes which are reabsorbed ("generalized transposition"). The genomic distance is $b(\text{breakpoints}) - c(\text{cycles})$.

Poster D-105

An Expectation-Maximization approach for dealing with imbalanced data for human splice site Prediction problem

Kihoon Yoon (The University of Texas at San Antonio); Stephen Kwek (The University of Texas at San Antonio)

Abstract: An accurate splice site predictor is important for a reliable gene finding system and also interesting from Machine Learning perspective due to the data imbalance. Most standard approaches cannot handle the problem well. Here, we proposed a better splice site predictor by reducing the imbalance.

Poster D-106

A quantitative analysis of conserved gene order in complete genomes

Natalia Khuri (The Carnegie Institution); Devaki Bhaya (The Carnegie Institute)

Abstract: We explore the conservation of gene order among the 44 complete bacterial and archaeal genomes revealing striking differences in spatial organization of related microbial strains. Synteny (gene order conservation) is not correlated with gene content and 16S rRNA identity. The genomes of thermophiles show far less synteny than mesophilic strains.

Poster D-107

Segmenting Eukaryote Genomes with the Generalised Gibbs Sampler

Jonathan Keith (Department of Mathematics, University of Queensland)

Abstract: This poster presents new statistical techniques, algorithms and software for detecting change-points in eukaryotic genomes. These are points at which some property of interest, such as GC content or degree of evolutionary conservation, changes abruptly, thus dividing the sequence into segments. Results for human chromosome 1 are presented.

Poster D-108

Tracking of NBS-LRR homologues in the recent allohexaploid, wheat.

Franco Du Preez (University of Pretoria); Anna-Maria Oberholster (University of Pretoria); Alexander Myburg (University of Pretoria)

Abstract: Nucleotide-binding site leucine-rich repeat (NBS-LRR) genes play a central role in recognizing intracellular avirulence proteins introduced by phytopathogenic organism to plant host cytoplasm. The NBS-LRR family is large (1% of Arabidopsis genes), and the evolution of this family was investigated by PCR and bioinformatic analysis of available sequence data. The results are interpreted in context of existing models of multi-gene family evolution.

Poster D-109

An alignment algorithm of stem candidates for comparing RNA sequences

Kiyoshi Asai (University of Tokyo); Yasuo Tabe (University of Tokyo); Taishin Kin (University of Tokyo)

Abstract: Our proposed algorithm for RNA sequence comparison considers both the secondary structures and the sequence similarities by simply aligning the potential stem candidates. The algorithm has compatible performance with existing methods for predicting common secondary structures and is much

better than plain sequence alignments for clustering non-coding RNAs.

Poster D-110

Mapping And Identification Of Sequence Flips In ESTs

Irene Gabashvili (Hewlett-Packard Labs); Richard Carter (Hewlett-Packard Labs); Peter Markstein (Hewlett-Packard Labs)

Abstract: Very-large-scale mapping of Expressed Sequence Tags (ESTs) is becoming more important with rapid accumulation of data. We employed our fast multiple-pattern search algorithm for both alignment of potentially spliced and error-containing sequences and genome-context-based analysis. Tissue-specific and disease EST findings will be discussed.

Poster D-111

Sequence Conservation across Presenilin Family Members in Transmembrane Domains

Darryl Gietzen (Accelrys, Inc.); Chris Lindley (Accelrys, Inc.)

Abstract: Presenilins are polytopic membrane proteins. PS1, APH2, nicastrin (NCT) and Presenilin Enhancer protein (Pen2), form a multi-protein complex called gamma-secretase, an active enzyme in Alzheimer's pathogenesis. In this study we used the tools of the GCG Wisconsin Package to identify and analyze homologs of presenilins.

Poster D-112

BioSPLASH: A sample workload from bioinformatics and computational biology for optimizing next-generation high-performance computer systems

David Bader (University of New Mexico); Vipin Sachdeva (University of New Mexico); Amitabh Trehan (University of New Mexico); Virat Agarwal (Indian Institute of Technology, Delhi); Gaurav Goel (Indian Institute of Technology, Delhi); Abhishek Singh (Indian Institute of Technology)

Abstract: We present BioSplash, our effort to assemble a suite of life science applications that span bioinformatics and computational biology and are useful for designing high-performance computing systems. We use the cycle-accurate IBM MAMBO simulator, and real performance monitoring using an Apple G5 workstation.

Poster D-113

Prediction of MHC Class II Binding Peptides Based on an Iterative Supervised Learning Model

Naveen Murugan (University of Illinois at Chicago); Yang Dai (University of Illinois at Chicago)

Abstract: An iterative supervised learning method based on a linear programming model for the prediction of MHC class-II binding peptides is developed. This model is fast and involves no parameter tuning. The results on 10 benchmark datasets demonstrate its competitive performance compared to other advanced methods.

Poster D-114

Classification of antimicrobial peptides using decision tree

Su Yeon Lee (Seoul National University Biomedical Informatics (SNUBI)); Ju Han Kim (Seoul National University Biomedical Informatics (SNUBI))

Abstract: The purpose of this study was to investigate the use of decision tree for the classification of antimicrobial peptides. We describe a successful application of decision tree that provides the understanding of the effects of physicochemical characteristics of peptides on bacterial membrane.

5 Genomics and Gene Expression

Poster E-1

Use Of Microarrays To Identify New Downstream Target Genes For Transcription Factors. Application To The Differentiation Of Adult Stem Cells (MSCs) Into Chondrocytes And Adipocytes

Jason R. Smith (Tulane National Primate Research Center); Joni Ylostalo (Center for Gene Therapy, Tulane University); Radhika Pochampally (Center for Gene Therapy, Tulane University); Robert Matz (Center for Gene Therapy, Tulane University); Ichiro Sekiya (Orthopedic surgery, Division of Bio-Matrix); Benjamin L. Larson (Center for Gene Therapy, Tulane University); Jussi T. Vuoristo (Biocenter Oulu, University of Oulu); Darwin J. Prockop (Center for Gene Therapy, Tulane University)

Abstract: We describe an oligonucleotide microarray based screen for transcription factor target genes. Samples of mRNA were obtained from human marrow stromal cells under three culture conditions (six total experiments). We demonstrate literature concordance and verify predicted interactions of PPARgamma protein with the promoters of four out of five predicted targets.

Poster E-3

Developing Software to Design SNP Probes for HLA Genotyping

Wenbo Xu (Bioinformatics Division, Center for Molecular Medicine & Genetics, School of Medicine, Wayne State University); David Womble (Bioinformatics Division, Center for Molecular Medicine & Genetics, School of Medicine, Wayne State University)

Abstract: A software was developed in our lab to automatically choose SNP-based probes from aligned human major histocompatibility complex (MHC, or HLA) DNA sequence based on GC content, sodium concentration, desired T_m and nucleotide length. These probes can be used for fast, and large scale MHC genotyping using microarray technology

Poster E-4

Functional genomic analysis of a hypertension QTL on Rat Chromosome 1

Richard Dixon (Dept. of Cardiovascular Sciences, Leicester University)

Abstract: 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.

Poster E-5

Predictive taxonomy of individual microarray samples of skeletal muscle in connection with insulin resistance and type II diabetes

Andrey Ptitsyn (Pennington Biomedical Research Center); Steven Smith (Pennington Biomedical Research Center); Olga Kutnenko (Sobolev Institute of Mathematics); Nikolay Zagoruiko (Sobolev Institute of Mathematics)

Abstract: We have developed a new classification algorithm ranking patients by the severity of symptoms and known risk factors. We have re-analyzed the microarray data related to type II diabetes and developed a predictive taxonomy, estimating risk of diabetes based on clinical data (BMI, insulin sensitivity, etc.) and individual genetic makeup.

Poster E-6

A Significant Difference in output among Microarray Experiment Front-end Tools

Ezekiel Adebiyi (Department of Computer and Information Sciences, Covenant University); Segun Fatumo (Department of Computer and Information Sciences, Covenant University); Victor Osamor (Department of Computer and Information Sciences, Covenant University)

Abstract: 25-mer oligo arrays (Affymetrix gene Chip) could be hybridization problematic and longer oligos might be needed to generate accurate expression profiles (Le Roch et al., 2003). A computational analysis of existing methods for finding long oligos shows there is a significant variant among the algorithms, in term of their output.

Poster E-7

Genome wide expression profiling of *Aspergillus fumigatus* against amphotericin B using a DNA Microarray Technology

Jata Shankar (Institute of Genomics and Integrative Biology); Taruna Madan (Institute of Genomics and Integrative Biology); Seemi Basir (Jamia Millia Islamia); Usha Sarma (Indian Agricultural Research Institute)

Abstract: Current study involves, analysis of 160 expressed sequence tags of *Aspergillus fumigatus* (Afu) and differential gene expression of Afu genome was examined with respect to Amphotericin B treatment using microarray technology in order to facilitate the gene identification of Afu relevant for pathogenesis and potential anti-fungal drug targets.

Poster E-8

Application of Genetic Algorithm/K-Nearest Neighbor Method to Cancer Classification Using Gene Expression Data

Dongqing Liu (Department of Computer Science, University of Akron); Zhong-Hui Duan (Department of Computer Science, University of Akron); Jianping Zhu (Department of Theoretical and Applied Mathematics, University of Akron)

Abstract: We use a genetic algorithm and k-nearest neighbor method to classify cancer subtypes based on microarray gene expression data. We show that the algorithm can efficiently identify a panel of discriminator genes. We also analyze the robustness, stability and sensitivity of the algorithm.

Poster E-9 (*There will also be an oral presentation of this poster.*)

TEPC: Total Evidence Phylogenetic Correlation of Microbial Phenotypes and Genotypes

Indra Sarkar (American Museum of Natural History); Paul Planet (American Museum of Natural History); Rob DeSalle (American Museum of Natural History)

Abstract: We propose an evolutionary framework for organizing information from completed microbial genomes and the study of genes or residues that are historically correlated with laboratory phenotypes, disease entities, and clinical symptoms. This framework uses phylogeny to account for statistical biases that may obfuscate the identification of relevant genes or residues.

Poster E-10

Microarray data analysis and pathway activity inference in PATIKA

Ozgun Babur (Bilkent University); Asli Ayaz (University of California, Irvine); Recep Colak (Bilkent University); Emek Demir (Bilkent University); Ugur Dogrusoz (Bilkent University)

Abstract: Pathway activity inference attempts to infer differential activity of cellular networks, given a qualitative state - transition model of the network and an expression profile of RNA molecules. We present an efficient algorithm for this problem, implemented as part of microarray data analysis component of PATIKA, a pathway analysis tool.

Poster E-11

ARTADE: a new tool for tiling-array-based estimation of transcriptional structures for coding and non-coding genes

Tetsuro Toyoda (RIKEN Genomic Sciences Center)

Abstract: Tiling arrays are powerful tools for gene discovery. However, it is difficult to determine the structure of new gene from noisy array signals only. We present a statistical method that estimates the precise exon/intron structure of a structurally unknown gene based on both tiling-array data and sequence data.

Poster E-12

Gene Function Prediction using Specific Gene Expression of Biological Functions

Satoshi Kamegai (INTEC Web and Genome Informatics Corporation); Takuya Oyama (INTEC Web and Genome Informatics Corporation); Mikio Yoshida (INTEC Web and Genome Informatics Corporation); Fumihito Miura (University of Tokyo); Kenji Satou (Japan Advanced Institute of Science and Technology); Takashi Ito (University of Tokyo)

Abstract: We studied to discover an association between gene expression and biological functions in order to reveal functions of uncharacterized genes. Our approach could discover several biological functions having specific gene expression, and adapted to function prediction of uncharacterized genes.

Poster E-13

Gene Expression, Connectivity and Neural Contributions: A Bridge Too Far?

Alon Kaufman (The Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem.); Eytan Ruppin (Schools of Computer Science and Medicine, Tel-Aviv University, Tel-Aviv.)

Abstract: Does the genetic expression of neurons and their connectivity architecture determine their functional role? We address this fundamental question utilizing micro-array data, electron microscopic reconstruction and neuron laser ablation experiments. We provide a positive answer to this question in the case of Amphid chemosensory neurons of *C. elegans*.

Poster E-14

Mapping Gene Expression Data to Pathways

Thomas Karopka (University of Rostock, Institute for Medical Informatics and Biometry); nne Glass (University of Rostock, Institute for Medical Informatics and Biometry)

Abstract: XMMapper - a software tool for mapping microarray gene expression data to pathway diagrams. The purposes are first, the generation of new or integration of existing biological pathways, second, the mapping of experimental data to these diagrams and third, the afterwards ranking of pathways according to the number of matches.

Poster E-15

Accurate estimation of microarray gene expression levels by physical models

Yongqing Zhang (St. Jude Children Research Hospital); Antonio Ferreira (St. Jude Children Research Hospital); Cheng Cheng (St. Jude Children Research Hospital); Yongchun Wu (St. Jude Children Research Hospital); Jiong Zhang (St. Jude Children Research Hospital)

Abstract: We present a physical model based on the chemical equilibrium of hybridization and the kinetic process of washing and implemented it in a parallel generic simulated annealing algorithm. This model considers non-specific hybridization, probe affinity effects and the non-linear signal response and performs well on Affymetrix gene chip data.

Poster E-16

Sequence dependence of cross hybridization on short oligo microarrays

Chunlei Wu (University of Texas - M.D. Anderson Cancer Center); Roberto Carta (University of Central Florida); Li Zhang (University of Texas - M.D. Anderson Cancer Center)

Abstract: We report a free energy analysis of cross hybridization on short oligo microarrays. Our analysis revealed that cross-hybridization on the arrays are mostly caused by oligo fragments with a run of 10 to 13 nucleotides and the oligo fragments tend to bind to the 5' ends of the probes.

Poster E-17

Distribution of exonic splicing enhancer elements in human genes

Yongchun Wu (St. Jude Children's Research Hospital); Yongqing Zhang (St. Jude Children's Research Hospital); Jiong Zhang (St. Jude Children's Research Hospital)

Abstract: ESE motifs for four SR proteins are enriched in the region 80 to 120 bases away from the ends of splice acceptor sites. Significant enrichment of ESEs is associated with weak splice acceptor sites but not weak donor sites. ESEs are enriched in introns with weak donor or acceptor sites.

Poster E-18

Operon Prediction in Mycobacterium tuberculosis (MTB) from Gene Expression Data

Joel Beard (St Olaf College); Kristin Henry (St Olaf College); Sara Krohn (St Olaf College); Heather Wiste (St Olaf College); Paul Roback (St Olaf College); Robert Rutherford (St Olaf College)

Abstract: Around 2 million people die each year from TB. We will describe the use of Bayesian classification methods to predict probabilities for pairs of MTB genes being in an operon based upon intergenic distance and the expression correlation between two genes as established from 500 DNA microarray experiments.

Poster E-19

Microarray expression analysis and statistical methods comparison for caloric restriction in Emory mouse

Bing Liu (Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University); Fu Shang (Laboratory for Nutrition and Vision Research, Jean Mayer USDA HNRC on Aging at Tufts University); Allen Taylor (Laboratory for Nutrition and Vision Research, Jean Mayer USDA HNRC on Aging at Tufts University); Ina Hoeschele (Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University); Karen Duca (Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University)

Abstract: To test how different statistical methods perform when determining changes in gene expression that are associated with lifespan extension, we used Affymetrix Genechips to generate an expression profile of caloric-restricted and control-fed Emory mice livers. Several statistical methods were used and compared based on a number of criteria.

Poster E-20

ERIC - Enteropathogen Resource Integration Center, an NIAID Bioinformatics Resource Center for Biodefense and Emerging/Re-emerging Infectious Disease

Thomas Hampton (SRA International); Robin Martell (SRA International); Matthew Shaker (SRA International); Lorie Shaull (SRA International); Panna Shetty (SRA International); Mary Wong (SRA International); Guy Plunkett III (University of Wisconsin - Madison); Valerie Burland (University of Wisconsin - Madison); Eric Cabot (University of Wisconsin - Madison); David Bowen (University of Wisconsin - Madison); Jeremy Glasner (University of Wisconsin - Madison); Paul Liss (University of Wisconsin - Madison); Michael Rusch (University of Wisconsin - Madison); Frederick R. Blattner (University of Wisconsin - Madison); Nicole T. Perna (University of Wisconsin - Madison); John Greene (SRA International)

Abstract: ERIC (Enteropathogen Resource Information Center) is an NIAID Bioinformatics Resource Centers for Biodefense. ERIC is a JetSpeed portal to integrate information on the genome, annotations, and biological data for five enteropathogens: diarrheagenic E. coli, Shigella, Salmonella, Yersinia enterocolitica, and Yersinia pestis. Annotation is now online at <http://www.ericbrc.org>.

Poster E-21

A Noise Reduction Procedure for Gene Expression Data

Satwik Rajaram (Department of Physics, University of Illinois at Urbana-Champaign); Yoshihiro Taguchi (Department of Physics, Faculty of Science and Technology, and Institute for); Yoshitsugu Oono (Department of Physics and Institute for Genomic Biology, University of Illinois at Urbana-Champaign)

Abstract: Science and Technology, Chuo University The raw data obtained by microarray and SAGE experiments are often strongly influenced by various disturbances including statistical noise. We propose a method to extract stable relations among genes through eliminating uninformative/noise-infested genes based on our nonmetric multidimensional scaling method.

Poster E-22

A bottom-up approach to transcriptomes from the genome

Tomokazu Konishi (Faculty of Bioresource Sciences, Akita Prefectural University)

Abstract: A thermodynamic model that describes the formation of transcriptome by genomic information, as well as some experimental results that support appropriateness of the model, is introduced. The model will provide the bases for decoding quantitative information written in the genome and for integrating knowledge obtained in transcriptome studies.

Poster E-23

Conserved cis-regulatory regions in *Bacillus subtilis* operons under sulfur-limitation conditions

YeeLeng Yap (Department of microbiology, The University of HongKong); CheMan Chan (HKU-Pasteur Research Centre); Agnieszka Sekowska (Commissariat l'Energie Atomique); Antoine Danchin (Institute Pasteur)

Abstract: 113 differentially-expressed genes were identified from global transcription profiles of *Bacillus-subtilis* grown with various sulfur sources. These genes were unevenly distributed among functional groups. The cis-regulatory regions of these genes/operons were discovered based on the congruency of the existence of DNA upstream motifs and their regulated downstream gene expression profiles.

Poster E-24 (*There will also be an oral presentation of this poster.*)

Are antisense transcripts prone to A-to-I RNA editing ?

Yossef Neeman (Faculty of life sciences, Bar Ilan University); Dvir Dahary (Compugen Ltd.); Rotem Sorek (Compugen Ltd.); Eli Eisenberg (School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University)

Abstract: Recent studies have hypothesized that sense-antisense RNA transcript anneal together creating dsRNA duplexes, which undergo A-to-I RNA editing. We find that editing level in antisense regions is negligible relative to the recently reported widespread editing phenomenon. Our results may cast doubt on suggested antisense regulatory mechanisms.

Poster E-25

Comparison of computational methods for the identification of cell cycle regulated genes

Ulrik de Lichtenberg (Center for Biological Sequence Analysis); Lars J. Jensen (European Molecular Biology Laboratory); Anders Fausbll (Center for Biological Sequence Analysis); Thomas S. Jensen (Center for Biological Sequence Analysis); Peer Bork (European Molecular Biology Laboratory); Sren Brunak (Center for Biological Sequence Analysis)

Abstract: The *Saccharomyces cerevisiae* cell cycle microarray data have been subjected to a wide range of bioinformatics analysis methods. Here, we provide the first thorough benchmark of such methods, revealing that most new and more mathematically advanced methods actually perform worse than a novel, simple permutation-based method.

Poster E-26

Accuracy of cDNA microarray to detect gene expression changes induced by neuregulin on human breast epithelial cells

Bin Yao (MCBI Wayne State Node/Applied Genomics Technology Center, Wayne State University); Sanjay Rakhade (Center for Molecular Medicine and Genetics, Wayne State University); Qunfang Li (Barbara Ann Karmanos Cancer Institute); Sharlin Ahmed (Department of Neurology, Wayne State University); Raul Krauss (AlphaGene); Sorin Draghici (Department of Computer Science, Wayne State University); Jeffrey Loeb (Department of Neurology/ Center for Molecular Medicine and Genetics, Wayne State University, Wayne State University)

Abstract: We compared two experimental designs for cDNA arrays, dye-swap design and one we refer to as the "control correction method". Using both designs, we were able to measure small gene expression changes of transformed human breast epithelial cells treated with neuregulin with high fidelity based on

Northern blot confirmation.

Poster E-27

Machine learning approaches for predicting methylation sensitive CpG island sequences and disease-free interval in ovarian cancer

Henry Paik (Indiana University); Susan Wei (Ohio State University); Yoo-sung Kim (Indiana University); Sandya Liyanarachchi (Ohio State University); Lang Li (Indiana University); Curtis Balch (Indiana University); Ramana Davuluri (Ohio State University); Sun Kim (Indiana University); Tim Huang (Ohio State University); Kenneth Nephew (Indiana University)

Abstract: Differential methylation hybridization (DMH) is an effective technique for monitoring DNA methylation profiles of tumors. Using DMH profiles from 19 ovarian tumors, subdivided into high- / low-methylation subgroups, we utilized supervised machine-learning techniques to successfully construct classifiers to discriminate long- / short-term survival with over 90% accuracy.

Poster E-28

Methodologies for Intra and Inter-Species Comparisons of Yeast Transcriptome States

Galle Lelandais (Ecole Normale Suprieure)

Abstract: Distinguishing the similar from the dissimilar in large-scale datasets, comparative analyses of several transcriptome states promises to improve fundamental understanding of both the universality and the specialization of biological mechanisms. We developed bioinformatic approaches to compare intra and inter-species genome expression programs operating in response to environmental changes.

Poster E-29

Genotyping of Campylobacter jejuni strains and analysis of methods for identifying absent or highly divergent genes

Ozan Gundogdu (London School of Hygiene & Tropical Medicine); Olivia Champion (London School of Hygiene & Tropical Medicine); Nick Dorrell (London School of Hygiene & Tropical Medicine); Brendan Wren (London School of Hygiene & Tropical Medicine)

Abstract: Comparative genomics techniques were used to study the food borne pathogen Campylobacter jejuni with the aim of identifying host specific genetic determinants. The efficacy of methods used for selecting "absent or highly divergent" genes was also investigated.

Poster E-30

PASA-2: The Next Generation of the PASA Pipeline, Applied Towards Automated Incorporation of Full-length cDNAs and ESTs into Gene Structure Annotations

Brian Haas (The Institute for Genomic Research); Jennifer Wortman (The Institute for Genomic Research); Aihui Wang (The Institute for Genomic Research); C. Robin Buell (The Institute for Genomic Research)

Abstract: The PASA-2 pipeline consists of software that incorporates expressed transcript alignments into gene structure annotations. Annotation updates include the creation of new gene models, modifications to existing gene structures, additions of alternative splicing variants, identification of polyadenylation sites, antisense transcripts, and complex annotation updates including gene merges and splits.

Poster E-31

Biophysical Modeling of Microarray Data

Hrishikesh Deshmukh (George Mason University/School of Computational Sciences/Bioinformatics); Jennifer Weller (George Mason University/School of Computational Sciences/Bioinformatics)

Abstract: Microarray signal is often misinterpreted as direct indications of concentration. Biophysical properties are not included in analyses. Affymetrix GeneChip probes (PM, MM) in linear range are used to model secondary structures of probes after filtering probes for cross-hybridization, SNPs presence. Biophysical properties account for signal differences for some probes.

Poster E-32

In Silico Identification of Transcription Factors Involved in Cellular Differentiation by Promoter Analysis of Co-regulated Genes

Peter Hoen (Center for Human and Clinical Genetics, Leiden University Medical Center); Jun Hou (Center for Human and Clinical Genetics, Leiden University Medical Center); Renee de Menezes (Center for Human and Clinical Genetics, Leiden University Medical Center); Ellen Sterrenburg (Center for Human and Clinical Genetics, Leiden University Medical Center); Johan den Dunnen (Center for Human and Clinical Genetics, Leiden University Medical Center); Gert-Jan van Ommen (Center for Human and Clinical Genetics, Leiden University Medical Center); Peter Taschner (Center for Human and Clinical Genetics, Leiden University Medical Center)

Abstract: We aimed to identify transcription factors responsible for differential gene expression during differentiation of human primary skeletal myoblasts by comparison of co-regulated gene promoters with transcription factor binding sites. Those promoters were significantly enriched in myogenic regulatory elements compared to control promoter sequences or promoters from genes with unchanged expression.

Poster E-33

Genomic analysis of RNA polymerase III terminators

Riccardo Percudani (Universit di Parma); Priscilla Braglia (Universit di Parma); Giorgio Dieci (Universit di Parma)

Abstract: A simple run of four thymines on the sense strand is believed to provoke efficient termination of RNAP III transcription in all organisms. We present here a more complex picture of RNAP III termination emerging from the statistical analysis of 3'-flanking regions of tRNA sequences in various eukaryotic genomes

Poster E-34

Leptin Action in Adipose Tissue: The Identification of Potential Efficacy Biomarkers Facilitated by The Pathway Prioritizer, a Novel Computational Algorithm.

Qing Liu (Gene Logic Inc.); Kory Johnson (Gene Logic Inc.); Prakash Kulkarni (Gene Logic Inc.); Lawrence Mertz (Gene Logic Inc.)

Abstract: We have investigated leptin action on gene expression in the adipose tissue in leptin-deficient ob/ob mice using microarray data and the Pathway Prioritizer, a novel algorithm used to identify key regulated biological pathways. Using these tools, we have identified potential efficacy biomarkers of leptin action in adipose tissue.

Poster E-35

Identifying biclusters by genetic algorithm

Hua-Sheng Chi (Department of Computer Science and Information Engineer-ing, National Taiwan University); Tao-Wei Huang (Department of Computer Science and Information Engineer-ing, National Taiwan University); Cheng-Yan Kao (Department of Computer Science and Information Engineer-ing, National Taiwan University)

Abstract: We propose a novel model and Biclustering by Iterative Genetic Algorithm (BIGA) to identifying condition-specific transcriptional modules, biclusters. The experimental results have shown the effectiveness and correctness of proposed approach. Comparisons among other methods are also indicated that our method associated with better performance in biclustering gene expression data.

Poster E-36

Visual Methods for Statistical Analysis of Microarray Clusters

Matthew Hibbs (Princeton University, Computer Science Department, Lewis-Sigler Institute for Integrative Genomics); Nathaniel Dirksen (Princeton University, Computer Science Department); Kai Li (Princeton University, Computer Science Department); Olga Troyanskaya (Princeton University, Computer Science Department, Lewis-Sigler Institute for Integrative Genomics)

Abstract: Appropriate visualization methods can complement existing numerical methods in the challenging task of verification of the results of clustering algorithms on microarray datasets. We present several techniques for visualization-based analysis of microarray clusters in a noise-robust manner to assess cluster quality, identify outliers, and examine cluster relationships.

Poster E-37

Improved MCC for Clustering Of Gene Expression Data

Youping Deng (Department of Biological Sciences, University of Southern Mississippi, Hattiesburg); Venkatachalam Chokalingam (Department of Industrial Engineering, Arizona State University, Tempe); Chaoyang Zhang (Department of Computer Science, University of Southern Mississippi, Hattiesburg); Mohamed O.Elasri (Department of Biological Sciences, University of Southern Mississippi, Hattiesburg)

Abstract: We introduce clustering of gene expression data based on the correlation of one-step Markov chain transition probabilities. The behavior of genes at each step of the experiment is taken into consideration and the series that have similar behavior at each point are grouped together using agglomerative clustering.

Poster E-38

Identifying Shared Functional Binding Site Modules In Sets Of Orthologs OR Co-expressed Genes

Hailong Meng (University of Florida); Arunava Banerjee (University of Florida); Lei Zhou (University of Florida)

Abstract: The goal of our project is to research and develop a methodology for identifying shared cis-regulatory modules (CRMs) in sets of orthologs or co-expressed genes. Our approach focuses on identifying shared CRMs among the sequences that have overall low similarity at the DNA level

Poster E-39

Genome-based analysis of expression microarray data - an application to alternative mRNA processing

Ann Loraine (Section on Statistical Genetics, University of Alabama); Xiangqin Cui (Section on Statistical Genetics, University of Alabama); Gary Churchill (The Jackson Laboratory)

Abstract: We have developed a genome-based analysis method that uses data from expression microarray experiments (Affymetrix platform) to investigate alternative mRNA processing, especially polyadenylation site choice. This approach revealed numerous examples of probe set groups whose expression patterns suggest strain-dependent differential mRNA processing in mouse.

Poster E-40

The detection of novel alternative promoters in the human genome and their importance in various cancers

Gregory Singer (The Ohio State University); Kristi Bennett (The Ohio State University); Christoph Plass (The Ohio State University); Ramana Davuluri (The Ohio State University)

Abstract: We scanned for novel promoters in cancer-related genes, yielding dozens of previously unknown first exons/promoters. Out of five test genes, four were confirmed to have novel first exons via RT-PCR. These previously unknown promoters may play a role in the formation of cancers associated with these genes.

Poster E-41

Testing Heuristics Concerning Sequence and Expression Similarity

Brent Hughes (The University of Akron); Zhong-Hui Duan (The University of Akron)

Abstract: The soundness of the following heuristic was tested: a set of highly similar sequence pairs will have similar changes in their protein expression levels under an experimental condition. A novel measure of sequence similarity was developed. By testing this measure statistically against expression similarity, we

determined the heuristic's overall reliability.

Poster E-42 (*There will also be an oral presentation of this poster.*)

Feasibility of genome-wide recognition of mutant single nucleotide polymorphisms (SNPs) with effects on constitutive mRNA splicing

Peter Rogan (Children's Mercy Hospital and Clinics, University of Missouri-Kansas City); Vijay Nalla (Children's Mercy Hospital and Clinics, University of Missouri-Kansas City)

Abstract: To assess the feasibility of predicting the effects of SNPs on mRNA splicing, we evaluated SNP-induced changes in individual information within the known boundaries of all genes and mapped transcripts on chromosome 22. Deleterious SNP alleles predicted to alter constitutive mRNA splicing occur in 3.5% of these genes.

Poster E-43

Transcriptome Analysis Tools: Visualization and Management of Ultra-High Volume of DNA Sequence Data

Irina Khrebtukova (Lynx Therapeutics, Inc., Hayward, CA); Christian D. Haudenschild (Lynx Therapeutics, Inc., Hayward, CA); Daixing Zhou (Lynx Therapeutics, Inc., Hayward, CA.); William Nelson (Lynx Therapeutics, Inc., Hayward, CA.); Selene M. Virk (Lynx Therapeutics, Inc., Hayward, CA.); Maria Johnson (Lynx Therapeutics, Inc., Hayward, CA.); Keith Moon (Lynx Therapeutics, Inc., Hayward, CA.); Thomas Vasicek (Lynx Therapeutics, Inc., Hayward, CA.)

Abstract: Massively Parallel Signature Sequencing (MPSS) is an extremely efficient method for generating short DNA sequences and the ideal technology for establishing reference transcriptome databases. This presentation will describe our pipeline for data processing and the genome browser developed for viewing short sequence data in the genome and transcriptome context.

Poster E-44

IslandPath 2: web application aiding integrated analysis of genomic islands

William Hsiao (Department of Molecular Biology and Biochemistry, Simon Fraser University); Parmit Chilana (Simon Fraser University); Amber Fedynak (Department of Molecular Biology and Biochemistry, Simon Fraser University); Fiona Brinkman (Department of Molecular Biology and Biochemistry, Simon Fraser University)

Abstract: Genomic islands (GIs) are frequently associated with adaptations of medically, agriculturally or environmentally important microbes. We will present a re-engineered version of our previously published GI identification and analysis software, Islandpath and some results from our analyses of GIs. New features of IslandPath-2 include incorporated GBrowse view and multiple data-tracks.

Poster E-45

Enhancing Searches for Molecular Markers: Normalization and Linear Modeling Improve the Sensitivity of ChIP-Chip Studies

Paul Boutros (Department of Medical Biophysics, University of Toronto); Romina Ponzielli (Division of Cellular and Molecular Biology, University Health Network); Igor Jurisica (Department of Medical Biophysics, University of Toronto); Linda Penn (Department of Medical Biophysics, University of Toronto)

Abstract: ChIP-Chip is a high-throughput technique for detecting transcription-factor binding. A 54-array validation study has allowed us to quantify the effects of different antibodies and microarray platforms on ChIP-Chip data. This dataset has also been used to test a panel of normalization and filtering algorithms for sensitivity and selectivity.

Poster E-46

Clustering through Transductive learning for Personalized Modeling

Liang Goh (Knowledge Engineering & Discovery Research Institute); Nikola Kasabov (Knowledge Engineering & Discovery Research Institute)

Abstract: A new approach to clustering by transductive learning to seek out large bi-clusters of vectors in local sub-space of the problem space and then uses the information elicited to infer the significance of the bi-cluster. It is a powerful way to use existing information to elicit better, local clusters and models.

Poster E-47

Discovering Protein-Protein Interaction via Domain-Domain Interactions

Sarah Javaid (The Ohio State University); Jingchun Chen (The Ohio State University); Fa Zhang (The Ohio State University); Hatice Gulcin Ozer (The Ohio State University); Bo Yuan (The Ohio State University)

Abstract: Protein-protein interactions are supposed to be mediated via conserved protein domains. However, existing protein interaction information is largely sequence based. We used the Maximum Likelihood Estimation method to infer which domains are involved in actual interactions. We were able to determine known and unknown domain interactions via this method.

Poster E-48

Multi-copy genes in the genomes

Seung Hoon Baek (Dept of Biochemistry, Han-yang University, Ansan, South Korea); Soo Young Cho (Dept of Biochemistry, Han-yang University, Ansan, South Korea); Young Seek Lee (Dept of Biochemistry, Han-yang University, Ansan, South Korea)

Abstract: We have found that many genes are present more than one copy on two different chromosomes or two different local in the same chromosome. To analyze function and evolutionary relationship of multi copy genes we classified those genes by using GO, KEGG, and COG in 4 species.

Poster E-49

ASePCR: Alternative splicing electronic RT-PCR in multiple tissues and organs

Namshin Kim (Ewha Womans University); Dajeong Lim (Seoul National University); Sanghyuk Lee (Ewha Womans University); Heebal Kim (Seoul National University)

Abstract: ASePCR is a web-based application that estimates the amplicon size for a given primer pair based on the transcript models identified by NCBI reverse e-PCR. ASePCR (<http://genome.ewha.ac.kr/ASePCR/>) supports the transcriptome models of RefSeq, Ensembl, ECGene and AceView for human, mouse, rat, and chicken.

Poster E-51

AffyMAPSDetector: A Tool To Detect SNPs In Affymetrix GeneChipTM Expression Arrays

Sunita Kumari (George Mason University, VA); Lalit Verma (Celera Genomics, MD); Jennifer Weller (George Mason University, VA)

Abstract: AffyMAPSDetector is a computational tool that assists in preprocessing Affymetrix GeneChipTM data. It can be used to characterize the potential contribution of SNPs, especially when a mismatch probe signal is higher than the corresponding perfectmatch signal. Identifying the presence of SNPs will improve post-processing and subsequent data-analysis.

Poster E-52

Heterologous Expression of Alpha-Galactosidases from Human Pathogen Aspergillus fumigatus

Betul Soyler (Department of Food Engineering, Middle East Technical University); Peter Biely (Institute of Chemistry, Slovak Academy of Sciences); Zumrut Ogel (Department of Food Engineering, Middle East Technical University)

Abstract: Previously, two alpha-galactosidase genes with different characteristics were cloned from *Aspergillus fumigatus*. After the completion of the *A. fumigatus* genome project, regions with homology to alpha-galactosidase genes were searched. The aim of this study is to express alpha-galactosidases of *A. fumigatus* in a safe host, *Aspergillus sojae*.

Poster E-53

Molecular Phenotypic Profiling of Clinical Features in Lung Cancer by Independent Component Analysis

Mi Hyeon Kim (Seoul National University Biomedical Informatics(SNUBI)); Mi-Ryung Han (Seoul National University Biomedical Informatics(SNUBI)); Ju Han Kim (Seoul Nat'l University College of Medicine)

Abstract: To identify relevant biological information, we analyzed cDNA microarray data from lung cancer patients using Independent Component Analysis (ICA). We associated resulting ICs with clinical data and revealed the difference between recurrence in each age groups by comparing genes with extreme (negative or positive) loading values of independent components.

Poster E-54

Model-based clustering of microarray expression profiles for identification of coordinately controlled gene clusters

Tra Vu (Inst. of Microbiology); Josef Panek (Inst. of Microbiology); Jiri Vohradsky (Inst. of Microbiology)

Abstract: We present new approach to identification of coordinately controlled gene clusters combining feature selection using singular value decomposition of time series data and a recurrent genetic network model-based clustering method. The approach has been applied to identification of coordinately controlled and successive gene clusters in the genome of eubacterium *S. coelicolor*.

Poster E-55

Igor: An integrated automated Affymetrix Microarray data Analysis and Knowledge Management System

Steven Osselaer (Johnson & Johnson Pharmaceutical Research and Development); Hinrich Goehlmann (Johnson & Johnson Pharmaceutical Research and Development); Luc Bijnens (Johnson & Johnson Pharmaceutical Research and Development); An De Bondt (Johnson & Johnson Pharmaceutical Research and Development); Rudi Verbeeck (Johnson & Johnson Pharmaceutical Research and Development); Andrew Stubbs (Johnson & Johnson Pharmaceutical Research and Development)

Abstract: To accelerate microarray data analysis, an automated pipeline with HTML result publication, was developed. QC modules allow for sample quality assessment. Data analysis modules (SAM, PAM, Q-value, spectralmap) result in annotated hyperlinked lists of significant genes. Results are stored in relational databases and the functionality can be used interactively.

Poster E-56

Gene Expression and Internalization of Mammalian Cell Entry (Mce) Proteins Encoded by mce3 Operon of Mycobacterium tuberculosis by Mammalian Cells

Sherief El-Shazly (Faculty of Medicine, Kuwait University.); Suhail Ahmad (Faculty of Medicine, Kuwait University.); Abu Salim Mustafa (Faculty of Medicine, Kuwait University.); Raja Al-Attiyah (Faculty of Medicine, Kuwait University.); Dimitrolos Krajci (Faculty of Medicine, Kuwait University.)

Abstract: This study aimed to demonstrate the expression of adhesin/invasin-like Mce3A-F proteins during in vivo and in vitro growth of Mycobacterium tuberculosis and their role in the internalization by mammalian cells by following the functional genomics approach, a process central to the molecular and immunological pathogenesis of tuberculosis.

Poster E-57

Gene Expression biclustering by Sparse Non-negative Matrix Factorization

Pedro Carmona-Saez (National Center of Biotechnology); Roberto Pascual-Marqui (The KEY Institute for Brain-Mind Research, University Hospital of Psychiatry); Francisco Tirado (Computer Architecture Department. Universidad Complutense de Madrid); Jose M. Carazo (National Center of Biotechnology); Alberto Pascual-Montano (Computer Architecture Department. Universidad Complutense de Madrid)

Abstract: In the present work we show the potential of a new data mining technique: Sparse Non-Negative Matrix Factorization, for biclustering analysis of gene expression data. This technique was applied to uncover the main gene expression modules associated to clusters of samples.

Poster E-58

Fast Fourier Transform Clustering (FFTC)

Gek Huey Chua (Research Associate (BioInformatics Institute)); Tong Seng Lim (PhD Student (BioInformatics Institute)); Siew Woh Choo (Research Associate (BioInformatics Institute)); Win King Sung (assistant professor (School of Computing, NUS))

Abstract: Fast Fourier Transform Clustering (FFTC aims to identify time-shifted and inverted expression profiles, which are not detected using conventional clustering approaches. The information is useful to decipher potential biological relationships, such as activation or inhibition. Besides, FFTC can identify co-expressed genes and predict the function of unknown genes.

Poster E-59 (*There will also be an oral presentation of this poster.*)

Transcriptional reprogramming in genetic backup circuits

ron kafri (Weizmann Institute of Science); tzachi pilpel (Weizmann Institute of Science)

Abstract: Gene duplicates have been proposed to function as backups against mutations, thus buffering the phenotype against genomic variations. Analyzing the yeast transcriptional regulatory network we provide evidence suggesting the existence of a fine tuned regulatory control that activates, and up-regulates paralogs in response to their partners' deletion.

Poster E-60

Computer approach in Disease Diagnosis

vinod kumar (Bioinformatics Institute of India); Dr. Vimrash Raina (Indraprashta Apollo Hospitals)

Abstract: Computer approach in Disease Diagnosis, Future methods of Disease diagnosis

Poster E-61

Sources of variation and reproducibility of microarray experiments

Stanislav Zakharkin (University of Alabama at Birmingham); Kyoungmi Kim (University of Alabama at Birmingham); Tapan Mehta (University of Alabama at Birmingham); Lang Chen (University of Alabama at Birmingham); Stephen Barnes (University of Alabama at Birmingham); Katherine Scheirer (University of Alabama at Birmingham); Rudolph Parrish (University of Louisville); David Allison (University of Alabama at Birmingham); Grier Page (University of Alabama at Birmingham)

Abstract: We estimated relative contributions of different sources of variation in Affymetrix microarray experiments and evaluated their reproducibility. The greatest source of variation was biological variability followed labeling variability and residual error. The hybridization and daily variability were significant while scanning order was not. Different image processing algorithms gave comparable estimates.

Poster E-62

Hierarchical Probes and Primers Design System for Microbial Identification

Junhyung Park (Busan Genome Center, College of Medicine, Pusan National University, Busan); Heekyung Park (Busan Genome Center, College of Medicine, Pusan National University, Busan); Eunsil Song (Institute for Genomic Medicine, GeneIn. Co., Ltd, Busan); Hyunjung Jang (Institute for Genomic Medicine, GeneIn. Co., Ltd, Busan); Byeongchul Kang (Division of Applied Bioengineering, Dongseo University, Busan); Seungwon Lee (Busan Genome Center, College of Medicine, Pusan National University, Busan); Hyunjin Kim (Busan Genome Center, College of Medicine, Pusan National University, Busan); Cheolmin Kim (Busan Genome Center, College of Medicine, Pusan National University, Busan)

Abstract: We have developed a hierarchical primer design system, MiProbe, for universal-, genus-, and species-specific primer. It is accessed by the web interface, a collection of perl scripts and running on the MySQL. MiProbe consists of sequence collection, multiple alignment, candidate primer finding, and

BLAST system.

Poster E-63

A Model-Based Scan Statistic For Identifying Chromosomal Patterns of SNP Effects

Yan Sun (Department of Epidemiology, University of Michigan); Albert Levin (Department of Human Genetics, University of Michigan); Eric Boerwinkle (Department of Human Genetics, University of Texas Health Sciences); Henry Roberston (Department of Biostatistics, University of Michigan); Julie Douglas (Department of Human Genetics, University of Michigan); Sharon Kardia (Department of Epidemiology, University of Michigan)

Abstract: We developed a model-based scan statistic that takes into account the complex landscape of the human genome and its variations to identify chromosomal regions with SNP effects. We compared this method with a sliding-window approach and identified regions which were highly associated with hypertension on human chromosome 19.

Poster E-64

Analyzing gene expression profiling data of neuroblastomas with signal transduction networks

Rainer Knig (Division Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg); Gunnar Schramm (Division Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg); Frank Westermann (Division Tumour Genetics, German Cancer Research Center (DKFZ), Heidelberg); Marcus Oswald (Institute of Computer Science, University of Heidelberg); Sebastian Sager (Interdisciplinary Center for Scientific Computing, University of Heidelberg); Andre Oberthr (Cologne Children's Hospital, Department of Pediatric Oncology and Hematology, University of Cologne); Matthias Fischer (Cologne Children's Hospital, Department of Pediatric Oncology and Hematology, University of Cologne); Benedikt Brors (Division Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg); Gerhard Reinelt (Institute of Computer Science, University of Heidelberg); Roland Eils (Division Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg)

Abstract: Applying literature-based and manually curated databases like Transpath (Biobase, Germany), we reconstruct the human signal transduction network. Gene-expression profiles of pediatric neuroblastomas with good and bad prognosis define interaction strengths of neighbouring molecules and are used to compare the topology of the respective networks.

Poster E-65

GenoSIS: the Application of Spatial Genomics

Mary E. Dolan (Mouse Genome Informatics, The Jackson Laboratory); Constance Holden (National Center for Geographic Information and Analysis, University of Maine); M. Kate Beard (National Center for Geographic Information and Analysis, University of Maine); Carol J. Bult (Mouse Genome Informatics, The Jackson Laboratory)

Abstract: In GenoSIS, we have adapted the concepts and tools of geographic information science to allow dynamic integration, querying, analysis and visualization of diverse genomic data. Using the layered map paradigm of geographic information systems, we have applied GenoSIS to explore the spatial organization of genomic features of the laboratory mouse.

Poster E-66

GeneDecks: A Systems Biology Facilitator With Combinatorial Genecards Outlook

Shany Ron (The Weizmann Institute of Science); Liora Strichman-Almashanu (The Weizmann Institute of Science); Michael Shmoish (The Weizmann Institute of Science); Asaf Madi (The Weizmann Institute of Science); Alexandra Sirota (The Weizmann Institute of Science); Karin Noy (The Weizmann Institute of Science); Naomi Rosen (The Weizmann Institute of Science); Orit Shmueli (The Weizmann Institute of Science); Marilyn Safran (The Weizmann Institute of Science); Doron Lancet (The Weizmann Institute of Science)

Abstract: We introduce a novel tool, GeneDecks, for exploiting the rich and varied integrated dataset of

human genes in GeneCards, as well as uncovering assorted gene relationships. Given a gene of interest, GeneDecks provides sets of associated genes that are similar with respect to combinatorial selected annotation categories.

Poster E-67

Protocols for the Assurance of Microarray Data Quality and Process Control

Lyle Burgoon (Michigan State University); Jeanette Eckel-Passow (Mayo Clinic Cancer Center); Chris Gennings (Virginia Commonwealth University); Darrell Boverhof (Michigan State University); Jeremy Burt (Michigan State University); Cora Fong (Michigan State University); Tim Zacharewski (Michigan State university)

Abstract: A comprehensive quality control protocol for cDNA microarrays was developed using in vivo and in vitro dose-response and time-course studies. The protocol combines: 1) diagnostic plots identifying feature misalignments, 2) boxplots to monitor microarray intensity distributions, and 3) a support vector machine (SVM) model to classify microarray quality.

Poster E-68

Analysis of Genomic Alterations in Tumor Samples with Affymetrix 100K SNP chips

Yuri Kotliarov (Neuro-Oncology Branch, National Cancer Institute, National Institutes of Health); Neil Christopher (Neuro-Oncology Branch, National Cancer Institute, National Institutes of Health); Mary Steed (Neuro-Oncology Branch, National Cancer Institute, National Institutes of Health); Jennifer Walling (Neuro-Oncology Branch, National Cancer Institute, National Institutes of Health); Angela Center (Neuro-Oncology Branch, National Cancer Institute, National Institutes of Health); Oliver Bogler (Hermelin Brain Tumor Center, Depts. Neurology & Neurosurgery, Henry Ford Hospital); Tom Mikkelsen (Hermelin Brain Tumor Center, Depts. Neurology & Neurosurgery, Henry Ford Hospital); Jean Zenklusen (Neuro-Oncology Branch, National Cancer Institute, National Institutes of Health); Howard Fine (Neuro-Oncology Branch, National Cancer Institute, National Institutes of Health)

Abstract: Loss of Heterozygosity (LOH) and copy number alterations were analyzed for 173 glioma tumor and 20 normal samples in order to develop standards of quality control and analysis techniques. Various algorithms for copy number smoothing were evaluated and optimized for use with Affymetrix 100K data.

Poster E-69

A multidisciplinary functional genomics strategy to identify treatment-related disease specific genes: The model of multiple sclerosis

Michael Gurevich (Multiple Sclerosis Center, Sheba Medical Center, Tel-Hashomer and Sackler School of Medicine, Tel-Aviv University); Yair Snir (Multiple Sclerosis Center, Sheba Medical Center, Tel-Hashomer and Sackler School of Medicine, Tel-Aviv University); Anat Achiron (Multiple Sclerosis Center, Sheba Medical Center, Tel-Hashomer and Sackler School of Medicine, Tel-Aviv University)

Abstract: We present a multidisciplinary functional genomics strategy to identify treatment-related disease specific genes. This strategy combined in-vivo and in-vitro gene-expression data of peripheral blood mononuclear cells and novel computational analyses methods. Our strategy was applied to analysis of interferon beta1-a treatment-effects in multiple sclerosis.

Poster E-70

Classification and model averaging: first stages in developing predictive models for disease onset.

Robert Podolsky (Medical College of Georgia); Christin Collins (Medical College of Georgia); Jin-Xiong She (Medical College of Georgia)

Abstract: We present and compare methods for identifying classification models during the development of prognostic biomarkers based on genomic and proteomic data. Our methods involve identifying and averaging that provide true prognostic ability.

Poster E-71

Gene Ontology Term Enrichment from Microarray Studies Using a Hyperbolic Viewer Facilitates Comparison Between Experiment Groups

Ernst Dow (Eli Lilly and Company); Amar Kumar (Eli Lilly and Company); Seppo Karrila (Lilly Systems Biology); Scott McAhren (Eli Lilly and Company); Adam West (Eli Lilly and Company); Nalini Kulkarni (Eli Lilly and Company); Ramneek Gupta (Lilly Systems Biology); Mahesh Guzuva Desikan (Lilly Systems Biology); Rakhi Bhat (Lilly Systems Biology); Nicholas Lewin-Koh (Eli Lilly and Company); Trent Stewart (Eli Lilly and Company USA); Jianyong Shou (Eli Lilly and Company USA); Sudhanshu Patwardhan (Bioinformatics Institute A*STAR Singapore); John Calley (Eli Lilly and Company USA); Jude Onyia (Eli Lilly and Company USA)

Abstract: Interpreting a list of differentially expressed genes from an array experiment is tedious and subjective. Interactive, multi-experiment hyperbolic views of enriched Gene Ontology terms allow rapid identification of key terms across experiments. Further, enrichment is relatively insensitive to the size of the gene list, reducing the likelihood of misinterpretation.

Poster E-72

A Two-dimensional Regression Tree Approach to the Modeling of Gene Expression Regulations

Jianhua Ruan (Dept. of Computer Science and Engineering, Washington University in St Louis); Weixiong Zhang (Dept. of Computer Science and Engineering and Dept. of Genetics, Washington University in St Louis)

Abstract: BDTree is a novel method for modeling the expression levels of multiple genes under multiple conditions, using TFBMs on gene promoters and the expression levels of TFs as predictors. The model provides hypotheses about condition-specific gene regulations, and can predict the expression levels of new genes under new conditions.

Poster E-73

Identification and Evaluation of Transcription Factor Binding Sites Responsive to PEG-IFN-alpha in PBMC in vitro

Yunlong Liu (Department of Biochemistry and Molecular Biology, Indiana University School of Medicine); Milton Taylor (Department of Biology, Indiana University); Howard Edenberg (Department of Biochemistry and Molecular Biology, Indiana University School of Medicine)

Abstract: A model-based procedure was developed to identify and evaluate transcription-factor binding sites from global gene expression data and genomic sequences. Using interferon-responsive genes in peripheral blood monocytes as a model system, we identified 15 positive binding sites and 5 negative binding sites in the interferon stimulation.

Poster E-74

Identification and analysis of Pescadillo homologues during a pathogenomics survey in Leishmania spp.

Elton J.R. Vasconcelos (Ncleo de Genmica e Bioinformtica, Universidade Estadual do Ceara); Joao J.S. Gouveia (Ncleo de Genmica e Bioinformtica, Universidade Estadual do Ceara); Ana C.L. Pacheco (Ncleo de Genmica e Bioinformtica, Universidade Estadual do Ceara); Diana M. Oliveira (Ncleo de Genmica e Bioinformtica, Universidade Estadual do Ceara)

Abstract: Analysis of predicted pescadillo homologues in *Leishmania* spp., identified during a pathogenomics survey, showed that the protein is well conserved among a variety of species (from 32 to 79% sequence homology with yeast, human, zebrafish, plasmodial and trypanosomal homologues) and contains intriguing structural motifs to parasite development and infectivity.

Poster E-75

PACdb: PolyA Cleavage Site and 3'-UTR Database

Priyam Singh (Boston University); Michael Brockman (Boston University); Donglin Liu (The Jackson Laboratory); Sean Quinlan (Boston University); Jesse Salisbury (University of Maine); Joel Graber (The Jackson Laboratory)

Abstract: The "PolyA Cleavage Site and 3'-UTR Database" (PACdb) is a web-accessible database that catalogs putative 3'-processing sites for multiple organisms. Using EST-genome alignments, we have identified and characterized the specificity and heterogeneity of 3'-processing across a broad range of genomes, including animals, plants, and fungi.

Poster E-76

Automated validation of polymerase chain reactions using amplicon melting curves

Tobias Mann (Dept. of Genome Sciences, University of Washington); Richard Humbert (Regulome); John Stamatoyannopolous (Regulome); William Noble (Dept. of Genome Sciences, University of Washington)

Abstract: We describe a robust solution to the problem of validating individual PCR reactions. Using a training set of 1,728 manually screened human genomic PCR amplicons, we developed a support vector machine classifier capable of discriminating single-product PCR reactions with better than 99% accuracy using features computed from amplicon melting curves.

Poster E-77

Performance Assessment and Integration of Oligonucleotide Microarray Design Software

Raad Gharaibeh (Virginia Polytechnic Institute and State University); Cynthia Gibas (Virginia Polytechnic Institute and State University)

Abstract: MODIT is a pipeline that merges the output from two oligo design programs ArrayOligoSelector and OligoArray, eliminates redundant and suboptimal probes, and applies uniform evaluation criteria to pick a set of optimal probes that are unique in sequence, uniform in T_m and free of stable secondary structure.

Poster E-78

Using fuzzy logic algorithm and gene expression database ONCOMINE in COPD outcome forecasting

Boris Shilov (Siberian State Medical University); Ekaterina Bukreeva (Siberian State Medical University); Natalya Shilova (Siberian State Medical University)

Abstract: Investigation included survey of 37 patients with COPD infectious exacerbation. In conditions of similar pathomorphological changes they had various genes expression level, there are: AT, 2M, MMP. Use the fuzzy logic algorithm has allowed patients classifying on the gene expression basis in groups with different forecast of the disease upshot.

Poster E-79

MAMA — A Meta-Analysis Platform for Expression Profiling of Cancer Tissues

Zhe Zhang (Dept. of Biomedical Informatics, Univ. of North Carolina at Chapel Hill); David Fenstermacher (Biomedical Informatics Facility, Univ. of Pennsylvania)

Abstract: MAMA, a data-mining platform supporting analysis of microarray data across multiple studies, is aimed to improve precise expression profiling of cancer tissues. It stores microarray datasets in a server-side database and allows users to investigate stored data with an application, which implements statistical methods like meta-analysis.

Poster E-80

An Extensible Automated System For Genomic Study of Bacteria

Ying Fong (McMaster University); Brian Golding (McMaster University)

Abstract: An automated system for processing and management of data generated from random library sequences, gene fusion expression experiment and microarray experiment in a high-through-put environment. Function prediction of gene is based on integration of experimental data and Nearest

Neighbor Genes Hypothesis-A computational approach to predict gene function.

Poster E-81

Use of Exact-Match Methodology for Focussing the Functional Comparison of Bacterial Genomes

Wilfred Cuff (Public Health Agency of Canada); Raymond Tsang (Public Health Agency of Canada)

Abstract: Exact Matches+Structural Variables+MUMmer+Neisseria meningitidis Comparative genomics papers concentrate quickly on functional information (genes, proteins, metabolic pathways). The complexity of resulting papers risks scientific verification of conclusions and tends to exclude small group participation. An effort was made to use simpler structural variables to help focus functional studies.

Poster E-82

GEMAT: Genomic Experiment Management and Analysis Tool

Venkatarajan Mathura (Roskamp Institute); Deepak Kolippakkam (Roskamp Institute); Fiona Crawford (Roskamp Institute); Michael Mullan (Roskamp Institute)

Abstract: GEMAT is a web-based software tool that facilitates data management in microarray experiments performed in multiple platforms. Gene mining/analysis tools, hyperlinked annotation and automatic web-based content delivery are available. Customizable Biological Pathway Visualization (BPV) tool can represent simple networks and allow researchers to overlay their gene expression data.

Poster E-83

Prediction of Functional Modules Based on Comparative Genome Analysis and Gene Ontology Application

Hongwei Wu (Computational Biology Institute, Oak Ridge National Laboratory); Zhengchang Su (Computational Biology Institute); Fenglou Mao (Department of Biochemistry and Molecular Biology, The University of Georgia); Victor Olman (Department of Biochemistry and Molecular Biology, The University of Georgia); Ying Xu (Department of Biochemistry and Molecular Biology, The University of Georgia)

Abstract: A computational framework has been developed to predict modules in microbial organisms through identifying "linkage clusters" of genes that are predicted to be functionally linked. The functional linkages are predicted by combining the comparative genome analysis and GO annotations using the Bayesian inference. For E. coli K12, 185 functional modules are predicted that are highly consistent with databases.

Poster E-84

Using GEMS for Cancer Diagnosis and Biomarker Discovery from Microarray Gene Expression Data

Alexander Statnikov (Discovery Systems Laboratory, Department of Biomedical Informatics, Vanderbilt University); Ioannis Tsamardinos (Discovery Systems Laboratory, Department of Biomedical Informatics, Vanderbilt University); Constantin Aliferis (Discovery Systems Laboratory, Department of Biomedical Informatics, Vanderbilt University)

Abstract: We introduce a system GEMS for automated development and evaluation of high-quality cancer diagnostic models and biomarker discovery from microarray gene expression data. The development of GEMS was informed by a rigorous algorithmic evaluation. The system is freely available for non-commercial use from <http://www.gems-system.org>

Poster E-85

Hybridization prediction model for cDNA microarrays

Yian A. Chen (Medical University of South Carolina); Xinghua Lu (Medical University of South Carolina); Jonas S. Almeida (Medical University of South Carolina)

Abstract: Cross-hybridization was modeled and cross-validated for cDNA microarrays using three approaches: multiple linear regression, regression tree, and artificial neural network to improve the

univariate polynomial model. The results agreed that percent identity was a good predictor, and further showed that GC content and the alignment-free measurements were significant nonlinear predictors.

Poster E-86

Combining expression data and genotyping to hunt for susceptibility genes for EAE, the murine model of Multiple Sclerosis: expression QTLs and epistatic effects

Steffen Mller (University of Rostock, Institute of Immunology); Patrik Wernhoff (University of Rostock, Institute of Immunology); Uwe K. Zettl (University of Rostock, Department of Neurology); Hans-Jrgen Thiesen (University of Rostock, Institute of Immunology); Dirk Koczan (University of Rostock, Institute of Immunology); Saleh M. Ibrahim (University of Rostock, Institute of Immunology)

Abstract: Experimental autoimmune encephalomyelitis (EAE) is a mouse model that serves for studying the etiology, pathogenesis and new therapeutic approaches for treatment of multiple sclerosis (MS). We present our approach to combine gene expression profiling and linkage analysis to identify new putative genetic pathways that contribute to the pathogenesis of EAE.

Poster E-87

Computational Characterization of mRNA 3'-Processing in Mouse Spermatogenesis

Donglin Liu (The Jackson Laboratory); Brinda Dass (Texas Tech University); John McCarrey (University of Texas at San Antonio); Clinton MacDonald (Texas Tech University); Joel Graber (The Jackson Laboratory)

Abstract: Post-transcriptional mRNA processing in mouse spermatogenesis is highly specialized. Alternate forms of the RNA-binding protein CstF2 are differentially expressed throughout spermatogenesis. Analysis of ESTs from five tissue-specific libraries demonstrates that variants of CstF2 result in altered binding specificity and stage specific alternative polyadenylations.

Poster E-88

A Neuro- Fuzzy Analysis of A Microarray Study on Human Macrophage Immune Responses to Bacteria

Chin-Fu Chen (Dept. Genetics & Biochemistry, Clemson University); Xin Feng (Dept Electrical & Computer Engineering, Marquette University)

Abstract: We employed a neuro-fuzzy approach to explore gene signatures of microarray data. We utilized a new Impact Rating for evaluating the gene responses and implemented a bootstrap permutation. Our methodology has successfully produced signature genes for each experimental condition and revealed new results not seen from the hierarchical clustering.

Poster E-89

Effective Information Integration from Disparate Microarray Datasets

Duygu Ucar (The Ohio State University - CSE department); Sarah Javaid (The Ohio State University - Biophysics program); Srinivasan Parthasarathy (The Ohio State University - CSE department)

Abstract: To discover genes that might play a role in the existence of lung cancer, an analysis was done on two disparate microarray lung cancer datasets. The datasets were integrated based on common probesets, and co-clustered both individually and together. MedlineR and Common Subsets were used to evaluate the final clusters.

Poster E-90

Novel Meta-Analysis of Disparate Datasets in Stem Cell Culture

Michael Marin (University of British Columbia, Department of Statistics & Michael Smith Laboratories); Clive Glover (University of British Columbia, Michael Smith Laboratories); James Piret (University of British Columbia, Michael Smith Laboratories & Dept. of Chemical & Biological Engineering); Jennifer Bryan (University of British Columbia, Department of Statistics & Michael Smith Laboratories)

Abstract: We conducted a novel meta-analysis of disparate datasets investigating a common biological

question. Mouse genes were ranked for their potential as ‘stem cell markers’ based on microarray data from diverse differentiation experiments. We present a method for combining evidence across datasets that are impossible to directly analyze together.

Poster E-91

Tissue Specificity Modulation of NF κ B Target Genes

Ji Chen (Bioinformatics, Univ. of Michigan); David States (Bioinformatics, Univ. of Michigan)

Abstract: A novel strategy for integrating microarray data across multiple species and sources is presented and applied to analyze NF κ B target genes. Gene/tissue biclusters are defined based on composite human and mouse expression data. Bootstrap validation reveals several reliable clusters including immune expressed cytokines and liver expressed acute phase responses.

Poster E-92

Comparison of machine learning techniques to identify biomarkers for colorectal cancer in publicly available data

Lawrence LaPointe (CSIRO Div. of Molecular Science, Flinders University of South Australia); Robert Dunne (CSIRO)

Abstract: Using published, publicly-available gene expression data we applied a support vector machine and the proprietary GeneRave algorithm to identify low dimensional gene set predictors discriminating neoplastic and non-neoplastic colorectal tissues. We also explore the utility of LDA to understand the low-dimensional feature space for quality control purposes.

Poster E-93

Classification of Breast Tumor Progression States from Cell-Based Protein Expression Data

Gregory Pennington (Department of Computer Science, Carnegie Mellon University); Stanley Shackney (Department of Human Oncology, Allegheny Singer Research Institute); Russell Schwartz (Department of Biological Sciences, Carnegie Mellon University)

Abstract: We apply clustering and phylogeny methods to identify cancer progression pathways from heterogeneous cell populations. We find that individual cells are separable into groupings identifiable in heterogeneous data sets. Phylogenetic methods can separate distinct states and identify putative progression pathways.

Poster E-94

Induction of rules for classification of breast cancer from gene expression data

Gaye Hattem (University of Missouri, Kansas City); Deendayal Dinakarpanidyan (University of Missouri, Kansas City)

Abstract: We are analyzing a combination of gene expression and clinical data from breast cancer patients to assess their relationship to clinical outcome. Inductive logic programming, which creates human-readable classification rules, is used to create a profile of genes according to lymph node metastasis and gene ontology annotation.

Poster E-95

Comparing Pre-Processing Protocols

Rob Dunne (CSIRO Mathematical & Information Sciences); Glenn Stone (CSIRO Mathematical & Information Sciences); George Miklos (Secure Genetics Pty. Ltd); Ryszard Maleszka (Research School of Biological Sciences, The Australian National University)

Abstract: We have considered several pre-processing and gene selection protocols on a number of data sets. Our conclusion is that there is little commonality in the genes selected as differentially regulated by different processing regimes. The exact degree of commonality depends on the data set and the regimes compared.

Poster E-98

Temporal Discrimination of Microarray Data for Toxicogenomics

Adam A. Smith (Department of Computer Science, University of Wisconsin Madison); Mark Craven (Department of Biostatistics and Medical Informatics, University of Wisconsin Madison)

Abstract: Analysis of microarray data may help make toxic chemical assays more efficient. We are investigating the task of learning models for classifying uncharacterized chemicals and identifying chemicals of novel classes (i.e. anomaly detection). Our model incorporates splines and clustering to interpolate the missing data and reduce noise.

Poster E-100 (*There will also be an oral presentation of this poster.*)

A Transcriptome Atlas of the Mouse Brain at Cellular Resolution

James Carson (Baylor College of Medicine); Tao Ju (Rice University); Hui-Chen Lu (Baylor College of Medicine); Christina Thaller (Baylor College of Medicine); Musodiq Bello (University of Houston); Ioannis Kakadiaris (University of Houston); Gregor Eichele (Max Planck Institute of Experimental Endocrinology); Joe Warren (Rice University); Wah Chiu (Baylor College of Medicine)

Abstract: GeneAtlas.org is a molecular atlas for 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 and comparisons that can assist in identifying genes involved in biological pathways.

Poster E-101

freeman@u.washington.edu

Theodore Freeman (University of Washington); Michael Wasnick (University of Washington); Mitchell Brittnacher (University of Washington); Laurence Rohmer (University of Washington); Gregory Taylor (University of Washington)

Abstract: We present a new genome annotation tool that simultaneously displays multiple lines of annotation information for all six translation frames. Open reading frames predicted by Glimmer are displayed along with sequence alignment search results, identified protein domains, potential ribosome binding sites and peptides identified by whole cell proteomics experiments.

Poster E-102

In Silico Characterization of Novel Alternative Splice Products

Ritesh Agrawal (Washington University); Gary Stormo (Washington University)

Abstract: Experimental characterization of novel alternative splice products originally discovered as unexpected gel bands after a PCR reaction can be time consuming and costly. We report a computational method for the direct identification of unknown splice products and test the algorithmic accuracy on known alternatively spliced genes from *Caenorhabditis elegans*.

Poster E-103

Transformation of Expression Intensities Across Generations of Affymetrix Microarrays using Sequence Matching and Regression Modeling

Soumyaroop Bhattacharya (Pulmonary Medicine, Brigham and Women's Hospital and Pulmonary Bioinformatics, The Lung Biology Center, Harvard Medical School)

Abstract: A regression based transformation model for 5069 sequence-matched probesets across different generations of Affymetrix human arrays was developed using previously published datasets describing technical replicates performed across generations of arrays. Model-based Expression measurements showed significant improvement in inter-generation correlations between sample-wide means and individual probeset pairs.

Poster E-104

Wavelet analysis of array CGH data

Taku Tokuyasu (UCSF Comprehensive Cancer Center); Donna Albertson (UCSF Comprehensive Cancer Center); Dan Pinkel (UCSF Comprehensive Cancer Center); Ajay Jain (UCSF Comprehensive Cancer Center)

Abstract: Understanding the nature of array CGH profiles and their relation to phenotypic effects remains a challenge. We apply wavelet transforms to the decomposition of such profiles by length scale, the detection of copy number transitions, noise reduction, and sample clustering based on frequently aberrant genomic regions.

Poster E-105

Strategies For Leveraging Genomic Content to Improve the Quality and Coverage of Microarray Design

Brant Wong (Affymetrix, Inc.); Gangwu Mei (Affymetrix, Inc.); Christopher Davies (Affymetrix, Inc.); Harley Gorrell (Affymetrix, Inc.); Alan Williams (Affymetrix, Inc.)

Abstract: Gene expression microarrays are typically designed using only transcript data. We show that genomic sequence data can also be used to improve the quality and coverage of the design through the clustering of observed transcript data, use of predicted transcripts, verification of observed transcript quality and orientation, and syntenic annotations.

Poster E-106

Candidate ortholog clusters in human, mouse and chicken genomes

Akshay Vashist (Dept. of Computer Science, Rutgers, The State University of New Jersey); Casimir Kulikowski (Dept. of Computer Science, Rutgers, The State University of New Jersey); Ilya Muchnik (DIMACS, Rutgers, The state university of New Jersey)

Abstract: We extract ortholog clusters from multiple genomes as quasi-cliques, using combinatorial optimization, in a multipartite graph whose vertex classes represent genomes. We found 14,254 clusters in the human, mouse and chicken genomes. According to Pfam-family organization our clusters are homogeneous, and compare well with manually curated orthologs.

Poster E-107

Discovering tissue-specific and/or cancer-specific alternative splicing events from human and mouse by genome-based EST clustering

Seung-Jae Noh (Bioinformatics Lab. Korea Research Institute of Bioscience and Biotechnology); Cheol-Goo Hur (Bioinformatics Lab. Korea Research Institute of Bioscience and Biotechnology)

Abstract: In conjunction with genome-based clustering, graph-based algorithm, statistical analysis and comparative genomic study, we discovered evolutionarily conserved alternative splicing events and tissue-specific and/or cancer-specific transcript variants from human and mouse.

Poster E-108

Unravelling the Architecture of Duplications in Tumor Genomes

Benjamin Raphael (University of California, San Diego); Pavel Pevzner (University of California, San Diego)

Abstract: We describe a computational method for reconstructing the architecture of duplicated regions in tumor genomes. Our method relies on data from End Sequence Profiling experiments and a model of breakage-fusion-bridge (BFB) cycles. We demonstrate our technique on the MCF7 breast tumor cell line.

Poster E-109

Boosting-based Transcription Factor Binding Site Prediction

Lu-yong Wang (Siemens Corporate Research); Dr. Amit Chakraborty; Dr. Dorin Comaniciu

Abstract: To understand gene transcription regulation, a robust method is required to identify the transcription factor binding sites (TFBS) to overcome the high-false-positive limitation of traditional computational methods. We are motivated to propose a robust TFBS prediction algorithm based on

boosting. It utilizes the non-binding sites information and correlation between positions systematically. It gives more robust detection performance and minimizes the false positive rate.

Poster E-110

ChipQC: microarray artifact visualization tool

Peter Henning (Georgia Institute of Technology and Emory University); David Stiles (National Institutes of Health / NIDDK Microarray Core Facility); Todd Stokes (Georgia Institute of Technology); Geoffery Wang (Georgia Institute of Technology); David Wheeler (NCBI / National Library of Medicine / National Institutes of Health); Igor Sidorov (National Cancer Institute / Laboratory of Experimental and Computational Biology); Paul Tan (National Institutes of Health / NIDDK Microarray Core Facility); Margaret Cam (National Institutes of Health / NIDDK Microarray Core Facility); May Wang (Georgia Institute of Technology and Emory University)

Abstract: ChipQC is a web-based software tool developed to perform normalization, statistical analysis, and error analysis for experiments consisting of any number of microarrays. Use of this visualization tool has revealed localized areas of high variability in technical replicate arrays consistent with pin spotting errors, air bubbles, and edge effect.

Poster E-111

Optimal Tag SNP Selection for Haplotype Reconstruction

Jin Jun (University of Connecticut); Ion Mandoiu (University of Connecticut)

Abstract: In this poster, we propose optimum tag single nucleotide polymorphism (SNP) selection methods based on integer linear programming. Experimental results on simulated data show that haplotype reconstruction based on tag SNPs is nearly as accurate as reconstruction based on all SNPs.

Poster E-112

Deducing the isoform concentration from experimental microarray data

pora kim (bioinformatics lab./molecular life science/ewha womans university); sanghyuk lee (bioinformatics lab./molecular life science/ewha womans university)

Abstract: The junction microarray data was analyzed to obtain the isoform concentration using the ECgene modeling for alternative splicing. A probe-transcript matrix was constructed for each gene model and the relative concentration of isoforms was obtained using the Non-negative least squares (NNLS) fitting of experimental data

Poster E-113

Independent Information Gain for Microarray Feature Selection: A simple metric performs well

Brian Hare (School of Computing & Engineering, University of Missouri - Kansas City); Deendayal Dinakarpanthian (School of Computing & Engineering, University of Missouri - Kansas City)

Abstract: We examined different methods of selecting features from microarrays for further analysis. Information gain was the most effective of the methods studied, and was validated using a different data set. Many genes selected had been identified as important via other methods, providing further confidence in the method.

Poster E-114

Biological Significance of Clusters of Microarray Data

Sachin Mathur (School of Computing and Engineering, University of Missouri Kansas City); Deendayal Dinakarpanthian (School of Computing and Engineering, University of Missouri Kansas City)

Abstract: We have developed a new method to biologically evaluate clusters of microarray data based on gene ontology information that takes into account relationships between various biological processes. We show that this approach gives better biological significance to clusters than with assuming biological processes to be independent entities.

Poster E-116

Selecting a clustering algorithm for statistical consistency and biological relevance for gene expression data

Susmita Datta (Georgia state University); Somnath Datta (University of Georgia)

Abstract: Clustering is commonly used to group genes based on similarity of their expression profiles. However, different clustering algorithms cluster genes differently. This report presents some criteria to select the most stable or consistent clustering algorithm which also groups genes into clusters with meaningful biological functions.

Poster E-117

HBIMD: Host-Brucella Interaction Microarray Database

Yongqun He (Virginia Bioinformatics Institute); Wenjie Zheng (Virginia Bioinformatics Institute)

Abstract: HBIMD is a web-based relational database system that contains analyzed microarray data from studies of host-Brucella interaction and provides a variety of data mining and visualization tools for better understanding of Brucella pathogenesis and host defense mechanisms.

Poster E-118

Bayesian framework for integration of microarray data and binary gene-to-gene regulatory relationships

Andrey Sivachenko (Ariadne Genomics, Inc); Anton Yuryev (Ariadne Genomics, Inc); Nikolai Daraselia (Ariadne Genomics, Inc); Ilya Mazo (Ariadne Genomics, Inc)

Abstract: We describe an integrative analysis of microarray data with regulatory relations extracted from entire Medline database (12,000,000 abstracts and full-text articles). Based on Markov Random Field model, our method estimates Bayesian posterior probabilities of differential expression and suggests a putative regulatory subnetwork to explain observed expression pattern.

Poster E-119

Distinguishing protein-coding from non-coding RNAs through SVM

Jinfeng Liu (Columbia University); Julian Gough (RIKEN Genomic Sciences Centre); Burkhard Rost (Columbia University)

Abstract: Distinguishing protein-coding from non-coding transcripts is an important problem in genome annotation. We have developed an SVM-based method to classify RNAs according to protein features of the potential peptides the RNAs are transcribed into. Ten-fold cross validation showed that the method can distinguish the coding and non-coding RNA at sensitivity of 95% and specificity of 95%.

Poster E-120

Microarray Probe Design with Minimal Cross-hybridization

Subramanian Ajay (Bioinformatics Graduate Program, University of Michigan, Ann Arbor, MI); Inhan Lee (Michigan Center for Biological Information, Department of Psychiatry, University of Michigan, Ann Arbor, MI); Brian Athey (Michigan Center for Biological Information, Department of Psychiatry, University of Michigan, Ann Arbor, MI)

Abstract: To reduce cross-hybridization we have developed a probe design algorithm that can introduce base changes that efficiently discriminate between genes that have similar sequence. From a test set of 100 RefSeq genes we were able to produce 97 target-specific probes.

Poster E-121

Sequential Classification Combining Microarray and Clinical Data

Guenter Tusch (Grand Valley State University)

Abstract: We investigate the problem of constructing a sequential classification procedure allowing for a classification into two (risk) groups at the earliest possible time using clinical and microarray data. The

quality of the procedure is maintained by specified upper bounds for the conditional errors of the entire procedure.

Poster E-122

Genome-wide survey of domain changes due to alternative pre-mRNA splicing

Younghee Lee (Ewha Womans University); Youngah Shin (Ewha Womans University); Sanghyuk Lee (Ewha Womans University)

Abstract: Domain changes due to alternative splicing (AS) were examined on a genome-wide scale using the ECgene transcript models. The origin of changes in functional domains was analyzed in terms of the AS types and frame shifts, and their phenotypic consequences were explored by inspecting EST sequences.

Poster E-123

GRAMPA: Gene Research Annotation and Microarray Profile Archive

Chris Topinka (University of Missouri); SB Bhuiyan (University of Missouri); AA Khambati (University of Missouri); RV Patel (University of Missouri); WS Spollen (University of Missouri); H Sanchez-Villeda (University of Missouri); SG Schroeder (University of Missouri); GK Springer (University of Missouri)

Abstract: GRAMPA is the description of a normalized relational object model database for molecular cloning research derived from ongoing exploratory EST genome projects coordinated by the University of Missouri Bioinformatics Consortium. The purpose is the characterization and documentation of a logical, generalized and robust schema for data generated in this domain.

Poster E-124

Data Rich But Information Poor: Applying Data Mining for Protein Function Annotation

Sivakumar Kannan (Robert Cedergren Centre for Genomics and Bioinformatics, Dpartment de Biochimie, Universit de Montral, Montral, Qubec H3C 3J7.); Gertraud Burger (Robert Cedergren Centre for Genomics and Bioinformatics, Dpartment de Biochimie, Universit de Montral, Montral, Qubec H3C 3J7.)

Abstract: Assigning function to hypothetical proteins is a classical problem in functional genomics. We developed an analysis procedure to identify protein function using predictive data mining. The goal is to detect hidden signatures and patterns in the integrated biological data, and to employ this new knowledge for deciphering genomic data at a large scale.

Poster E-125

Partial Residual Analysis of Gene Expression Data with Survival

Young Chul Kim (Seoul National University); Seung Yeoun Lee (Sejong University); Sin Ho Jung (Duke University); Taesung Park (Seoul National University)

Abstract: In the area of cancer clinical trials, the gene expression data are commonly observed along with clinical information such as survival times and covariates. We propose a testing procedure to identify prognostic genes. The proposed method uses a permutation test based on the partial residuals from Weibull regression models.

Poster E-126 (*There will also be an oral presentation of this poster.*)

Clustering 2D mRNA Expression Patterns of Drosophila Embryos

Hanchuan Peng (Genomics / Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA.); Fuhui Long (Genomics / Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA.); Michael Eisen (Genomics / Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA.); Eugene Myers (Computer Science Division, University of California, Berkeley, CA.)

Abstract: The spatio-temporal patterning of gene expression in early embryos is important in understanding functions of genes involved in development. We introduce a new method for clustering 2D images of gene expression patterns in *Drosophila melanogaster* embryos. Results on 463 genes suggest previously unobserved gene clusters with biologically interesting patterns.

Poster E-127

Alternate Splicing in Cancer: Challenges in the Statistical Analysis of EST Data

Marc A. Schaub (School of Computer Science and Department of Biological Sciences, Carnegie Mellon University); Eric P. Xing (School of Computer Science, Carnegie Mellon University); Jeff Schneider (School of Computer Science, Carnegie Mellon University); Dannie Durand (Department of Biological Sciences and School of Computer Science, Carnegie Mellon University); A. Javier Lopez (Department of Biological Sciences, Carnegie Mellon University)

Abstract: We evaluate the use of Expressed Sequence Tag data to screen for modifications of splicing patterns in cancer. Our study identifies a set of potential splicing alterations in specific tissues and demonstrates the risk of false positives due to hidden variables, sample size biases and violations of the independence assumption.

Poster E-128

Development of a Functionally Sensitive Hybrid Approach to Locating Nuclear Matrix Association Regions

Adrian Platts (Applied Genomics Technology Center, Wayne State University, Detroit, MI); Daniel Liu (Applied Genomics Technology Center, Wayne State University, Detroit, MI); Amelia Quayle (Center for Molecular Medicine & Genetics, Wayne State University, Detroit, MI); Norman Doggett (3Life Sciences Division and Center for Human Genome Studies, Los Alamos National Laboratory, Los Alamos, NM); Juergen Bode (GBF (German Research Centre for Biotechnology), Braunschweig); Stephen Krawetz (Department of Ob/Gyn, Wayne State University, Detroit, MI)

Abstract: The nuclear matrix serves as a structural and functional framework that supports several elements of nuclear organisation. We describe our efforts to identify functionally differentiated matrix attachment regions (MARs) using a combined wet-bench, in-silico strategy. This has produced a hybrid algorithm with both high specificity and sensitivity.

Poster E-129

dnaMATE: a consensus DNA melting temperature prediction server

Tomas Norambuena (Pontificia Universidad Catolica); Alejandro Panjkovich (Pontificia Universidad Catolica); Francisco Melo (Pontificia Universidad Catolica)

Abstract: There are several methods for calculating the melting temperature (T_m) of short DNA oligonucleotides. A recent study demonstrated that significant differences exist when T_m values calculated by different methods are compared. A novel consensus method that gives the most accurate estimation of experimental T_m values has been developed. This work describes a web server that implements this new method.

Poster E-130

Apoptosis as a Mediator of Delayed Tissue Damage in Progressive Stroke: A Computational Study

Kenneth Revett (University of Westminster)

Abstract: A computational model of ischemic stroke is presented that incorporates both necrosis and apoptosis as mediators of tissue damage. The model predicts that cortical spreading depression causes acute and apoptosis delayed tissue damage. The model predicts apoptosis mediates a secondary bout of CSD waves occurring on a delayed time scale.

6 Gene Regulation, microRNA's

Poster F-1

REGULATION OF COPPER METALLOTHIONEIN GENE EXPRESSION: Use of bioinformatic tools towards isolation of putative transcriptional factors.

Satish Kumar Kalari (Department of Biochemistry, University College of Science, Osmania University); Pavan Kumar Manikonda (Department of Biochemistry, University College of Science, Osmania University); Dayananda Siddavattam (Department of Animal Sciences, School of Life Sciences, Hyderabad Central University); Subramanyam Chivukula (Department of Biochemistry, University College of Science, Osmania University)

Abstract: BLAST search for upstream regions of Neurospora copper-metallothionein gene and bioinformatic screening revealed a calcineurin dependent response element, two metal response elements and an antioxidant response element in upstream regions (FEMS Letters 242: 45-50, 2005). Cloning these sequences permitted amplification of cis elements and affinity purification of putative trans factors.

Poster F-2

AMOD: A morpholino oligonucleotide selection tool

Eric Klee (University of Minnesota); Kyong Jin Shim (University of Minnesota); Michael Pickart (University of Wisconsin - Stout); Stephen Ekker (University of Minnesota); Lynda Ellis (University of Minnesota)

Abstract: AMOD is a web-based program that aids in the functional evaluation of nucleotide sequences through sequence characterization and antisense morpholino oligonucleotide (target site) selection.

Poster F-3

Using pixel-derived spot weighting to enhance motif discovery with REDUCE

Ron Tepper (Columbia University); Harmen J. Bussemaker (Columbia University)

Abstract: The REDUCE software algorithm uses regression to discover cis-regulatory motifs that explain a genome-wide set of expression ratios. Here we used pixel-level statistics to estimate the signal variance for each probe on a cDNA microarray, and found that weighted regression analysis utilizing these variances provides more accurate and sensitive motif detection.

Poster F-4

Intergenic splicing across the human genome

Pinchas Akiva (Bar-Ilan university, Compugen Ltd.); Amir Toporik (Compugen Ltd.); Sarit Edelheit (Compugen Ltd.); Yifat Peretz (Compugen Ltd.); Alex Diber (Compugen Ltd.); Ronen Shemesh (Compugen Ltd.); Amit Novik (Compugen Ltd.); Rotem Sorek (Compugen Ltd.)

Abstract: We report more than 200 cases of intergenic splicing in the human genome, where chimeric transcripts are formed by transcription of two consecutive genes into one fused RNA. We show unique characteristics of the fused transcripts and suggest that this mechanism might contribute to the evolution of protein complexes.

Poster F-5 (*There will also be an oral presentation of this poster.*)

Inferring Splicing Regulatory Activity of Short Oligonucleotides from Sequence Neighborhood

Michael Stadler (Department of Biology, Massachusetts Institute of Technology, Cambridge, MA); Gene Yeo (Crick-Jacobs Center, Salk Institute, La Jolla, CA); Noam Shomron (Department of Biology, Massachusetts Institute of Technology, Cambridge, MA); Christopher Burge (Department of Biology, Massachusetts Institute of Technology, Cambridge, MA)

Abstract: In higher eukaryotes, the degeneracy of the canonical splice signals necessitates accompanying splicing enhancers and silencers. Using a new algorithm (Neighborhood-Inference), we predicted such elements based on sequence neighbors of known function, and predictions are being experimentally verified. Our approach could be applied to other classes of antagonistic elements.

Poster F-6 (*There will also be an oral presentation of this poster.*)

Genome-wide comparative analysis of alternative splicing in plants

Bing-Bing Wang (Iowa State University); Volker Brendel (Iowa State University)

Abstract: We reported Alternative splicing (AS) in two plants: Arabidopsis and rice. 22% expressed genes showed AS in both plants. 15% of AS events were read-through, 22%-30% occurred in UTR. 36%-43% of AS events generated NMD candidates. ASIP is available at: <http://www.plantGDB.org/prj/SiP/ASIP/>.

Poster F-7

Post-transcriptional regulation of protein expression on a genome-wide scale

Andreas Beyer (IMB-Jena); Jens Hollunder (IMB-Jena); Heinz-Peter Nasheuer (National University Galway); Thomas Wilhelm (IMB-Jena)

Abstract: Based on large scale proteomics data for yeast we investigate the relation of transcription, translation, and protein turnover on a genome-wide scale. We elucidate variations between different spatial cell compartments and functional modules by comparing protein-to-mRNA ratios, translational activity, and a novel descriptor for protein specific degradation.

Poster F-10

Duoblasr - a MicroRNA search tool.

Keith Satterley (The Walter and Eliza Hall Institute of Medical Research); Leonie Gibson (The Walter and Eliza Hall Institute of Medical Research); Jerry Adams (The Walter and Eliza Hall Institute of Medical Research)

Abstract: Duoblasr is a search tool for microRNAs. Duoblasr determines all the possible stem loops in the first or query sequence. It filters out the less likely looking stem loops and seeks alignment in a syntenic area in another genome. Further filtering is done on pairs of homologous sequences found.

Poster F-11

Bioinformatic studies of gene regulation involving SOX9 and the SOX family

Angel C.Y. Mak (Department of Biochemistry, Faculty of Medicine, The University of Hong Kong); Annie Y.N. Ng (Department of Biochemistry, Faculty of Medicine, The University of Hong Kong); Sarah L. Wynn (Department of Biochemistry, Faculty of Medicine, The University of Hong Kong; Current address: National Institute for Medical Research, London, UK); Dagmar Wilhelm (Institute for Molecular Bioscience, The University of Queensland); Peter Koopman (Institute for Molecular Bioscience, The University of Queensland); Kathryn S.E. Cheah (Department of Biochemistry, Faculty of Medicine, The University of Hong Kong); David K. Smith (Department of Biochemistry, Faculty of Medicine, The University of Hong Kong)

Abstract: SOX9 is a transcription factor involved in development and disease. To better understand the role of SOX9, chromatin immunoprecipitation and computational analyses have been used to identify genes likely to be regulated by SOX9. Cross genome comparisons to identify conserved non-coding sequences and regulatory signals have been used.

Poster F-12

Is strict regulation of mRNA levels critical for translational regulation?

Nicole Cloonan (Griffith University); Matthew Crampton (Griffith University); Gillian Bushell (Griffith University)

Abstract: Protein abundance can be regulated at transcription, post-transcription, translation, or post-translation. It has been suggested that proteins regulated at translation require only minimal regulation at the RNA level. Our observations do not support this; we propose a model where regulation of mRNA levels is critical for translational regulation.

Poster F-13

Prediction of mammalian microRNA genes by comparative genome analysis

Ying Sheng (Center for Genomics and Bioinformatics, Karolinska Institute); Pr Engstrm (Center for Genomics and Bioinformatics, Karolinska Institute); Boris Lenhard (Center for Genomics and

Bioinformatics, Karolinska Institute)

Abstract: MicroRNAs are endogenous small noncoding RNAs with important regulatory roles in animals and plants. We present an efficient microRNA prediction method that uses sequence conservation profiles and secondary structure characteristics. The method predicts an extensive set of potential human and mouse microRNAs. Comparisons with additional genomes are being incorporated.

Poster F-14

Dynamic responses of the Intracellular Metabolite Concentrations of the Wild Type and pykA mutant Escherichia coli against Pulse Addition of Glucose or NH₃ under Those Limiting Continuous culture

Md. Aminul Hoque (Keio University); Haruo Ushiyama (Biott Corporation); Masaru Tomita (Keio University); Kazuyuki Shimizu (Kyushu Institute of Technology)

Abstract: We developed a new automated rapid sampling device which enables us to take samples rapidly within a second. Then the dynamics of intracellular metabolite concentrations were investigated for wild-type and pykA gene knock-out mutant Escherichia coli in responses to substrate pulse addition during different nutrient limited continuous cultures.

Poster F-15

Evolution of plant miRNA families through duplication events

Christopher Maher (Cold Spring Harbor Laboratory); Lincoln Stein (Cold Spring Harbor Laboratory); Doreen Ware (Cold Spring Harbor Laboratory / USDA)

Abstract: We have developed an approach to detect novel miRNAs (noncoding regulatory RNAs) within plant genomes, and have investigated the evolutionary conservation and expression of these miRNAs in order to begin to understand the evolution of miRNA families and their role in the regulation of plant development.

Poster F-16

Catalogs of sequence elements associated with transcript stability and sub-cellular localization derived from 3' UTRs of yeast mRNAs

Reut Shalgi (The Weizmann Institute of Science); Arren Bar-even (The Weizmann Institute of Science); Ron Shamir (Tel Aviv University); Yitzhak Pilpel (The Weizmann Institute of Science)

Abstract: We present a novel attempt to mine the regulatory motifs lying inside the terra-incognita of 3' UTRs. We derived a catalog of stability-associated motifs, and an extremely significant motif associated with mitochondrial localization, through analyses of yeast 3' UTR sequences, genome-wide mRNA half-life, and localization data.

Poster F-17

Target prediction of differentially expressed microRNAs during prion-induced neurodegeneration

G Sorensen (Division of Host Genetics and Prion Diseases, Public Health Agency of Canada, National Microbiology Laboratories, Winnipeg, MB, R3E 3R2); R Saba (Division of Host Genetics and Prion Diseases, Public Health Agency of Canada, National Microbiology Laboratories, Winnipeg, MB, R3E 3R2); MB Coulthart (Division of Host Genetics and Prion Diseases, Public Health Agency of Canada, National Microbiology Laboratories, Winnipeg, MB, R3E 3R2); SA Booth (Division of Host Genetics and Prion Diseases, Public Health Agency of Canada, National Microbiology Laboratories, Winnipeg, MB, R3E 3R2)

Abstract: In this study we predict potential targets for microRNAs known to be differentially expressed during prion-induced neurodegeneration. Functional classification and pathway analysis was performed on candidate targets to determine if these miRNAs could lead to coordinated effects, which together may have a biologically significant effect in prion-induced neurodegeneration.

Poster F-18

Predicting binding of transcriptional regulators with a two-way latent grouping model

Samuel Kaski (University of Helsinki, Department of Computer Science); Eerika Savia (Helsinki University of Technology, Laboratory of Computer and Information Science); Kai Puolamki (Helsinki University of Technology, Laboratory of Computer and Information Science)

Abstract: Binding of transcriptional regulators can be measured genome-wide to reveal regulatory networks. The measurements are noisy and expensive, however. We model existing binding data in order to predict binding for new factors or genes, assuming groups of genes and groups of transcription factors have similar binding patterns.

Poster F-19 (*There will also be an oral presentation of this poster.*)

Regulatory modules: identification and verification

Guoyan Zhao (Washington University in St. Louis); Gary Stormo (Washington University)

Abstract: We developed a method that uses a set of *C. elegans* muscle specific genes to identify muscle specific regulatory modules. It accurately predicted genes expressed in muscle from genomic sequences. This method only needs a set of co-expressed genes and are applicable to many other tissue-specific module identifications.

Poster F-20

Predicting Transcription Regulatory Mechanisms by Systematic Promoter Analysis

Li-Wei Chang (Washington University); Rakesh Nagarajan (Washington University); Jaffrey Magee (Washington University); Jeffrey Milbrandt (Washington University); Gary Stormo (Washington University)

Abstract: Identification of bona fide transcription factor binding sites in mammalian organisms remains a challenging problem. Here a systematic and statistical model of promoter analysis was established and implemented in a graphical user interface termed the promoter analysis pipeline (PAP). PAP was tested using known co-regulated gene sets, and the performance was promising.

Poster F-21

Chipper: discovering transcription-factor targets from chromatin immunoprecipitation microarrays using variance stabilization.

Francis Gibbons (Dept. of Biological Chemistry and Molecular Pharmacology, Harvard Medical School); Markus Proft (Dept. of Biological Chemistry and Molecular Pharmacology, Harvard Medical School; Current address: Instituto de Biología Molecular y Celular de Plantas (IBMCP), Universidad Politécnica de Valencia); Kevin Struhl (Dept. of Biological Chemistry and Molecular Pharmacology, Harvard Medical School); Frederick Roth (Dept. of Biological Chemistry and Molecular Pharmacology, Harvard Medical School)

Abstract: 'Variance stabilization' avoids fundamental problems of ratio-based differential-abundance measures. We use it to assess significance of ChIP results. Asymmetry in the ChIP method allows us to learn error model parameters internally, without a separate control experiment. We apply false-discovery rate analysis and have validated the method experimentally.

Poster F-22

Computational Core Promoter Prediction

Junwen Wang (Center for Bioinformatics & Department of Genetics/University of Pennsylvania); Sridhar Hannenhalli (Center for Bioinformatics & Department of Genetics/University of Pennsylvania)

Abstract: We will describe two novel algorithms aiming for core promoter identification: A generalized Markov model that uses variable model unit length and gap size, and a Positional Specific Propensity Analysis (PSPA) based model to localize the transcription start site. Both methods showed significant improvements over the traditional algorithms.

Poster F-24

A Nuclear Receptor DNA Binding Site Database

Timothy Breen (Division of Biostatistics, Indiana University School of Medicine); Lang Li (Division of Biostatistics, Indiana University School of Medicine)

Abstract: A database has been developed for nuclear receptor (NR) DNA binding sites and the corresponding NR DNA binding domain. DNA binding sites and NR DNA footprints are obtained from original peer reviewed research publications. The purpose of the database is to provide a resource for the study of NR-DNA interactions.

Poster F-25 (*There will also be an oral presentation of this poster.*)

Ab initio computational prediction of microRNA genes

Alain Sewer (University of Basel); Sebastien Pfeffer (The Rockefeller University); Alexei Aravin (The Rockefeller University); Pablo Landgraf (The Rockefeller University); Thomas Tuschl (The Rockefeller University); Erik van Nimwegen (University of Basel); Mihaela Zavolan (University of Basel)

Abstract: To identify novel (possibly species-specific) microRNAs, we designed a method that detects genomic regions forming suitable RNA secondary structures and scores them according to their similarity to known microRNAs. We illustrate the utility of our method for discovering miRNAs in viral genomes and those co-transcribed with known microRNAs.

Poster F-26 (*There will also be an oral presentation of this poster.*)

A computational method to identify amino acid residues involved in protein-DNA interactions

Changhui Yan (Iowa State University); Michael Terribilini (Iowa State University); Feihong Wu (Iowa State University); Robert Jernigan (Iowa State University); Drena Dobbs (Iowa State University); Vasant Honavar (Iowa State University)

Abstract: We present a computational method to predict interface residues involved in protein-DNA interactions. The method achieved 76% accuracy with 0.33 correlation coefficient, 39% specificity for interface residues, and 59% sensitivity for interface residues based on five-fold cross validation using a set of 56 proteins.

Poster F-27

Natively Unstructured Regions in Transcription Factors

Jiangang Liu (School of Informatics, Indiana University - Purdue University at Indianapolis); Narayanan Perumal (School of Informatics, Indiana University - Purdue University at Indianapolis); Vladimir Uversky (Center for Computational Biology and Bioinformatics, Indiana University - Purdue University at Indianapolis); A. Keith Dunker (Center for Computational Biology and Bioinformatics, Indiana University - Purdue University at Indianapolis)

Abstract: Transcription factors (TFs) regulate transcription through DNA binding and activation/inhibition of other proteins. The different domains of TFs have various secondary structures but some regions are intrinsically disordered. An analysis of a large data set of TFs shows almost a 2-fold abundance of disorder compared to random proteins.

Poster F-28

A Specialized Learner for Inferring Structured Cis-Regulatory Modules

Keith Noto (University of Wisconsin - Madison); Mark Craven (University of Wisconsin - Madison)

Abstract: We present an approach to identifying cis-regulatory modules (CRMs) in terms of binding site motifs and the arrangement of their locations relative to the transcriptional start site. It is expressive enough to capture important structural aspects of a CRM, yet the search algorithm is specifically tailored to this context.

Poster F-29

The MAPPER platform for the computational identification of transcription factor binding

sites

Voichita D. Marinescu (Children's Hospital Boston, Harvard Medical School); Isaac S. Kohane (Children's Hospital Boston, Harvard Medical School); Alberto Riva (Children's Hospital Boston, Harvard Medical School)

Abstract: MAPPER (<http://bio.chip.org/mapper>) allows the identification and visualization of putative transcription factor binding sites in a given gene, sequence or genome and uses an extensive library of HMM models built from alignments of known binding sites. We present an evaluation of the method as well as the functionality of the platform.

Poster F-31

microRNA target prediction and validation: a combined approach

Jeffrey Schageman (UT Southwestern Medical Center); Alexander Pertsemlidis (UT Southwestern Medical Center)

Abstract: We have developed a method for identifying microRNA targets based on computational predictions, in vitro validation assays and expression profiling. We have applied this method to predict and validate targets among the set of lung cancer tumor suppressor genes and the set of sterol transporters in the human liver.

Poster F-32

Sequence analysis of human alternative splices predicted from exon junction arrays

Katherina Kechris (University of California, San Francisco); Jean Yee Hwa Yang (University of California, San Francisco); Ru-Fang Yeh (University of California, San Francisco)

Abstract: We introduce a new method for predicting alternatively spliced exons from exon-junction arrays. Using contrast word counts and regression-based methods, we identify candidate splicing enhancers and silencers. Ab initio motif finding algorithms are also applied to identify motifs that are relevant for tissue-specific splicing during development.

Poster F-33

A Probabilistic Approach to Simultaneously Predicting Operons in Multiple Bacterial Genomes

Joseph Bockhorst (University of Wisconsin); Mark Craven (University of Wisconsin)

Abstract: We present an approach to simultaneously predicting operons in multiple genomes using graphical probability models. Our approach incorporates both local evidence from one genome and comparative evidence from multiple genomes. One key advantage is that local evidence for a genome, such as expression experiment results, can influence predictions on others.

Poster F-34

Computational Detection of Transcriptional Control Elements in Lymphocyte Development

Narayanan Perumal (School of Informatics, Indiana University - Purdue University at Indianapolis); Vandana Singh (School of Informatics, Indiana University - Purdue University at Indianapolis)

Abstract: Mammalian B and T lymphocyte development is transcriptionally regulated by varying gene expression patterns. Computational detection of transcriptional control elements in upstream sequences of genes specifically expressed in these cells has been attempted employing three different data sets. Biological identification of such elements will be aided by our efforts.

Poster F-35

GenePro: A Cytoscape plugin for the analysis of the transcriptional regulation and expression of protein complexes

Chris Orsi (Hospital for Sick Children, Toronto); Mark Superina (Hospital for Sick Children, Toronto); Gina Liu (Hospital for Sick Children, Toronto); Shoshana Wodak (Hospital for Sick Children, Toronto)

Abstract: GenePro is a Cytoscape plugin that provides several integrative and interactive analysis

capabilities for protein and gene networks. In particular it enables mapping of condition-specific gene expression onto the networks. We illustrate its use in the analysis of the transcriptional regulation and expression of multi-protein complexes in yeast.

Poster F-36

Revealing Predictive Gene Regulation with Interval Graph Recognition

Noppadon Khiripet (National Electronics and Computer Technology Center)

Abstract: Inferring gene regulatory networks from microarray data and other evidence has met with limited success. One of the main problems is the ambiguity in the gene control sequence. An interval graph recognition algorithm is implemented to reveal the correct gene regulation by analyzing clusters of gene interactions.

7 Databases

Poster G-1

Biomarker KnowledgeTree - A Flexible And Versatile Visualization Tool For Hierarchical Data

Mary Gaylord (Eli Lilly and Company); John Calley (Eli Lilly and Company); Huahong Qiang (Eli Lilly and Company); Birong Liao (Eli Lilly and Company)

Abstract: we describe a platform where non-hierarchical biological data can be visualized through the application of a customized hierarchy incorporating MeSH classifications. This platform gives users flexibility in update and ease of manipulation, and can facilitate fresh scientific insight by highlighting biological information through cross-referencing in different hierarchical branches.

Poster G-2

SSAHA2 Trace server - Overcoming the Computational Challenges of Searching a Rapidly Growing Archive.

Adam Spargo (The Wellcome Trust Sanger Institute); Steven Leonard (The Wellcome Trust Sanger Institute); Mark Rae (The Wellcome Trust Sanger Institute); Antony Cox (The Wellcome Trust Sanger Institute); Zemin Ning (The Wellcome Trust Sanger Institute)

Abstract: The Ensembl trace repository holds over 600 million DNA sequences with 20 million submitted each month. Current methods of database indexing are not sustainable against such rapid growth. We present a novel client-server system, exploiting commodity hardware, which is fully scalable with respect to both performance and data capacity.

Poster G-4

Gene Duplication Detection (GDD): Web-based phylogenetic analysis

Allison Hooi Chien Soo (Temasek LifeSciences Laboratory); Juguang Xiao (Temasek LifeSciences Laboratory); Alan Christoffels (Temasek LifeSciences Laboratory)

Abstract: Gene duplication detection (GDD), a web-based phylogenetic application to identify gene duplicates in protein families. GDD is designed with flexibility to configure the underlining algorithms. The workflow is divided into various stages to run in a high-performance Load Sharing Facility for parallel processing to achieve better job performance

Poster G-5

grainSAGE - A database system for the management and analysis of SAGE tag sequences

Giovanni Cordeiro (Centre for Plant Conservation Genetics, Southern Cross University); Daniel Barbary (Centre for Plant Conservation Genetics, Southern Cross University); Peter Bundock (Centre for Plant Conservation Genetics, Southern Cross University); Robert Henry (Centre for Plant Conservation Genetics, Southern Cross University)

Abstract: We have developed grainSAGE, a data warehousing facility to store and analyse large sample sets of tag sequences derived from Serial Analysis of Gene Expression (SAGE) for wheat and barley grain. Expression profiles of specific tag sequences across multiple time point libraries are visualised graphically with SWF.

Poster G-6

Life Science Information Management: from Theory to Practice

Yuerong Zhu (BioInfoRx, Inc.)

Abstract: BxAF is a data management application framework featuring open source, web-based, easy to implement, unlimited expandability, and great portability. With BxAF, designers can focus on their project-specific data management questions. BxAF is especially useful for life sciences and has been used to implement several systems, e.g. BioInfoMan.

Poster G-7

Phytome: A Plant Comparative Genomics Resource

Todd Vision (Department of Biology, University of North Carolina at Chapel Hill); Stefanie Hartmann (Department of Biology, University of North Carolina at Chapel Hill); Dihui Lu (Department of Biology, University of North Carolina at Chapel Hill); Jason Phillips (Department of Biology, University of North Carolina at Chapel Hill)

Abstract: Phytome is an internet resource for plant comparative genomics. The current version enables phylogenetic and functional exploration of protein sequences. In the future, Phytome will expand to include data and tools for comparative mapping in plants. Phytome is designed to address a number of barriers that prevent wet lab biologists from taking advantage of comparative genomics. Large amounts of heterogeneous data require considerable expertise and effort to compile for multiple species. With few exceptions, many applicable software tools are not well known outside the bioinformatics community. And many methods require computational resources that are not at the disposal of a typical laboratory. Phytome enables individual researchers to utilize the tools of plant comparative genomics by centralizing the necessary data and making the results of its computationally intensive analysis pipeline available through an intuitive web-based graphical user interface. The first public release of the database (Sep 04) is now available at <http://www.phytome.org>. Currently, there are 39 species in Phytome, including 33 angiosperms and six other land plants. Protein sequences (called Unipeptides) have been derived from a variety of different plant DNA sequence databases. Collectively, there are nearly three quarters of a million Unipeptides, of which over half a million have been classified into appx. 26,000 protein families of size two or more. For each family, a multiple sequence alignment and phylogenetic tree have been inferred. Subfamilies have been determined based on the phylogenetic trees. Interpro and Gene Ontology assignments have been made for one representative from each subfamily. In its current form, Phytome is a powerful tool for studies of functional diversification of protein families and organismal lineages. In addition, Phytome serves as a glue between otherwise disjoint plant DNA sequence databases and taxon-specific genome databases. In the future, novel functionality will be added through the inclusion of comparative mapping data and tools into the analysis pipeline and web interface.

Poster G-8

ASD:A Bioinformatics Resource on Alternative Splicing

Alphonse Thanaraj Thangavel (European Bioinformatics Institute); Stefan Stamm (University of Erlangen); Jean-Jack Marco Riethoven (European Bioinformatics Institute); Vincent Le Texier (European Bioinformatics Institute); Chellappa Gopalakrishnan (European Bioinformatics Institute); Vasudev Kumanduri (European Bioinformatics Institute)

Abstract: ASD: A Bioinformatics Resource on Alternative Splicing T.A. Thanaraj^{1*}, S. Stamm^{2*}, J.J.M. Riethoven¹, V. Le Texier¹, C. Gopalakrishnan¹, and V. Kumanduri¹ ¹European Bioinformatics Institute, UK. ²University of Erlangen, Germany. *For Correspondence: thanaraj@ebi.ac.uk, stefan@stamms-lab.net ABSTRACT ASD (<http://www.ebi.ac.uk/asd>) presents data on alternative splicing as derived through (i)

computational delineation using available transcript sequences, (ii) collecting experimentally determined data from peer-reviewed journals, and (iii) integrating from other similar resources. The reported splice events and transcript/peptide isoforms are annotated for various biological features including cross-species comparisons. The ASD resource also presents RNA splice analysis tools.

Poster G-9

SNP Prioritization Using FASTSNP

Hsiang-Yu Yuan (Institute of Biomedical Sciences, Academia Sinica); Po-he Tseng (Institute of Information Science, Academia Sinica); Jiann-Jyh Lu (Institute of Information Science, Academia Sinica); Jen-Jie Chiou (Institute of Information Science, Academia Sinica); Chun-Nan Hsu (Institute of Information Science, Academia Sinica); Shuen-Iu Hung (National Genotyping Center, Academia Sinica); Ming-Jing Hwang (Institute of Biomedical Sciences, Academia Sinica); Yuan-Tsong Chen (Institute of Biomedical Sciences, Academia Sinica); Adam Yao (National Genotyping Center, Academia Sinica)

Abstract: FASTSNP is an always up-to-date and extendable Web SNP prioritization tool for complex disease association studies. It integrates information from eleven external Web biological resources at query time to predict the queried SNPs' functional impact on transcription, pre-mRNA splicing, protein structures, etc. and prioritize the SNPs for genotyping.

Poster G-10

NMR Manager - Metabolomics Database Application for Interpreting NMR spectra

Robert Stones (Central Science Laboratory); Adrian Charlton (Central Science Laboratory); John Godward (Central Science Laboratory)

Abstract: NMR Manager is a metabolite profiling database for the management and visualisation of Nuclear Magnetic Resonance spectral metabolite data. Automated peak detection of experimental NMR and various database search engines enable retrieval of spectra in a graphical interface. Allowing rapid comparisons between database spectra against spectra generated from complex mixtures.

Poster G-11

The Cancer Biomedical Informatics Grid (caBIG)

Peter Covitz (National Cancer Institute); Sue Dubman (National Cancer Institute); Leslie Derr (National Cancer Institute); R. Mark Adams (Booz Allen Hamilton); Kenneth Buetow (National Cancer Institute)

Abstract: The NCI caBIG program is a highly coordinated biomedical informatics development and deployment organization. Participants are supported for development, adoption, training, and standards review work. Many applications are being supported, and a nationwide grid of data and analytic resources for clinical trials, translational research, and tissue banking is being constructed.

Poster G-12

In Silico Estimation of Missing Data for High Throughput Genotyping Experiments

Joel Parker (Constella Health Sciences); Myung Lee (University of North Carolina - Chapel Hill); J. Stephen Marron (University of North Carolina - Chapel Hill); Venetia Raheja (Constella Health Sciences); Ivan Rusyn (University of North Carolina - Chapel Hill); David Threadgill (University of North Carolina - Chapel Hill)

Abstract: High-throughput genotyping has generated much interest for finding disease genes. However, current methods produce 5-25% missed calls. Typing these missing points increases cost, but is necessary for statistical evaluation. A modified k-NN algorithm estimates missing data with 95-100% accuracy. Therefore, imputation may significantly reduce cost in high-throughput typing experiments.

Poster G-13

PATIKAweb: A Web service for querying, visualizing and analyzing a graph-based pathway database

Emine Zeynep Erson (Bilkent University); Asli Ayaz (University of California, Irvine); Ozgun Babur

(Bilkent University); Ahmet Cetintas (Bilkent University); Emek Demir (Bilkent University); Ugur Dogrusoz (Bilkent University); Erhan Giral (Bilkent University); Cagri Aksay (Bilkent University); Fatma Arik (Bilkent University); Esra Ataer (Bilkent University); E. Belviranli; R. Colak; G. Cozen; A. Dilek; E. Kaya; H. Yildirim

Abstract: PATIKAweb provides a Web service for retrieving and analyzing biological pathways in PATIKA database, which currently contains data integrated from popular public pathway databases like Reactome. It features a user-friendly interface, dynamic visualization, advanced graph-theoretic queries for extracting biologically important phenomena and exporting facilities to various exchange formats.

Poster G-14

Viral Bioinformatics Resource Center

Chris Upton (University of Victoria); Elliot Lefkowitz (University of Alabama, Birmingham)

Abstract: The Viral Bioinformatics Resource Center (www.biovirus.org) is one of eight NIH-sponsored Bioinformatics Resources. This VBRC was established to study Arenaviridae, Bunyaviridae, Filoviridae, Filoviridae, Paramyxoviridae, Poxviridae, and Togaviridae families. It shares resources with Viral Bioinformatics -Canada (www.virology.ca) which supports research on Herpesviruses, Baculoviruses, Coronaviruses and Adenoviruses.

Poster G-15

Developing One Step Program (SSR manager) for Rapid Identification Clones with SSRs and Marker Design

Kyu-Won Kim (National Institute of Agricultural Biotechnology); Jae-Woong Yu (National Institute of Agricultural Biotechnology); Eun-Gi Cho (National Institute of Agricultural Biotechnology); Nam-Cheon Paek (Seoul National University); Yong-Jin Park (National Institute of Agricultural Biotechnology)

Abstract: Simple Sequence Repeat (SSR) Marker system is very useful for comparing genetic characteristics and finding markers related to phenotypes of economic importance and mapping genes of interest through constructing of genetic linkage map. These merits made developing SSR markers in different crops. We have developed an one step program (SSR Manager) for rapid identification clones with SSRs and marker design to save time for tedious repeating work.

Poster G-16

A new approach for developing core sets with maximized genetic diversity and minimized redundancy using a heuristic algorithm in rice (*Oryza sativa* L.)

Hun-Ki Chung (National Institute of Agricultural Biotechnology); Gyu-Won Kim (National Institute of Agricultural Biotechnology); Jung-Ro Lee (National Institute of Agricultural Biotechnology); Eun-Ho Kim (National Institute of Agricultural Biotechnology); Hee-Kyung Kang (Kongju National University); Kenneth L. McNally (International Rice Research Institute); N.R.S. Hamilton (International Rice Research Institute); Eun-Gi Cho (National Institute of Agricultural Biotechnology); Yong-Jin Park (National Institute of Agricultural Biotechnology); Kyung-Ho Ma (National Institute of Agricultural Biotechnology)

Abstract: In this study, we introduce a new and independent program using an admissible heuristic method of A* algorithm to effectively obtain a core set with high genetic diversity and minimum redundancy. Using this program, a core set could be developed with a minimum size and maximum genetic diversity compare with conventional clustering methods.

Poster G-17

SeqServe - a web-based trace file archiving and pre-processing system

Keaogile Bezuidt (University of Pretoria); Renate Zipfel (University of Pretoria); Fourie Joubert (University of Pretoria)

Abstract: A web-based system is being developed for the archiving and pre-processing of trace files from small to medium-sized sequencing facilities. This includes instrument booking and accounting functionality.

Poster G-19

Automatic Identification and Classification of Protein Domains

Elon Portugaly (School of Computer Science Engineering, The Hebrew University of Jerusalem); Nathan Linial (School of Computer Science Engineering, The Hebrew University of Jerusalem); Michal Linial (Dept. of Biological Chemistry, Inst. of Life Sciences, The Hebrew University of Jerusalem)

Abstract: We present EVEREST, an automatic system that identifies and classifies domains within a database of protein sequences, using a random known set of families for automatic parameter tuning. The system recovers 63

Poster G-20

Current Comparative Table (CCT) automates customized searches of dynamic biological databases

Benjamin Lansteiner (St. Olaf College); Michael Olson (St. Olaf College); Robert Rutherford (St. Olaf College)

Abstract: Current Comparative Table (CCT) software enables working biologists to automate custom bioinformatic searches, typically of remote sequence or HMM databases. CCT currently supports BLAST, hmmpfam, and other programs useful for gene and ortholog identification. CCT is particularly useful for scientists studying large sets of molecules in the current evolving information landscape.

Poster G-21

GQA - Point and Click SQL Query Tool for Scientists

Gerard Hammond (Peter Wills Bioinformatic Centre)

Abstract: The Garvan Query Application (GQA) is a point and click SQL query tool which runs natively on Macintosh and Windows computers and features: Authentication and information access based on user permissions and ethics clearance; Version control; Dynamic On-line help; Extendable reporting system; User entered SQL.

Poster G-22

Efficiently Mining Sequence Patterns With Variable-Length Wildcard Regions Using An Extended Modified PrefixSpan Method

Shigetaka Tono (Hiroshima City University); Hajime Kitakami (Hiroshima City University); Keiichi Tamura (Hiroshima City University); Yasuma Mori (Hiroshima City University); Susumu Kuroki (Hiroshima City University)

Abstract: This paper proposes an extended method called the Extended Modified PrefixSpan Method for extracting these patterns. It adds a maximum error count as one of the input parameters to the existing Modified PrefixSpan method. Furthermore, in addition to constructing projected databases to perform pattern growth, our method also constructs scope databases to find variable-length wildcard regions.

Poster G-24

e-Fungi: An e-Science Infrastructure for Comparative Functional Genomics in Fungal Species

Michael Cornell (School of Computer Science, University of Manchester); Intikhab Alam (School of Computer Science, University of Manchester); M. Nedim Alpdemir (School of Computer Science, University of Manchester); Darren Soanes (Department of Biological Sciences, University of Exeter); Han Min Wong (School of Engineering, Computer Science and Mathematics, University of Exeter); Norman Paton (School of Computer Science, University of Manchester); Magnus Rattray (School of Computer Science, University of Manchester); Simon Hubbard (Faculty of Life Sciences, University of Manchester); Brian Lings (School of Engineering, Computer Science and Mathematics, University of Exeter); David Hoyle (School of Engineering, Computer Science and Mathematics, University of Exeter); Nick Talbot (Department of Biological Sciences, University of Exeter, UK.); Stephen G. Oliver (Faculty of Life Sciences, University of Manchester, UK.)

Abstract: e-Fungi integrates sequence and functional data from multiple fungal species, facilitating

systematic study of less well understood species with reference to model organisms. e-Fungi consists of a data warehouse and a library of bioinformatics analyses. Both warehouse and analysis libraries will be available within a service-oriented Grid.

Poster G-26

Systems for Biodegradation Pathway Prediction

Lynda Ellis (University of Minnesota); Dave Roe (University of Minnesota); Larry Wackett (University of Minnesota)

Abstract: We have developed a 250-rule, UNIX, web-based microbial Pathway Prediction System (PPS) (<http://umbbd.ahc.umn.edu/predict/>) to predict microbial catabolism. The PPS is expanding under Reactor (ChemAxon, Inc.), and as a stand-alone PC application based on METEOR (Lhasa, Ltd.). We compare PPS, METEOR, and Reactor.

Poster G-27

The Ensembl comparative genomics gene orthology prediction algorithm and production system

Jessica Severin (EBI/EMBL); Cara Woodwark (EBI/EMBL); Javier Herrero (EBI/EMBL); Ewan Birney (EBI/EMBL); Abel Ureta-Vidal (EBI/EMBL)

Abstract: The field of comparative genomics is inherently a non-linear problem. To solve this challenge within Ensembl, a new fault-tolerant, autonomous-agent, network-distributed processing system (ensembl-hive) has been developed. A fully-automatic implementation of this for orthology prediction is comparing 15 genomes in 24 hours.

Poster G-28

An highly optimized file format for sets of peptide sequences with practical benefits to MSMS-based database search algorithms

Jayson Falkner (University of Michigan, Dept. of Biological Chemistry, Program in Bioinformatics); Philip Andrews (University of Michigan, Dept. of Biological Chemistry)

Abstract: Presented here is an open-source, free file format for peptide sequences that uses minimal physical disk space and requires minimal parsing/loading time by algorithms. The format requires significantly less physical disk space versus FASTA and speeds up existing MSMS search algorithms, including Mascot and XTandem.

Poster G-29

A functional hierarchical organization of the protein sequence space

Noam Kaplan (The Hebrew University); Moriah Friedlich (The Hebrew University); Menachem Fromer (The Hebrew University); Michal Linial (The Hebrew University)

Abstract: ProtoNet (www.protonet.cs.huji.ac.il) is an automatic unsupervised hierarchical clustering of 1000000 proteins. We show ProtoNet captures functional aspects of the protein world. An automatic procedure is used to reduce the hierarchy to 12% its original size, while retaining the system's predictive power concerning biological function.

Poster G-30

Building and surveying transcript-based alternative splicing databases

Noboru Jo Sakabe (Ludwig Institute for Cancer Research/Department of Biochemistry, Institute of Chemistry, University of Sao Paulo); Pedro Alexandre Favoretto Galante (Ludwig Institute for Cancer Research/Department of Biochemistry, Institute of Chemistry, University of Sao Paulo); Sandro de Souza (Ludwig Institute for Cancer Research)

Abstract: ASC is a database builder designed to automatically annotate alternative splicing on cDNA clusters aligned to the genome. The user is allowed to input his own sequences and to alter the annotation process. Data is stored in a MySQL database and can be accessed through an (exon) object oriented

interface.

Poster G-31

The Use of Dynamically Defined Data Tables in the GeneX Schema

D. Andrew Carr (George Mason University/School of Computational Sciences/Bioinformatics); Hrishikesh Deshmukh (George Mason University/School of Computational Sciences/Bioinformatics); Jennifer Weller (George Mason University/School of Computational Sciences/Bioinformatics)

Abstract: The rapid increase in large expression data microarray experiments has created the need for scientifically sound and flexible data-storage and data-analysis tools with the capability of maintaining data lineage. As a solution the GeneX system has implemented constructs that partition storage and record the data processing tasks.

Poster G-32

An interactive and integrative webtool for microarray data visualization and pathway analysis

Wen Luo (Ligand Pharmaceutical Inc)

Abstract: We have set up a relational database that stores data from microarray experiments, and built an interactive website to allow biologists to easily access and analyze the microarray data. Moreover, the pathway analysis is directly linked to microarray data for the users to instantly capture biologically meaningful information.

Poster G-33

BioNote: wiki-based knowledge base and collaborative environment

Marcus Breese (Center for Medical Genomics, Indiana University School of Medicine); Robert George (Center for Medical Genomics, Indiana University School of Medicine); Matthew Grow (Center for Medical Genomics, Indiana University School of Medicine)

Abstract: BioNote combines unstructured wiki technology with semi-structured annotation to form a knowledge base and collaborative environment. With user authorization and revision controls, BioNote can serve as a repository for laboratory documents. The configurable and extendable nature of the annotations makes BioNote suitable for use in virtually any laboratory environment.

Poster G-34

Customer-oriented Dynamic Bio-database Integration Systems for analyzing the cDNA microarray

Myunguen Chung (Electronics and Telecommunication Research Institute); Myungeun Lim (Electronics and Telecommunication Research Institute); Myungnam Bae (Electronics and Telecommunication Research Institute); Sunhee Park (Electronics and Telecommunication Research Institute)

Abstract: The integration of biological data is just one phase of the entire molecular biology and genomic hypothesis discovery process. For decade, several integration methods were proposed and integration products were developed already. However there are many challenges of the usage. We developed the database integration of biological and genomic source.

Poster G-35

Design of Specific Primers for Aspergillus fumigatus by Random Selection of cDNA Clones and Application on Clinical Isolates

Alper Soyler (Department of Food Engineering, Middle East Technical University); Ayse Kalkanci (Department of Medical Microbiology, Gazi University Faculty of Medicine); Zumrut Ogel (Department of Food Engineering, Middle East Technical University)

Abstract: During the last decade Aspergillus fumigatus has become the most prevalent cause of airborne pathogenic infections and of invasive infections in humans. In this study, three specific primers were designed by developing nonconserved regions finding software to detect A. fumigatus with PCR from serum, bronchoalveolar lavage fluid and blood samples.

Poster G-36

Genome Information Broker for Viruses genome (GIB-V): Viruses genome database for comparative analysis

Masaki Hirahata (Center for Information Biology and DNA Data Bank of Japan National Institute of Genetics); Satoru Miyazaki (Faculty of Pharmaceutical Sciences, Tokyo University of Science); Takashi Abe (Center for Information Biology and DNA Data Bank of Japan National Institute of Genetics); Yasumasa Shigemoto (Life Science Systems Division, Fujitsu Limited); Hideaki Sugawara (Center for Information Biology and DNA Data Bank of Japan National Institute of Genetics)

Abstract: In this study, we developed a virus genome database "Genome Information Broker for Viruses genomes (GIB-V)" GIB-V provides functions of retrieval, browsing, downloading and comparison of virus genomes. We will also introduce the general view of the Genome Information Broker that include microbe genomes.

Poster G-37

Integrating Web Biological Resources Without Programming

Hariato Siek (Institute of Information Science, Academia Sinica); Chih-Yuan Chien (Institute of Information Science, Academia Sinica); Jen-Jie Chiou (Institute of Information Science, Academia Sinica); Chang-Keng Lee (Institute of Information Science, Academia Sinica); Tien-Yu Lin (Institute of Information Science, Academia Sinica); Jiann-Jyh Lu (Institute of Information Science, Academia Sinica); Chih-Hung Kao (Institute of Information Science, Academia Sinica); Chun-Nan Hsu (Institute of Information Science, Academia Sinica)

Abstract: A Web wrapper agent is a script that automates interactions with target Websites. It is useful for integrating external Web biological resources in an analysis pipeline but fragile and difficult to maintain. This poster presents Agent Toolbox, a tool for users to quickly configure and maintain the agent without programming.

Poster G-38

INDIAN ETHNO HEALTH DATABASE

Veerapandi Srinivasan (Rajalakshmi Engineering College); Perumal Kumar (Rajalakshmi Engineering College)

Abstract: Web enabled Databases, search and inference engines and matching with even partial data - from Information Technology are useful to bring out the valuable but voluminous unstructured data from Ayurveda, (Science of life), compiled from the ancient Indian scriptures, on herbs, plants and trees for healthcare, curative processes and environmental health.

Poster G-39

The Predictive Power of CluSTr database

Robert Petryszak (European Bioinformatics Institute); Ernst Kretschmann (European Bioinformatics Institute); Daniela Wieser (European Bioinformatics Institute); Rolf Apweiler (European Bioinformatics Institute)

Abstract: The CluSTr database employs a fully automatic single-linkage hierarchical clustering method, based on a similarity matrix. We describe a set of automated annotation experiments that quantify the predictive power and hence its biological relevance. Our results show that this approach is a valuable alternative to traditional protein classifications.

Poster G-40

Finding Development Specific Alternative Splicing Using EST Database

Tien-hsiung Ku (Changhua Christian Hospital); Fang Rong Hsu (Feng Chia University)

Abstract: Among 190863 splicing sites, 163 Development specific alternative splicing sites were found. Development stage were coded for embryonic, fetus, infantile, juvenile and adult. Numbers of ESTs for the

specific development stage and splicing isoform at each splicing site were queried and analyzed with Fisher exact test.

Poster G-41

An XML Based Management System for Biological data and Protein Version Sequence using Local Sequence Alignment

Kwang Su Jung (PHD Student/Chungbuk National Universtiy); Hyo Soung Cha (Master Course/Chungbuk National Universtiy); Sung Hee Park (PHD candidate/Chungbuk National Universtiy); Keun Ho Ryu (Professor/Chungbuk National Universtiy)

Abstract: we proposed a technique of managing protein version sequences based Smith-Waterman Algorithm which is one of local sequence alignment method. We also developed a XML based system that manages a mount of data including version sequence and integrated biological information, and transfers flat-files into other formats.

Poster G-42

MicroArray Data Explorer - mining and visualizing microarray data

Roel Verhaak (Erasmus MC); Mathijs Sanders (Erasmus MC); Maarten Bijl (Erasmus MC); Ruud Delwel (Erasmus MC); Bob Lowenberg (Erasmus MC); Peter Valk (Erasmus MC)

Abstract: MADEx is a database system that can store and visualize expression data together with results of different types of analysis, such as cluster analysis, clinical data or differential expression tests, allowing it to function as a central repository for microarray studies. Combined with several dynamic analysis- and visualization functionalities this allows researchers to quickly access microarray data on different levels.

Poster G-43

Inflammation BioKW: A Case Study in Knowledge Discovery from Heterogeneous Biological Databases through Warehousing.

Muralidharan Kannan (Department of Computer Science,IUPUI); Mathew Palakal (Department of Computer Science,IUPUI); Sudhanshu Patwardhan (Bioinformatics Institute); Santosh Kumar Mishra (Bioinformatics Institute); Subhra Kumar Biswas (Bioinformatics Institute); Jake Chen (Department of Computer Science,IUPUI)

Abstract: We introduce a new Knowledge Warehousing paradigm BioKW which integrates existing bioinformatics data sources in the knowledge level thereby presenting an environment which allows automatic analysis of datasets across multiple domains. A novel architecture for this system as well as results from a preliminary case study are reported here.

Poster G-44

Fungal Plant Pathogen Database - A cyber-infrastructure for cataloging, identification and tracking of fungal plant pathogens

Narayanan Veeraraghavan (Center for Computational Genomics, The Huck Institutes of Life Sciences, Pennsylvania State University); Junyan Luo (Department of Geography, Pennsylvania State University); Shea Paterson-Burch (Center for Computational Genomics, The Huck Institutes of Life Sciences, Pennsylvania State University); Lori Kroiss (Department of Plant Pathology, Pennsylvania State University); Alexander Richter (Center for Computational Genomics, The Huck Institutes of Life Sciences, Pennsylvania State University); Mark Gahegan (Department of Geography, Pennsylvania State University); Seogchan Kang (Department of Plant Pathology, Pennsylvania State University); Izabela Makalowska (Center for Computational Genomics, The Huck Institutes of Life Sciences, Pennsylvania State University)

Abstract: The Fungal Plant Pathogen Database (FPPD) is an internet resource for the identification of fungal plant pathogens based on the DNA sequence of genetic markers. The database cross-links digitized genotypes and phenotypes of fungal plant pathogens at both the species and population levels. It allows users to perform phylogenetic analyses in order to visualize the evolutionary relationship between a newly isolated pathogen and other isolates archived in the database, to visualize the geographic origin of

individual pathogen isolates via a map, and to search for species related information including host range, pathogen distribution, taxonomy, available markers. Built around the XML, Perl and MySQL technologies on a Sun Solaris platform, the FPPD is a multi-faceted tool for cataloging, identification and tracking of fungal plant pathogens.

Poster G-45

Omics Data Reader

Antoine Janssen (Keygene N.V.); Willem Mestrom (Keygene N.V.); Joris van Aart (Keygene N.V.); Harold Verstegen (Keygene N.V.)

Abstract: We are developing a new framework to visualize, integrate and report about different types of omics data. This includes the development of a data standard based on xml which we call the Portable Omics Format

Poster G-46

3D-protein C mutation database: integration of structural, functional and clinical data of natural Protein C mutants

Ermanna Rovida (Istituto di Tecnologie Biomediche-CNR, Milano); Pasqualina D'Ursi (Istituto di Tecnologie Biomediche-CNR, Milano); Francesca Marino (Istituto di Tecnologie Biomediche-CNR, Milano); Andrea Caprera (Istituto di Tecnologie Biomediche-CNR, Milano); Luciano Milanesi (Istituto di Tecnologie Biomediche-CNR, Milano); Giuliana Merati (I.R.C.C.S. Ospedale Policlinico, Milano); Elena Faioni (University of Milano)

Abstract: We present a specialized relational database and search tool for natural mutants of Protein C that integrates structural, functional and clinical information useful to gain insight into the relationship between a molecular defect and pathology. Mutations are mapped on the structure and 3D images are visualized by VRML and RasMol.

Poster G-47

NRC as a formal model for expressing bioinformatics workflows

Anna Gambin (Warsaw University); Jan Hidders (University of Antwerp); Natalia Kwasnikowska (Limburgs Universitair Centrum); Slawomir Lasota (Warsaw University); Jacek Sroka (Warsaw University); Jerzy Tyszkiewicz (Warsaw University); Jan Van den Bussche (Limburgs Universitair Centrum)

Abstract: Using high-throughput biotechnologies, like mass spectrometry, results in large amounts of data that require automatic analysis. In silico experiments, modeled as workflows, can provide such support. We propose Nested Relational Calculus as a formal model for expressing workflows along with a Petri net based graphical representation.

Poster G-48

The TIGR Rice Genome Annotation Database

John Hamilton (The Institute for Genomic Research); Wei Zhu (The Institute for Genomic Research); Shu Ouyang (The Institute for Genomic Research); Aihui Wang (The Institute for Genomic Research); Haining Lin (The Institute for Genomic Research); Rama Maiti (The Institute for Genomic Research); Brian Haas (The Institute for Genomic Research); Razvan Sultana (The Institute for Genomic Research); Foo Cheung (The Institute for Genomic Research); Jennifer Wortman (The Institute for Genomic Research); C. Robin Buell (The Institute for Genomic Research)

Abstract: The TIGR rice genome annotation database (<http://rice.tigr.org>) is a community resource containing sequence and annotation information. The annotation is derived from ab initio prediction and experimental evidence. The website features a Rice Genome Browser, sequence and annotation download, and a tool under development for community annotation.

Poster G-50

The Scriptome: A minimal-learning toolbox for manipulating biological data

Amir Karger (Bauer Center for Genomics Research, Harvard University); Christopher Botka (Bauer Center for Genomics Research, Harvard University); Eitan Rubin (Bauer Center for Genomics Research, Harvard University)

Abstract: Formatting and low-level analysis of data are challenging problems for many biologists. The Scriptome provides an easy-to-use toolbox of tiny Perl scripts that perform simple manipulations of biological data. Non-programming biologists build "protocols" by breaking their problems into steps, and implementing each step with a script.

Poster G-51

ERTargetDB: an integral information resource of transcription regulation of ER target genes

Victor Jin (Human Cancer Genetics Program, Comprehensive Cancer Center, Department of Molecular Virology, Immunology & Medical Genetics, The Ohio State University); Hao Sun (Human Cancer Genetics Program, Comprehensive Cancer Center, Department of Molecular Virology, Immunology & Medical Genetics, The Ohio State University); Twyla Pohar (Human Cancer Genetics Program, Comprehensive Cancer Center, Department of Molecular Virology, Immunology & Medical Genetics, The Ohio State University); Sandya Liyanarachchi (Human Cancer Genetics Program, Comprehensive Cancer Center, Department of Molecular Virology, Immunology & Medical Genetics, The Ohio State University); Saranyan Palaniswamy (Human Cancer Genetics Program, Comprehensive Cancer Center, Department of Molecular Virology, Immunology & Medical Genetics, The Ohio State University); Tim Huang (Human Cancer Genetics Program, Comprehensive Cancer Center, Department of Molecular Virology, Immunology & Medical Genetics, The Ohio State University); Ramana Davuluri (Human Cancer Genetics Program, Comprehensive Cancer Center, Department of Molecular Virology, Immunology & Medical Genetics, The Ohio State University)

Abstract: ERTargetDB provides researchers with an integral information resources of ER direct target genes. The current version contains 40 genes with experimentally verified ERE binding sites, 36 experimentally verified ERE tethering sites, 42 genes identified by the ChIP-chip, 355 genes from gene expression microarray and 2659 genes from computational prediction.

Poster G-52

Applying well-established software engineering practices to data-centric biological applications: two case studies

Richard Kamuzinzi (Universit Libre de Bruxelles); Morgane Thomas-Chollier (Vrije Universiteit Brussel - Universit Libre de Bruxelles); Albert Herzog (Universit Libre de Bruxelles); Valrie Ledent (Universit Libre de Bruxelles)

Abstract: The two case studies presented here illustrate a way to enforce current software engineering practices in biological application developments. The particular aspects of our design are the use of the Object/Relational mapping framework Hibernate, a specific organisation of the Domain Object Model and the use of Eclipse Rich Client Platform.

Poster G-53

SMID-Genomes: Predicted domain-small molecule interactions for complete genomes

Michel Dumontier (Blueprint Initiative, Samuel Lunenfeld Research Institute); Howard Feldman (Blueprint Initiative, Samuel Lunenfeld Research Institute); Kevin Snyder (Blueprint Initiative, Samuel Lunenfeld Research Institute); Christopher Hogue (Blueprint Initiative, Samuel Lunenfeld Research Institute)

Abstract: SMIDGenomes organises predicted protein domain-small molecule interactions for complete genomes, based on interactions of the Small Molecule Interaction Database (SMID). SMIDGenomes (<http://smid.blueprint.org>) provides search functionality as well as browsing by protein, domain or small molecule and maps small molecule binding sites to domains and genomic proteins.

Poster G-54

Generic Relational Database Structures for Biological Databases

James Kadin (Mouse Genome Informatics - The Jackson Laboratory); Lori Corbani (Mouse Genome Informatics - The Jackson Laboratory); Richard Baldarelli (Mouse Genome Informatics - The Jackson Laboratory); Joel Richardson (Mouse Genome Informatics - The Jackson Laboratory); Judith Blake (Mouse Genome Informatics - The Jackson Laboratory); Carol Bult (Mouse Genome Informatics - The Jackson Laboratory); Martin Ringwald (Mouse Genome Informatics - The Jackson Laboratory); Janan Eppig (Mouse Genome Informatics - The Jackson Laboratory); Jonathan Beal; Sharon Cousins; Peter Frost; Jill Lewis; Michael McCrossin,; David Miers; Michael Walker; David Walton; Joshua Winslow

Abstract: A central problem in maintaining biological databases is evolving the database schema to keep up with new data types, relationships, and new paradigms in the science. Mouse Genome Informatics (MGI) uses generic, object-oriented-like database structures in our relational schema allowing new data types/relationships with minimal schema changes.

Poster G-55

Data Integration in the Comparative Toxicogenomics Database (CTD)

GT Colby (Mount Desert Island Biological Laboratory); CJ Mattingly (Mount Desert Island Biological Laboratory); MC Rosenstein (Mount Desert Island Biological Laboratory); JN Forrest (Mount Desert Island Biological Laboratory); JL Boyer (Mount Desert Island Biological Laboratory)

Abstract: CTD (<http://ctd.mdibl.org>) integrates sequence, chemical, reference, taxonomic and Gene Ontology data to identify gene-chemical interactions and to annotate genes of toxicological significance in vertebrates and invertebrates. CTD also includes "Gene Sets," which place sequences in a unique, comparative context by grouping sequences for related genes.

Poster G-56

BioExtract Server - A Federated Database Service for Biological Data

Carol Lushbough (University of South Dakota); Carolyn Lawrence (Iowa State University); Brent Anderson (University of South Dakota); Volker Brendel (Iowa State University)

Abstract: The BioExtract server is a federated database service designed to consolidate and serve data subsets from accessible, heterogeneous, biomolecular databases. It offers a central distribution point for uniformly formatted data from various distributed data sources including relational databases, data sources hosted by Web servers and specialized proprietary data warehouses.

Poster G-57

PIRSF Protein Family Classification System

Anastasia Nikolskaya (Protein Information Resource (PIR), Georgetown University Medical Center); Sehee Chung (Protein Information Resource (PIR), Georgetown University Medical Center); Hongzhan Huang (Protein Information Resource (PIR), Georgetown University Medical Center); Raja Mazumder (Protein Information Resource (PIR), Georgetown University Medical Center); Darren Natale (Protein Information Resource (PIR), Georgetown University Medical Center); Lai-Su Yeh (Protein Information Resource (PIR), Georgetown University Medical Center); Cathy Wu (Protein Information Resource (PIR), Georgetown University Medical Center)

Abstract: The PIRSF protein classification system reflects evolutionary relationship of full-length proteins and domains. PIRSF families are extensively curated using a bioinformatics infrastructure implemented in a J2EE framework. Fully curated families and their protein members provide basis for rich and accurate functional annotation of protein sequences in the UniProt Knowledgebase.

Poster G-58

Hardware-accelerated Protein Identification for Mass Spectrometry

Anish Alex (Blueprint Initiative); Michel Dumontier (Blueprint Initiative); Jonathan Rose (University of Toronto); Christopher Hogue (Blueprint Initiative)

Abstract: An ongoing issue in mass spectrometry is the time taken to search DNA databases with MS/MS peptide fragments. We present a custom circuit implemented on Field Programmable Gate Arrays

(FPGAs) which improves the speed and lowers the cost of these searches compared to large computing clusters.

Poster G-59

The Pathos and PathoGene Databases, Analytical Environments to Support Biodefense Research

Elizabeth Glass (Argonne National Laboratory); Dinanath Sulakhe (Argonne National Laboratory); Mark D'Souza (Argonne National Laboratory); John Peterson (Argonne National Laboratory); Rick Stevens (Argonne National Laboratory); Natalia Maltsev (Argonne National Laboratory)

Abstract: The PathoGene and Pathos DB are two complementary systems to support biodefense research. PathosDB provides an integrated resource with interactive sequence analysis capabilities and PathoGene is a database containing information regarding pathogenic pathways and their components derived from the literature. Together they allow efficient and comprehensive analysis of pathogenicity.

Poster G-60

WSCUA: Web Services for Codon Usage Analysis

Denis Shestakov (Turku Centre for Computer Science)

Abstract: Codon usage analysis is successfully used in studies of molecular evolution. WSCUA (Web Services for Codon Usage Analysis) provide a program-friendly interface to codon usage analysis. WSCUA includes a web service for statistical analysis (e.g., correspondence analysis). Web services input and output are represented in XML format.

Poster G-61

Web Services for PIR/UniProt Databases

Baris Suzek (Georgetown University); Hongzhan Huang (Georgetown University); Sehee Chung (Georgetown University); Hsing-Kuo Hua (Georgetown University); Peter McGarvey (Georgetown University); Zhangzhi Hu (Georgetown University); Cathy Wu (Georgetown University)

Abstract: We have developed web services for scientific community to access PIR/UniProt protein databases in the framework of NCI/caBIG using open-source, common-standard and J2EE technology. To address data interoperability, we develop controlled vocabularies and common data elements, and adopt proteomic data standard for NIAID Proteomics Research Program.

Poster G-62

Structural Bioinformatics of Protein-Bound Water

Christopher Bottoms (University of Missouri-Columbia); Tommi White (University of Missouri-Columbia); John Tanner (University of Missouri-Columbia)

Abstract: Water plays essential roles in protein structure and function. Certain water-binding sites are particularly important, as evidenced by their conservation among proteins sharing a common three-dimensional fold. We have developed a computational method that allows the rapid identification and preliminary analyses of such sites.

Poster G-63

UniProt: the Universal Protein Resource

Lai-Su Yeh (Protein Information Resource (PIR) , Georgetown University Medical Center); UniProt Consortium (EBI/PIR/SIB)

Abstract: UniProt is the most comprehensive catalog of protein sequence and function, produced by EBI, PIR and SIB. It has three components optimized for different uses. UniProt Knowledgebase is an expertly curated database. UniProt Archive provides comprehensive sequence repository. UniProt Reference Clusters merge sequences based on sequence identity to speed searches.

Poster G-64

A Data Cleaning and Annotation Framework for Genome-wide studies

Ranjani Ramakrishnan (OGI School of Science and Engineering, Department of Computer Science& Engineering, Oregon Health and Science University); Shannon McWeeney (OGI School of Science and Engineering, Department of Computer Science& Engineering; OHSU Cancer Institute; Division of Biostatistics, Department of Public Health & Preventive Medicine. Oregon Health and Science University)

Abstract: Addressing data integration issues is critical for genome-wide studies, which overlay computational and experimentally derived genomic features. Integration issues are addressed by mapping between data sources and creating a framework to present alternate lines of evidence. Results are presented for a SACO transcription factor binding map in the mouse genome (mm5).

Poster G-65

A CASE STUDY: Web-Based Informatics System for Mouse Knockout Production

Rong Su (Stony Brook University); Aditi Pandit (Stony Brook University); Liqun Zhu (InGenious Targeting Laboratory, Inc); Wei Weng (InGenious Targeting Laboratory, Inc); Klaus Mueller (Stony Brook University)

Abstract: A high-throughput, web-based informatics system for animal knockout production is presented in this poster. The system incorporates an adaptive WfMS, a mobile-extensible LAMS module, and multi-level data organization with interfaces to various analysis and reporting tools. The system has been successfully deployed at iTL.

Poster G-66

The Biozon system for analysis of heterogeneous interrelated biological data: ranking, fuzzy searches, and topologies

Aaron Birkland (Cornell University); Golan Yona (Cornell University)

Abstract: Biozon is a knowledge resource of heterogeneous biological data that merges databases of proteins, interactions, pathways (and more) into single graph schema. Biozon provides complex searches over the graph, and supports also "fuzzy" queries that use similarity relationships. Biozon also ranks results, using a similar algorithm to Google's PageRank

Poster G-67

Knowledge Acquisition From Autonomous, Distributed, Semantically Heterogeneous Data Sources

Doina Caragea (Iowa State University); Adrian Silvescu (Iowa State University); Jyotishman Pathak (Iowa State University); Jie Bao (Iowa State University); Carson Andorf (Iowa State University); Changhui Yan (Iowa State University); Drena Dobbs (Iowa State University); Vasant Honavar (Iowa State University)

Abstract: We present INDUS - a system for collaborative discovery from multiple autonomous, distributed, semantically heterogeneous data sources. INDUS employs ontologies and user-supplied inter-ontology mappings, to support integrative analysis of data in such a setting, in the absence of a centralized data warehouse or a global ontology.

Poster G-68

A new bioinformatics strategy for biological data analysis using DNA Data Bank of Japan (DDBJ) Web services.

Hideaki Sugawara (Center for Information Biology and DNA Data Bank of Japan National Institute of Genetics); Satoru Miyazaki (Faculty of Pharmaceutical Sciences, Tokyo University of Science); Takashi Abe (Center for Information Biology and DNA Data Bank of Japan National Institute of Genetics); Yasumasa Shigemoto (Life Science Systems Division, Fujitsu Limited)

Abstract: We set up DDBJ Web services by use of SOAP server (<http://www.xml.nig.ac.jp/>). We demonstrate workflows in biological data analysis using the DDBJ Web Service; specifically introduce a workflow for the analysis of proteins or proteomics data sets. The workflow could find "hidden" linkages among database.

Poster G-69

PA-GOSUB: A model organism database for subcellular and general function predictions

Alona Fyshe (University of Alberta); Roman Eisner (University of Alberta); Russell Greiner (University of Alberta); Paul Lu (University of Alberta); Brandon Percy (University of Alberta); Brett Poulin (University of Alberta); Duane Szafron (University of Alberta); David Wishart (University of Alberta)

Abstract: PA-GOSUB is a database of popular model organism proteomes with precomputed predictions for protein function and subcellular localization. Since our last publication, the database has more than doubled in size. The predictions available in the database are searchable, browsable and easy to understand using PA's updated "Explain" facility.

Poster G-70

META-database as the collection of access methods and sets of parameters to manipulate CGIs of the molecular biological databases on the Internet.

Satoru Miyazaki (Tokyo University of Science (RIKADAI)); Toshihiko Asano (Information & Science Techno- System Co.,Ltd.); Satoshi Kitadate (Fujitsu Limited.); Hideaki Sugawara (National Institute of genetics)

Abstract: To realize efficient utilization of data resources on the Internet, we developed the database, which is the collection of publicly available databases and their access methods including parameters accepted by CGIs provided by the sites. This database on databases will be a core resource to design the semantic Web services.

Poster G-71

Rembrandt: Synergistic Leverage of High-Throughput Molecular Profiling, Clinical Data and Bioinformatics to Improve patient Outcome

Subha Madhavan (National Cancer Institute); Sahni Himanso (Science Applications International Corporation (SAIC)); Nick Xiao (Science Applications International Corporation (SAIC)); Yuri Kotliarov (Science Applications International Corporation (SAIC)); James Luo (Science Applications International Corporation (SAIC)); Mervi Heiskanen (National Cancer Institute); Sue Dubman (National Cancer Institute); Jean Claude Zenklusen (National Cancer Institute); Howard Fine (National Cancer Institute); Kenneth Buetow (National Cancer Institute)

Abstract: Rembrandt (Repository for Molecular BRAin Neoplasia DaTa) is an informatics effort led by NCI that employs datawarehousing technology and N-tiered architecture to integrate clinical/genomics data from cancer clinical trials. This application integrates expression patterns with genetic aberrations and clinical observations for classifying brain tumors into biological categories, thus improving diagnosis.

Poster G-72

CDTree: a tool to analyze and annotate protein subfamily hierarchies

Chunlei Liu (, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health)

Abstract: CDTree is an interactive graphical application designed to discover and create hierarchical relationships among domain families in a consistent, coherent fashion. Consistent and extensible GUI-based interfaces allow one to easily perform the analyses such as alignment-editing, phylogenetic tree, and domain architecture, and correlate different sources of information.

Poster G-73

caArray Data Management and Analysis Tools at the National Cancer Institute (NCI) Center for Bioinformatics

Mervi Heiskanen (National Cancer Institute); Scott Gustafson (Science Applications International Corporation (SAIC)); Xiaopeng Bian (National Cancer Institute); Juergen Lorenz (Science Applications International Corporation (SAIC)); Sumeet Muju (Science Applications International Corporation)

(SAIC)); Phu Tran (Science Applications International Corporation (SAIC)); John Moy (Science Applications International Corporation (SAIC)); Subha Madhavan (National Cancer Institute); Sue Dubman (National Cancer Institute); Ken Buetow (National Cancer Institute); Zim Zhou (Science Applications International Corporation (SAIC)); Beth Neuberger (Science Applications International Corporation (SAIC)); Hangjiong Chen (Science Applications International Corporation (SAIC)); David Bauer (Science Applications International Corporation (SAIC)); Andrew Shinohara (Science Applications International Corporation (SAIC)); Ye Wu (Science Applications International Corporation (SAIC)); Durga Addepalli (Science Applications International Corporation (SAIC)); Sharon Settnek (Science Applications International Corporation (SAIC)); Johnpaul Mejjorada (Science Applications International Corporation (SAIC))

Abstract: caArray is an open source data management system that features MIAME 1.1 compliant data annotation forms, controlled vocabularies, and MAGE-ML import and export. caArray also provides interfaces for programmatic access to data, and analytical tools. caArray database and tools can be accessed at <http://caArray.nci.nih.gov>

Poster G-74

cPath: Pathway Database Software for Systems Biology

Ethan Cerami (Memorial Sloan-Kettering Cancer Center); Gary Bader (Memorial Sloan-Kettering Cancer Center); Chris Sander (Memorial Sloan-Kettering Cancer Center)

Abstract: cPath is open-source pathway database software that eases data integration from multiple sources. It currently supports the PSI-MI Level 1 protein interaction standard and is being extended to support the BioPAX pathway format. It can be locally installed and can connect with Cytoscape for network visualization and analysis.

Poster G-75

Comparative Genome Analysis in Gramene

Chengzhi Liang (Cold Spring Harbor Laboratory); Will Spooner (Cold Spring Harbor Laboratory); Kiran Kumar (Cold Spring Harbor Laboratory); Payan Canaran (Cold Spring Harbor Laboratory); Ken Youens-Clark (Cold Spring Harbor Laboratory); Pankaj Jaiswal (Department of Plant Breeding, Cornell University); Wei Zhao (Cold Spring Harbor Laboratory); Immanuel Yap (Department of Plant Breeding, Cornell University); Doreen Ware (Cold Spring Harbor Laboratory); Lincoln Stein (Cold Spring Harbor Laboratory)

Abstract: Gramene is a resource for comparative genome analyses among major crop plants from grass family. It fosters the use of rice as anchor model for comparison to unravel the genome organization, function and evolution of these crop plants. We will present the comparative analysis infrastructure currently in use by Gramene.

Poster G-76

Semantic minimal spanning trees for semi-automatic annotation in custom databases

Nivritha Gopathi (University of Missouri, Kansas City); Deendayal Dinakarpanian (University of Missouri, Kansas City)

Abstract: A recurring requirement in the construction of custom bioinformatic databases to support experimental projects is to add annotation information based on external databases to the local data. We have developed an approach that automates this process and uses embedded semantics to deal with redundancy and to categorize the annotation.

Poster G-77

BioWarehouse: A Bioinformatics Database Warehouse Toolkit

Thomas Lee (SRI International); Yannick Pouliot (SRI International); Valerie Wagner (SRI International); David Stringer-Calvert (SRI International); Jessica Tenenbaum (Stanford University); Peter Karp (SRI International)

Abstract: We introduce BioWarehouse, an open source toolkit for constructing bioinformatics database (DB) warehouses using MySQL and Oracle. BioWarehouse integrates its component DBs into a common representational framework within a single DB management system. BioWarehouse currently supports the integration of UniProt, GenBank, ENZYME, KEGG, BioCyc, NCBI Taxonomy, and CMR.

Poster G-78

Using the Object-Relational Model for Integrating Biological Databases

Xin Li (Department of Information Systems, University of Maryland Baltimore County); Hongfang Liu (Department of Information Systems, University of Maryland Baltimore County); Aryya Gangopadhyay (Department of Information Systems, University of Maryland Baltimore County); George Karabatis (Department of Information Systems, University of Maryland Baltimore County)

Abstract: We propose a methodology based on the object-relational data model, addressing the database integration issues in bioinformatics. This methodology suggests a loosely-coupled system which integrates source biological databases; it features a plug-and-play architecture where new databases can be added with minimal modification on the integrated database.

Poster G-79

Putative Protein Network of *Candida albicans* Derived From *Saccharomyces cerevisiae*'s Protein-protein Interactions

Chia-Ling Chen (National Health Research Institutes, Taiwan); Chung-Yen Lin (National Health Research Institutes, Taiwan); Fan-Kai Lin (National Health Research Institutes, Taiwan); Chi-Shiang Cho (National Health Research Institutes, Taiwan); Chia-Ming Chang (National Health Research Institutes, Taiwan); Pao-Yang Chen (National Health Research Institutes, Taiwan); Chieh-Hua Lin (National Health Research Institutes, Taiwan); Chen-Zen Lo (National Health Research Institutes, Taiwan); Chao A. Hsiung (National Health Research Institutes, Taiwan)

Abstract: *Candida albicans* is a successful pathogen causing broad infection sites on human body and could be life-threatening with clinical evidences. We propose the Hybrid Model of Association and MLE to decrypt *Saccharomyces cerevisiae*'s protein-protein interactions into domain level for building the putative protein network of *Candida albicans*.

Poster G-80

LymphTF-DB: A Database of Transcription Factors involved in Lymphocyte Development

Paul Childress (School of Informatics, Indiana University - Purdue University at Indianapolis); Narayanan Perumal (School of Informatics, Indiana University - Purdue University at Indianapolis)

Abstract: Lymphocyte development is controlled by a significant number of transcription factors (TFs) acting on target genes. We present LymphTF-DB, a database of lymphocyte TFs and their target genes. The DB can be queried from a variety of angles including a list of TFs and target genes for developmental stages.

Poster G-81

TKB: Toxin Knowledge Base for Discovering Bio-Engineered Threats

Michael Kifer (Stony Brook University); I.V. Ramakrishnan (Stony Brook University); Arvind Ramanathan (Stony Brook University); Chang Zhao (Stony Brook University); Seetharaman Jayaraman (Brookhaven National Laboratory); Subramanyam Swaminathan (Brookhaven National Laboratory)

Abstract: We have established a Toxin Knowledge Base (TKB), a bioinformatic resource of molecular information about toxins, for discovering bio-threats. The TKB system consists of a powerful data-acquisition system and an ad-hoc query and reasoning system that is able to discover potential biological warfare agents which are missed out by other approaches.

Poster G-82

SpliceML: an XML file format for alternative splicing data exchange.

Gregory Tyrelle (Bioinformatics Research Center, National Yang-Ming University); Gloria Fu (Institute of Bioinformatics, National Yang-Ming University); Ueng-Cheng Yang (Bioinformatics Research Center, Institute of Bioinformatics, National Yang-Ming University)

Abstract: Alternative splicing (AS) is an important biological phenomenon. Available AS data, including microarray data and annotations, is growing rapidly: it is therefore critical that this body of knowledge is organized and made accessible to researchers. We propose SpliceML: an XML file format for exchange, annotation and integration of AS data.

Poster G-83

EcoCyc: a comprehensive database resource for Escherichia coli K-12

Alexander Shearer (SRI International); Ingrid Keseler (SRI International); Julio Collado-Vides (CIFN, National Autonomous University of Mexico); John Ingraham (University of California, Davis); Ian Paulsen (The Institute for Genomic Research); Peter Karp (SRI International)

Abstract: EcoCyc is a comprehensive model organism database for Escherichia coli K-12. Maintained and expanded through active literature-based curation, EcoCyc is a valuable resource for many research applications, especially computational biology projects.

Poster G-84

Development of a Data Extraction Module Supporting Bio Data Integration

Myungeun Lim (Electronics and Telecommunications Research Institute); Myungguen Chung (Electronics and Telecommunications Research Institute); Myungnam Bae (Electronics and Telecommunications Research Institute); Sunhee Park (Electronics and Telecommunications Research Institute)

Abstract: Biological data integration is an important topic because data are scattered in various sources. Data integration system can be composed of query processing, data extraction, and result integration. Among these, we designed biological data extraction module. To implement dynamic data translation from heterogeneous databases, we followed wrapper based extraction model.

Poster G-85

Cancer informatics - high speed real-time cancer cell tracking

Geoffrey Wang (Dept. of Biology, Georgia Institute of Technology); May Wang (Dept. of BME, Georgia Institute of Technology)

Abstract: Specific designed nano probes have the features to detect the cancer on the cell level. While it generates exciting results, it also needs more computing power to process the high throughput data. This research provides a real-time cancer cell data gathering and processing model.

Poster G-86

A framework for describing domain-specific microarray data with an attribute-value approach: an application in toxicogenomics data

Ji Yeon Park (Division of Genetic Toxicology, National Institute of Toxicological Research(NITR), KFDA); Misun Park (Division of Genetic Toxicology, National Institute of Toxicological Research(NITR), KFDA); Chang Yong Yoon (Division of Genetic Toxicology, National Institute of Toxicological Research(NITR), KFDA); Bang Hyun Kim (Division of Genetic Toxicology, National Institute of Toxicological Research(NITR), KFDA); Hyugsung Kwon (Division of Genetic Toxicology, National Institute of Toxicological Research(NITR), KFDA); Hai Kwan Jung (Division of Genetic Toxicology, National Institute of Toxicological Research(NITR), KFDA)

Abstract: We propose a database design using an attribute-value approach to describe heterogeneous biomaterials in the evolving MIAME standards. To implement these standards in a flexible, extensible fashion, we constructed a framework for capturing more specific information of various biomaterials, and also utilized it for managing toxicogenomics data.

Poster G-87

RIP: Integrated analysis database for the DNA repeats in primates

Taeha Woo (National Genome Information Center /Korea Research Institute of Bioscience and Biotechnology); Jungmin Seo (National Genome Information Center/Korea Research Institute of Bioscience and Biotechnology); Byungchul Kim (National Genome Information Center/Korea Research Institute of Bioscience and Biotechnology); Sangsoo Kim (National Genome Information Center/Korea Research Institute of Bioscience and Biotechnology); Chang-Bae Kim (National Genome Information Center/Korea Research Institute of Bioscience and Biotechnology)

Abstract: Integrated analysis database for the DNA repeats in primates (RIP) provides a web interface accessible at that allows the identification, visualization and selection of interspersed repetitive DNA elements in primates. RIP can search for repeat element that has a given keyword in chromosome, exon, intron and UTR region.

Poster G-88

Database of Splice Variants for Druggable Targets

Paolo Guarnieri (Axxam srl); Tod Flak (Axxam srl); Gyorgy Simon (Axxam srl); Paola Tarroni (Axxam srl)

Abstract: We developed a custom extraction and annotation procedure which populate a database of druggable targets comprising alternative transcripts information. The developed interface presents data allowing the user to gather easy access to information regarding splice variants, and designing specific probes for expression analysis experiment able to discriminate different isoforms.

Poster G-89

Mouse Genome Informatics (MGI): the integrated knowledgebase for mouse

Judith Blake (The Jackson Laboratory); Carol Bult (The Jackson Laboratory); James Kadin (The Jackson Laboratory); Joel Richardson (The Jackson Laboratory); Martin Ringwald (The Jackson Laboratory); Janan Eppig (The Jackson Laboratory)

Abstract: Mouse Genome Informatics (MGI, <http://www.informatics.jax.org>) provides integrated access to data on the genetics, genomics, and biology of the laboratory mouse. As the international community database for mouse, we develop data resources spanning from sequence to genotype to phenotype for data mining, hypothesis generation, and knowledge building.

Poster G-90

Codon64 - a new paradigm for gene sequence management

Dennis Maeder (Center of Marine Biotechnology, U. of MD Biotechnology Institute); Warren Gish (Department of Mol. Genetics, Washington University School of Medicine)

Abstract: A mnemonic code for all 64 codons is proposed. Preliminary studies generated databases of codon sequences, usage tables and blocks derived from Swissprot and cognate nucleotide sequences. A BLOCDN matrix series analogous to BLOSUM has been used for sequence alignment tests using modified blast and clustalw software.

Poster G-91

A Bayesian Network Analysis of Breast Pathology Diagnoses

Susan Maskery (Windber Research Institute); Yonghong Zhang (Windber Research Institute); Hai Hu (Windber Research Institute); Craig Shriver (Walter Reed Army Medical Center); Jeffery Hooke (Walter Reed Army Medical Center); Michael Liebman (Windber Research Institute)

Abstract: To quantitatively analyze heterogeneity in breast disease, we constructed a Bayesian network from breast disease diagnoses contained within 891 breast pathology reports from a single pathologist. This type of quantitative study of clinical associations in breast disease will enable the characterization of complex pathologic associations within samples and between patients.

Poster G-92

A property-based model for multi-disciplinary biological knowledge representation and early cancer diagnosis

Alma Barranco-Mendoza (School of Computing Science, Simon Fraser University/Computing Science Department, Trinity Western University); Deryck Persaud (Genome Sciences Centre, BC Cancer Agency); Veronica Dahl (School of Computing Science, Simon Fraser University); Gregory Eppel (Computing Science Department, Trinity Western University); Bernerd Farrant (Computing Science Department, Trinity Western University)

Abstract: We introduce the Probabilistic Property-Based Model, method for multidisciplinary biological knowledge representation, applied to early cancer diagnosis. This model is based on Concept Formation Rules, a constraint-based formalism, expanding it to enable probabilistic analysis to associate the impact of each constraint on patient's likelihood to develop cancer.

Poster G-94

Applying Bioinformatics Tools to Nutrition Genomics

Fekux Barron (Clemson University, Food Science and Human Nutrition); Vivian Haley-Zitlin (Clemson University, Food Science and Human Nutrition)

Abstract: Nutrition related diseases such as obesity, insulin resistance and hyperglycemia are affecting millions of people around the world. Nutrition genomics or Nutrigenomics has been identified as a potential area to alleviate these problems, and the efficient application of Bioinformatics tools will be critical to provide the right solutions. Our goal is to provide a framework of reference to integrate the use of Bioinformatics, genomics, and proteomics that can be used by nutrition researchers.

Poster G-96

MiMI: Michigan Molecular Interactions

Adriane Chapman (University of Michigan); Magesh Jayapandian (University of Michigan); Cong Yu (University of Michigan); H.V. Jagadish (University of Michigan)

Abstract: Michigan Molecular Interactions (MiMI) assists scientists digging through protein interaction data. MiMI merges well-known protein interaction datasets in a lossless process. A provenance model tracks where each data is from. A complementary user interface aids query formation. Thus, MiMI allows scientists to query all data, corroborative and contradictory.