Tutorials

TUTORIAL AM1
Comparative analysis of protein structures: Principles, tools, and applications for establishing evolutionary relationship and predicting function
Saturday, July 21 8:30 a.m. – 12:30 p.m. Room: Hall L

PRESENTER(S):
Raja Mazumder, Georgetown University Medical Center, USA
Sona Vasudevan, Georgetown University Medical Center, USA

ABSTRACT: This tutorial will demonstrate how to analyze and evaluate structure-sequence data for establishing evolutionary relationships and predicting functional residues (and protein function) especially for sequences that have low sequence identity (<30%) to template. How to present the analysis data convincingly with tangible evidence will also be shown. We assume a basic understanding of biology, and familiarity with the basic concepts of three-dimensional structures and protein sequences.

TUTORIAL AM2
Workflow Approaches to Transcriptomics Analysis
Saturday, July 21 8:30 a.m. – 12:30 p.m. Room: E2

PRESENTER(S):
Katy Wolstencroft, University of Manchester, UK
Georgina Moulton, University of Manchester, UK

ABSTRACT: The quantity and size of bioinformatics data is continually growing, providing rich resources to researchers, but also presenting problems of interoperability and data management. Workflows offer a solution to this problem as they enable the automated and systematic use of distributed bioinformatics data and applications from the scientist’s desktop. This provides a fast and efficient methodology for conducting large-scale experiments without the overhead of installing and maintaining local resources. Additionally, data and metadata management capabilities facilitate the support of the whole in silico experiment life cycle. This tutorial gives an overview of implementing biological workflows for comprehensive, integrated data analysis, focusing on a transcriptomics case study. We will present a review of workflow systems and successful project outcomes along with interactive demonstrations of designing and building workflows.

The session would benefit anyone (postgraduate students and researchers) wishing to explore new methods of designing complex, and/or repetitive, in silico experiments. No prior knowledge of workflows is required for this tutorial. Some knowledge of standard bioinformatics resources and transcriptomic data analyses will be useful.

TUTORIAL AM3
Ontologies for Biomedicine – How to make and use them
Saturday, July 21 8:30 a.m. – 12:30 p.m. Room: Hall M

PRESENTER(S):
Barry Smith, University at Buffalo, USA
Nigam Shah, Stanford University, USA

ABSTRACT: Ontologies are becoming essential as the quantity and types of data in the molecular biology domain rise. Though the need to use ontologies is widely appreciated, the right manner in which they should be developed and applied is not well understood. Researchers still resort to ad hoc methods in developing and using ontologies, losing opportunities for integration and cross-domain reasoning. This tutorial will provide an overview of the various ways in which ontologies are used in bioinformatics and biomedicine, along with pointers to innovative applications of ontologies such as to validate pathway data and enable cross-database reasoning. It will educate the participants on what ontologies are and in how they are currently used in molecular biology. It will provide guidance on how to create ontologies using OWL and on best practices for ontology development and use.

This tutorial will be aimed at advanced graduate students and researchers who need to understand the basic principles underlying ontology development in order to use ontologies more effectively for interpreting their own data (e.g. interpreting microarray data using the Gene Ontology or pathway information from Reactome). It will be of value, too, to researchers who need to participate in ontology development efforts in their scientific community.

Participants should be at least aware of ontologies such as the Gene Ontology and BioPAX. Familiarity with basic biology terms such as genes, proteins, promoter, intron and exon is required.
TUTORIAL AM4  
Exploring Computational Biology with a Massively Parallel High Performance Computing Environment  
Saturday, July 21  8:30 a.m. – 12:30 p.m.  Room: Hall N  

PRESENTER(S):  
Kirk Jordan, IBM, USA  
Srinivas Aluru, Iowa State University, USA  

ABSTRACT:  
Computation is playing an ever increasing and vital role in biology creating demand for new machines. Vendors strive to meet demands with advanced computer architectures such as IBM’s Blue Gene machine. In this tutorial, we will give an overview of the Blue Gene architecture. We will briefly describe both the hardware and software architecture and the central philosophy behind the development of the Blue Gene that makes it easy to use on ultrascalable problems. We will emphasize the key features that allow thousands of processors to work together on a user’s problem. We will present the programming model used on Blue Gene. We will explain ways to take advantage of the Blue Gene nodes and their associated networks. We hope to provide a foundation for attendees to begin to think about problems and how to design and implement them so they will scale out and take full advantage of the computational power in Blue Gene.

TUTORIAL AM5  
Genomes, Browsers and Databases: Tools for Integrating Sequence and Annotation Data From Multiple Genomes  
Saturday, July 21  8:30 a.m. – 12:30 p.m.  Room: Hall F2  

PRESENTER(S):  
Peter Schattner, University of California, Santa Cruz, USA  

ABSTRACT:  
The UCSC, Ensembl and NCBI genome databases integrate data from multiple, disparate sources in a uniform manner. However, developing effective interactive and programmatic queries to access these integrated databases has a considerable learning curve. Using realistic examples, participants will learn to design such queries for the analyses of genomic data. The tutorial is self-contained; however, exposure to the UCSC, ENSEMBL or NCBI Genome Browsers would be helpful. Some experience with sequence similarity searching, multiple sequence alignment, and relational databases would be helpful for parts of the tutorial. Programming skills would be useful for the last third of the tutorial. No specific knowledge of biology is required; however, familiarity with the typical molecular biology sequence analysis tasks — such as retrieving sequences from databases, parsing database files and comparing EST data to genomic sequences — will make it easier for the student to appreciate the advantages of the tools being presented.

TUTORIAL AM6  
Implementing phylogenetic workflows for comparative genomics using BioPerl  
Saturday, July 21  8:30 a.m. – 12:30 p.m.  Room: Hall O  

PRESENTER(S):  
Jason Stajich, University of California, Berkeley, USA  
Albert Vilella, European Bioinformatics Institute, UK  

ABSTRACT:  
Large-scale phylogenetic analysis of gene and protein families is an increasingly important application within comparative genomics. This tutorial demonstrates the use of BioPerl to automate phylogenomics workflows including inference of orthology, tests of natural selection on coding sequences and tests of lineage-specific gene family diversification. Participants should have at least some prior experience with phylogenetic trees, Perl and object-oriented programming.

TUTORIAL AM7  
Genomic data fusion for gene prioritization and function prediction  
Saturday, July 21  8:30 a.m. – 12:30 p.m.  Room: Hall F1  

PRESENTER(S):  
Yves Moreau, University of Leuven, Belgium  

ABSTRACT:  
With the wave of omics technologies (sequencing, microarrays, yeast two-hybrid, etc.), systems biology must now efficiently integrate (or fuse) these multiple and heterogeneous types of data. This tutorial describes computational methods to prioritize biological hypotheses and predict biological function; and demonstrates their use in optimizing biological validation and medium-throughput assays.  
Participants should have an elementary knowledge of high-throughput methods and of the basic concepts of systems biology (interaction networks, regulatory networks, and so on). The tutorial will be essentially self-contained. No prior knowledge of data fusion techniques is expected, nor a detailed knowledge of techniques for the analysis of the different types of omics data. The tutorial will be at a level accessible to all ISMB participants with a general interest in systems biology or high-throughput data analysis. All computational concepts will be presented in an intuitive fashion, so as to be accessible to participants without an extensive computational background. Biological examples will be presented in their basic biological context, so as to be accessible to participants with a limited biological background or a biological background in a different area.
TUTORIAL PM8
Gene and Protein Networks
Saturday, July 21  1:30 p.m. – 5:30 p.m.  Room: Hall F2

PRESENTER(S):
Debra Goldberg, University of Colorado at Boulder, USA
Todd Gibson, University of Colorado School of Medicine, USA

ABSTRACT: Genomic network analysis is an invaluable tool for analyzing complex systems in diverse disciplines ranging from functional genomics and chemoinformatics to evolutionary biology. Most students and researchers have not yet explored genomic network analysis beyond noting the common catchphrases which characterize them such as “small world” and “scale-free”. However, researchers are increasingly realizing genomic networks’ potential to improve biological inferences by including the biological context of molecules. Predicting unknown protein function and interactions, classifying proteins by essentiality, modeling evolution, and genomic comparisons are just a few of the tasks which have been enhanced by incorporating graph theory into the underlying biological networks. Research in genomic networks has grown so rapidly in recent years that it is becoming difficult for interested researchers to enter the field. This tutorial is intended for students and researchers who could benefit from the genomic network concepts, algorithms, and techniques used in today’s molecular biology inference tasks. This tutorial is not an outdated primer. It includes methods and issues which have been published in recent months.

No prior knowledge of network (graph) theory or molecular networks is assumed, although an awareness of gene and protein interactions (e.g. regulatory networks or protein-protein interactions) is helpful.

TUTORIAL PM9
Automatic text analysis based on Web services
Saturday, July 21  1:30 p.m. – 5:30 p.m.  Room: Hall L

PRESENTER(S):
Dietrich Rebholz-Schuhmann, European Bioinformatics Institute, UK

ABSTRACT: The tutorial teaches state of the art text mining™ of biomedical literature and focuses on the automatic use of public Web services to this end. It gives an introduction to computational linguistics, information retrieval and information extraction. It will state current TM challenges and teach evaluation techniques. It will give an overview on publicly available TM resources and services, their usefulness and their integration into local bioinformatics solutions. The tutorial will in particular explain how public TM Web services can be combined for individual purposes. Selected bioinformatics solutions will be explained in more detail, e.g. functional annotation of proteins, identification of protein-protein interactions and gene-disease associations. It closes with an outlook on future developments in the TM field and in open access publishing.

Participants should have some experience in Java and some knowledge of TM tools. Knowledge of machine learning techniques (e.g. SVMs) would be also desirable. Attendees should also be familiar with bioinformatics resources such as UniProtKb, gene ontology, UMLS.

TUTORIAL PM10
Reverse engineering mammalian transcriptional regulatory circuits
Saturday, July 21  1:30 p.m. – 5:30 p.m.  Room: Hall F1

PRESENTER(S):
Pavel Sumazin, Columbia University Medical Center, USA
Andrew D. Smith, Cold Spring Harbor Laboratory, USA
Michael Q. Zhang, Cold Spring Harbor Laboratory, USA
Andrea Califano, Columbia University Medical Center, USA

ABSTRACT: We describe data and techniques used to reverse engineer transcriptional regulatory circuits, a major challenge in systems biology. These techniques use observations including gene expression and TF binding, together with genome sequences to produce putative direct and indirect regulatory interactions that predict observations, and help reverse engineer phenotype-specific regulatory-circuit components.

Suitable for those with some previous exposure to analysis of expression or binding data.

TUTORIAL PM11
Systems Biology of Host-Pathogen Interactions and Microbial Communities
Saturday, July 21  1:30 p.m. – 5:30 p.m.  Room: Hall M

PRESENTER(S):
Christian Forst, Los Alamos University, USA

ABSTRACT: Systems Biology of Host-Pathogen Interactions and Microbial Communities examines the interactions between the components of two distinct organisms, either a microbial or viral pathogen and its host or two (or more) different microbial species in a community. With the availability of complete genomic sequences of a variety of hosts and pathogens, together with breakthroughs in proteomics, metabolomics and other experimental areas, the investigation of host-pathogen systems and microbial communities on a multitude of levels of detail comes within reach.

The purpose of this tutorial is to provide the ISMB audience a reference and starting point in this fast moving research area. It will further expand the “classic” systems biology toward interacting systems. Emerging biothreats and
climate-change, among many other issues, require novel approaches and solutions. With respect to host-pathogen interactions the aim of the tutorial is to initiate novel approaches for the detection of threat-pathogens, the prevention and treatment of infections, including drug design. With respect to microbe-microbe interactions, the tutorial introduces methods for meta-genome analysis and assesses their predictive capabilities. The tutorial also provides fundamental systems knowledge to target homogeneous and heterogeneous microbial community relationships, e.g., by quorum sensing or in soil communities.

Basic knowledge in computational and molecular biology is required; basic knowledge in immunology would be useful.

TUTORIAL PM12
Comprehensive analysis of Affymetrix Exon expression data using BioConductor
Saturday, July 21 1:30 p.m. – 5:30 p.m. Room: Hall E2

PRESENTER(S):
Crispin Miller, Paterson Institute for Cancer Research, UK
Michal Okoniewski, Paterson Institute for Cancer Research, UK

ABSTRACT: Affymetrix exon arrays enable the analysis of mRNA expression at the level of individual exons, and provide an opportunity to identify and subsequently characterize alternatively spliced genes and novel isoforms. This tutorial will show how BioConductor can be used to analyze, interpret and explore exon array data.

Participants should be familiar with the basics of R programming, have some knowledge of alternative splicing, gene expression and transcript structure, genes and the genome. Some knowledge of oligonucleotide microarrays would be advantageous.

TUTORIAL PM13
Introduction to Phylogenetic Networks
Saturday, July 21 1:30 p.m. – 5:30 p.m. Room: Hall N

PRESENTER(S):
Daniel Huson, Tuebingen University, Germany

ABSTRACT: The evolutionary history of a set of related species is usually represented by a phylogenetic tree. However, in the presence of reticulate events such as hybridization, horizontal gene transfer or recombination, a tree may not suffice and a phylogenetic network may be more suitable. Moreover, even in the absence of such events, a network may provide a useful representation that can display incompatible signals in a dataset. There are many different types of phylogenetic networks, which can be separated into two main classes: “explicit” networks aim at explicitly describing a evolutionary scenario in terms of events such as speciations, mutations, reticulations, duplications etc., whereas “implicit” networks are used to visualize incompatible signals in a dataset.

This tutorial provides an introduction to phylogenetic networks, focusing on split networks as the most important class of implicit networks and reticulate networks as an important class of explicit networks. The tutorial covers the definition of different types networks, how they are computed and what type of data they can be applied to. Additionally, we will discuss existing software.

TUTORIAL PM14
An introduction to bioinformatics for glycomics research
Saturday, July 21 1:30 p.m. – 5:30 p.m. Room: Hall O

PRESENTER(S):
Kiyoko Aoki-Kinoshita, Soka University, Japan
Claus-wilhelm von der Lieth, German Cancer Research Center, Germany

ABSTRACT: The development and use of bioinformatics tools and databases for glycomics and glycomics research has increased considerably in recent years. However, the general development in this field can still be considered as being in its infancy when compared to the genomics and proteomics areas. In terms of bioinformatics in glycobiology, there are several paths of research that are currently in progress. The development of algorithms to reliably support the characterization of glycans is the most immediate demand of the glycomics community. Additionally, several major glyco-related projects (US Consortium for Functional Glycomics, KEGG Glycan, GLYCOSCIENCES.de) are becoming mature and provide well-structured glyco-related data which are waiting for intensive data mining and data analysis. With the exciting new developments in carbohydrate arrays and automated MS annotation the analysis of the glycome has reached a new step of sophistication, which require broader (bio)informatics support.

This tutorial aims to give an overview of the current status of carbohydrate databases, the newest analytical techniques as well as the informatics needed for rapid progress in glycomics research. No prior knowledge is assumed.