BioNetVisA

Biological
Network
Visualisation
Analysis

WO2

BioNetVisA workshop

From biological network reconstruction to data visualization and analysis in molecular biology and medicine

The **BioNetVisA** workshop will bring together different actors of network biology from database providers, networks creators, computational biologists, biotech companies involved in data analysis and modeling to experimental biologists, clinicians that use systems biology approaches. The participants will be exposed to the different paradigms of network biology and the latest achievements in the field.

The goal of **BioNetVisA** workshop is to build a discussion around various approaches for biological knowledge formalisation, data integration and analysis; compatibility between different methods and biological networks resources available the field; applicability for concrete research and clinical projects depending on scientific question and type of high-throughput data.

The **BioNetVisA** workshop aims at identifying bottlenecks and proposing short- and long-term objectives for the community as discussing questions about accessibility of available tools for wide range of user in every-day standalone application in biological and clinical labs. In addition, the possibilities for collective efforts by academic researchers, clinicians, biotech companies and future development directions in the field will be discussed during the round table panel.

Organizers

Inna Kuperstein (Institut Curie, Paris, France)

Emmanuel Barillot (Institut Curie, Paris, France)

Andrei Zinovyev (Institut Curie, Paris, France)

Hiroaki Kitano (Okinawa Institute of Science and Technology Graduate University, RIKEN Center for Integrative Medical Sciences)

Nicolas Le Novère (Babraham Institute, Cambridge)

Robin Haw (Ontario Institute for Cancer Research, Toronto)

Mario Albrecht (Graz University of Technology, Graz, Austria)

Benno Schwikowski (Institut Pasteur, France)

Venue

Forum of the Medicine Campus, 4 rue Kirschleger, Strasbourg

Web site

http://www.eccb14.org/program/workshops/bionetvisa
http://sysbio.curie.fr/bionetvisa

Contact

bionetvisa@curie.fr

BioNetVisA workshop program

9.00-9.40 Keynote talk

Network analysis of high dimensional data

Tom Freeman (University of Edinburgh, Edinburgh, UK)

Session1 (part 1)

Development, curation and maintenance of biological network databases

Chair: Emmanuel Barillot (Institut Curie, Paris, France)

9.40-10.00 Reactome: Linking Biological Pathways and Networks to Disease

Robin Haw (Ontario Institute for Cancer Research, Toronto, Canada)

10.00-10.20 IntAct - High Resolution Disease-centric Networks

Henning Hermjakob (EMBL-European Bioinformatics Institute, Hinxton, UK)

10.20-10.40 Coffee break and posters

Session1 (part 2)

Development, curation and maintenance of biological network databases

Chair: Emmanuel Barillot (Institut Curie, Paris, France)

10.40-11.00 Visualization and analysis of data using Atlas of Cancer Signalling Networks (ACSN) and NaviCell tools for integrative systems biology of cancer

Inna Kuperstein (Institut Curie, Paris, France)

11.00-11.20 Graphical data representation

Pauline Gloaguen (cea- CEA, Grenoble, France)

11.20-12.00 Keynote talk

Garuda- The way biology connects

Samik Ghosh (Systems Biology Institute, Tokyo, Japan)

12.00-12.20 Posters flash presentations

Poster10: Global Analysis of coregulation for the identification of functional modules

Rim Zaag (Plant Genomics Research, Evry, France)

Poster 14: Sparse factor models for gene co-expression networks

Yuna Blum (UCLA, Los Angeles, USA)

Poster 16: Streaming Visualisation for Raw Mass Spectrometry Data Based on a Novel Compression Algorithm

Yan Zhang (University of Manchester. Manchester, UK)

12.20-13.30 Lunch and poster session

Session 2

Data visualisation and analysis in the context of biological networks

Chair: Mario Albrecht (University Medicine Greifswald, Greifswald, Germany)

13.30-13.50 A Pathway-centric Approach to Multiomics Research Powered by GeneSpring Analytics

Nigel Skinner (Agilent, Santa-Clara, CA, USA)

13.50-14.10 An integrative network analysis pipeline in Cytoscape

Mohammed El-Kebir (Centrum Wiskunde & Informatica, Amsterdam, Netherlands)

14.10-14.30 Using network analysis: HyperSet is a novel framework for functional interpretation of 'omics' data in global networks

Andrey Alexeyenko (Stockholm Science for Life Laboratory, Solna, Sweden)

Session 3

Network biology in research and medicine

Chair: Robin Haw (Ontario Institute for Cancer Research, Toronto, Canada)

14.30-14.50 Gene networks, tumor subtypes and patient prognosis signatures associated with ovarian cancer mutations

Vladimir Kuznetsov (Bioinformatics Institute, A*STAR, Singapore)

14.50-15.10 Decoding Network Dynamics in Cancer

Rune Linding (Technical University of Denmark, Lyngby, Denmark)

15.10-15.30 Using Topological Analysis to Study Metabolism of Heterotrophic Plant Cell Network

Marie Beurton-Aimar (Université Bordeaux 1, Bordeaux, France)

15.30-15.50 Conserved cross-species network modules elucidate Th17 T cell differentiation in human and mouse

Hayssam Soueidan (NKI-AVL, Amsterdam, Netherlands)

15.50-16.10 Meta expression analysis of regulatory T cell experiments for gene regulatory network reconstruction

Stefan Kroeger (Humboldt-Universität zu Berlin, Berlin, Germany)

16.10-16.30 Coffee break and posters

16.30-17.10 Keynote talk

*****THE EMBO LECTURE: HER2 and EGFR: at last, cancer therapy meets systems biology**

Yosef Yarden (The Weizmann Institut of Science, Rehovot, Israel)

17.10-17.50 Round table - discussion; awarding poster prizes and conclusions

BioNetVisA workshop abstracts

Keynote talk

Network Visualisation and Analysis of Biological Data

Tom Freeman

University of Edinburgh, Edinburgh, United Kingdom

In recent years enormous amounts of data pertaining to the activity of genes and proteins and their interactions within a cell, have been generated by a variety of functional genomics analysis platforms. In general these data are large, noisy and complex, and therefore challenging to analyse by conventional approaches. In order to better understand data generated by these approaches, there has been increasing need to develop new mathematical and computational analytical methods. In this respect network-based analyses are playing an increasingly important role, primarily because the interaction or distance between biological components can be either measured experimentally or calculated and used as a basis to draw network graphs. I will discuss the work we have been carrying in this area, in particular with respect to our studies on the network visualisation and analysis of transcriptomics data and pathway systems. I will also talk about the development of the network analysis program BioLayout Express3D (www.biolayout.org) that supports these analyses.

Talk

Reactome Knowledgebase - Linking biological pathways, networks and disease

Robin Haw

OICR, Toronto, Canada

Modern health initiatives and drug discovery are focused increasingly on targeting diseases that arise from perturbations in complex cellular events. Consequently, there has been a tremendous effort in biological research to elucidate the molecular mechanisms that underpin normal cellular processes. A reaction-network pathway knowledgebase is the tool of choice for assembling and visualizing the “parts list” of proteins and functional RNAs, as a foundation for understanding cellular processes, function and disease. The Reactome Knowledgebase (www.reactome.org) is a publically accessible, open access bioinformatics resource that stores full descriptions of human biological reactions, pathways and processes. Curated pathway knowledgebases, like Reactome, are uniquely powerful and flexible tools for extracting biologically and clinically useful information from the flood of genomic data. Specific features of Reactome support the visualization of interactions of many gene products in a complex biological process, and the application of bioinformatics tools to find causal patterns in expression data sets. To maximize Reactome’s coverage of the genome, we have supplemented curated data with a conservative set of predicted functional interactions (FI), roughly doubling our coverage of the translated genome. We have developed a Cytoscape app called “ReactomeFIViz”, which utilizes this FI network to assist biologists to perform pathway and network analysis to search for gene signatures from within gene expression data sets or identify significant genes within a list. Pathway and network-based tools for building and validating interaction networks derived from multiple data sets will give researchers substantial power to screen intrinsically noisy experimental data in order to uncover biologically relevant information.

Talk

IntAct - High Resolution Disease-centric Networks

Henning Hermjakob[1], Pablo Porras Millan[1], Margaret Duesbury[1], Sandra Orchard[1], Christian Johannes Gloeckner[2]

Proteomics Services Team, European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Cambridge, UK[1]
Institute for Ophthalmic Research, University of Tuebingen, Germany[2]

Molecular interaction databases are an essential resource that enables access to a wealth of information on associations between proteins and other biomolecules. Network graphs generated from this data provide an understanding of the relationships between different proteins in the cell, and network analysis has become a widespread tool supporting –omics analysis. However, merging data extracted from different databases, choosing the appropriate level of detail in the data for a specific analysis and representing this information in a meaningful fashion remain far from trivial. Additionally, many classes of proteins are under-represented in the interaction databases, because their physico-chemical properties mean they are not amenable to many of the high-throughput protein interaction identification methodologies. A targeted curation approach is therefore necessary to find data generated by primarily low-throughput techniques in the literature and capturing this in the interaction databases. Here we present an example of the value of targeted curation, effective visualization and in depth analysis of an interactome focusing on LRRK2, a protein of largely unknown function linked to familial forms of Parkinson's disease, using data curated into the IntAct molecular interaction database.
IntAct and its partner IMEx databases curate interactions to a high level of detail, including interaction detection methodologies, expression systems, but also interacting domains and mutated variants of interactors whenever possible. We will demonstrate how this detailed annotation is subsequently used to score interaction confidence, and to provide a detailed view of the LRRK2 interaction network.

Talk and Poster 1

Visualization and analysis of data using Atlas of Cancer Signalling Networks (ACSN) and NaviCell tools for integrative systems biology of cancer

Inna Kuperstein[1,2,3], David Cohen[1,2,3], Hien-Anh Nguyen[1,2,3], Luca Grieco[1,2,3,5,6,7], Eric Bonnet[1,2,3], Eric Viara[4], Simon Fourquet[1,2,3], Laurence Calzone[1,2,3], Christophe Russo[1,2,3], Emmanuel Barillot[1,2,3],* and Andrei Zinovyev[1,2,3,*]

Institut Curie, 26 rue d'Ulm, F-75248 Paris France[1], INSERM, U900, Paris, F-75248 France[2], Mines ParisTech, Fontainebleau, F-77300 France[3], Sysra, Paris France[4], Ecole Normale Supérieure, IBENS, 46 rue d'Ulm, Paris, France5, CNRS, UMR8197, Paris, F-75005 France[6], INSERM, U1024, Paris, F-75005 France[7]

Studying reciprocal regulations between cancer-related pathways is essential for understanding signalling rewiring during cancer evolution and in response to treatments. To allow systematic analysis of cancer signalling, the knowledge about cell mechanisms dispersed in scientific literature can be collectively represented in the form of comprehensive maps of signalling networks amenable for computational analytical methods. The Atlas of Cancer Signalling Networks (ACSN, http://acsn.curie.fr) is a resource of cancer signalling tools with interactive web-based environment for navigation, curation and data visualisation supported by a user friendly Google Maps-based tool NaviCell (http://navicell.curie.fr). Construction and update of ACSN involves manual mining of molecular biology literature and participation of the experts in the corresponding fields. ACSN covers major mechanisms involved in cancer progression systematically represented in the form of comprehensive interconnected maps. Cell signalling mechanisms are depicted on the maps in great detail, together creating a seamless map of molecular interactions in cancer. The content of ACSN is visually presented in the form of a global 'geographic-like' molecular map browsable using the Google Maps engine and semantic zooming. The associated blog provides a forum for commenting and curating the ACSN maps content. ACSN and NaviCell are a systems biology tool for integration and visualization of cancer molecular profiles generated by high through-put techniques as genome, transcriptome, proteome or analysing results from drug screenings or synthetic interaction studies. Integration and analysis of these data in the context of ACSN may help in interpretation, understanding the biological significance of the data and rationalising the scientific hypothesis. This network-based approach will help in deciphering complex molecular characteristics of cancers, improving patients stratification, predicting responses and resistance to cancer drugs and proposing new treatment strategies.

Talk and Poster 2

Towards the virtual chloroplast

Pauline Gloaguen[1], Gilles Curien[2], Sylvain Bournais[1], Christophe Bruley[1], Florence Combes[1], Giovanni Finazzi[2], Marianne Tardif[1], Yves Vandenbrouck[1], Myriam Ferro[1] and Norbert Rolland[2]

Exploring the Dynamics of Proteomes (EDyP), BGE/U1038, INSERM/CEA/Université Grenoble Alpes, F-38054 Grenoble, France[1], D-Phy-Chloro team, CNRS, UMR 5168, CEA, DSV, iRTSV, Laboratoire de Physiologie Cellulaire et Végétale, 17 rue des Martyrs, F-38054 Grenoble, France[2]

The chloroplast is a complex and integrated metabolic network that produces a high number of metabolites of industrial interest (sugars, lipids…). One way to improve our knowledge of such a "metabolic factory" and how it can be successfully engineered by synthetic biology is to automatically build metabolic pathways with well-curated and integrated knowledge. Unfortunately, current knowledge of the plastidial metabolism is dispersed in the scientific literature. Existing protein and metabolite databases do not allow quantitative estimation of metabolic fluxes and do not take into account the suborganellar localization of molecules.
Thus we decided to create a virtual chloroplast, by integrating all the qualitative and quantitative data currently available. It will contain a user-friendly interface allowing visualization and virtual modulation of metabolic fluxes, for research or for teaching purposes. First, we built a series of metabolic maps of the *Arabidopsis thaliana* chloroplast using CellDesigner. These maps have been integrated into a web interface providing direct link with biological databanks and enabling the access to semi-quantitative data on protein abundance obtained from the AT_CHLORO database, a public resource dedicated to the sub-plastidial localization of proteins. Every map is connected with each-other allowing to follow a metabolite from one metabolism to another and each component is linked to their description
These maps are extremely useful for deep curation and for sharing knowledge as well. Graphical data representation and visualization functionalities allow to directly pinpoint the metabolic steps that still need to be completed/fulfilled at the protein level and provides a better understanding of the cross-talk and/or links between different metabolisms.
In the future, we expect to use the dynamic aspect of the Scalable Vector Graphics (creation of moving pictures) to visualize fluxes/quantities variations in this network. A first release of this database available to the community is expected by the end of 2014.

Keynote talk

Garuda – The way biology connects

Samik Ghosh, Yukiko Matsuoka, Hiroaki Kitano

The Systems Biology Institute, Tokyo, Japan

The complexity of living systems entails the study of biology at multiple dimensions of time and space. In recent years, we have seen a explosion is large scale omics data which aim to capture this complexity at a systems level, commonly known as biological big data. At the same time, in order to intepret the big data, filter the signal from the noise and lead to actionable insights, researchers need to focus on the biological small data - data which leads to meaningful information and contributes to knowledge about living systems. With the ever-increasing diversity of omics-scale experimental data, a key challenge is the ability to discover the right tools for a specific analysis and navigate through their specific formats. The complexity and multi-dimensionality of data and analytics make it extremely challenging to develop a single, stand-alone system which can integrate diverse sets of data and analysis techniques. Instead, a platform whichconnects the diverse components through open interfaces, and allows the community to discover and navigate through them, can provide a unique collaborative eco-system.The vision of a community-driven, connected platform drives the Garuda Platform. Garuda is an open, community-driven, common platform that provides a framework to connect, discover and navigate through different applications, databases and services in biology and medicine.Garuda provides language independent API to connect software as gadgets, explore them through the gateway and operate them through the dashboard, all the while supported by a global alliance of leaders in computational biology and informatics.

Talk

A Pathway-centric Approach to Multiomics Research Powered by GeneSpring Analytics

Nigel Skinner

Agilent Technologies, London, UK

The talk will discuss the advantages of analyzing multiple omics data sets, enabling such data to be visualized in a biological context through the use of bioinformatics. Several examples will be provided that illustrate the acceleration of biomarker discovery & drug safety research. One example will show how the integration of transcriptomic and metabolomics data enables improved experimental design of proteomics workflows, and accelerated identification of druggable targets forming the basis for subsequent structural studies and characterization.

Talk and Poster 3

An integrative network analysis pipeline in Cytoscape

Mohammed El-Kebir[1,2], Gunnar W. Klau[1,2]

Centrum Wiskunde & Informatica, Life Sciences group, Amsterdam, the Netherlands[1]

VU University Amsterdam, IBIVU, Amsterdam, the Netherlands[2]

Traditionally, methods for the analysis of differential gene expression data were gene-centric and resulted in ranked gene lists. Later, these methods were complemented by pathway- and network-based approaches. Pathway-based approaches identify pathways whose genes are more differentially expressed than expected by chance. Integrative, network-based analysis methods result in active modules, which are connected subnetwork modules that are differentially expressed. In contrast to pathway-based approaches that identify individual pathways, an active module may span several pathways and thus reveals their crosstalk. To interpret an active module, typically an enrichment analysis using the Gene Ontology or KEGG is performed. This results in set-based annotations that may provide novel mechanistic insights in the functioning of the active module. Recently we have developed a visual analysis tool that facilitates the interpretation of annotated modules [1]. However, the entire process of obtaining, analyzing and interpreting an active module involves different tools and frameworks and remains very tedious.

Here, we present a complete and easy-to-use integrative network analysis pipeline in Cytoscape. Our pipeline consists of three modular phases. Given a network whose nodes are labeled by p-values, we identify an active module in the first phase using, for example, our newly implemented Cytoscape app for Heinz. In the second phase, we perform an enrichment analysis on the identified module in terms of the Gene Ontology and KEGG pathways. In the final phase, the resulting annotated module serves as input to our tool eXamine, which displays set membership as contours on top of a node-link layout.

We demonstrate our pipeline by analyzing a data set on deregulated signaling by the human cytomegalovirus encoded G-protein coupled receptor US28. The analysis (see http://youtu.be/LFGKek_pgjw) results in a new hypothesis about non-canonical deregulated signaling of beta-catenin.

Talk and Poster 4

Using network analysis: HyperSet is a novel framework for functional interpretation of 'omics' data in global networks

Andrey Alexeyenko

Bioinformatics Infrastructure for Life Sciences, Science for Life Laboratory;
Department of Microbiology, Tumour and Cell biology, Karolinska Institutet, Stockholm.

Several years ago we developed FunCoup, a heavily optimized Bayesian tool to predict global gene interaction networks by multi-facetted data integration (Alexeyenko and Sonnhammer, 2009; Alexeyenko et al., 2011). It used evidence from multiple species to reconstruct global networks in eukaryotic organisms, including the human. Further, to enable statistically sound network analysis to test biological hypotheses, we proposed a new method of network enrichment analysis (Alexeyenko et al., 2012; McCormack et al., 2013) where topology was employed to evaluate the functional impact of experimentally determined genes and gene sets. This analytic paradigm can assist in exploring molecular landscapes in both hypothesis-driven and hypothesis-free manner. The method extends the gene set enrichment analysis into the network domain, and was applied to e.g. validation of candidate genes in Alzheimer's disease (Reynolds et al., 2010; Hong et al., 2010; Bennet et al., 2011), characterization of cancer transcriptomes (Alexeyenko et al., 2012), and analyses of experimental results in cell lines (Szatmári et al., 2012; Akan et al., 2012).
We have proved and will demonstrate that thus obtained pathway scores are much more superior to gene expression and mutation profiles when used the as biomarkers of clinical and experimental features, such as overall survival, drug sensitivity etc.
We also address the user-friendliness of the network analysis. In our view, the novel users - experimentalists - are looking for (and rarely find) 1) biological interpretability and transparency, 2) clear, sequential analytic procedure, and 3) statistically rigorous hypothesis testing. In order to satisfy this demand, we have developed and will present the new generation on-line tool for the network enrichment analysis http://research.scilifelab.se/andrej_alexeyenko/HyperSet/.
Based on this and by comparing to existing tools, novel paradigms and problems of network analysis will be discussed with the audience.

Talk and Poster 5

Gene networks, tumor subtypes and patient prognosis signatures associated with ovarian cancer mutations

Ghim Siong Ow[1], Anna Ivshina[1], Gloria Fuentes[2], Colombie Sophié[3] and Vladimir Kuznetsov[1]

*Bioinformatics Institute, A*STAR, Singapore[1]*
RIKEN, Saitama, Japan[2]

High-grade serous ovarian cancer (HG-SOC), a major histologic type of epithelial ovarian cancer (EOC), is a poorly-characterized, heterogeneous and lethal disease where somatic mutations of TP53 are common and inherited loss-of-function mutations in BRCA1/2 predispose to cancer in 9.5-13% of EOC patients. However, the overall burden of disease due to either inherited or sporadic mutations is not known. We performed integrative genomics and network analyses of mutational and clinical data of 334 HG-SOC tumor samples from The Cancer Genome Atlas to identify novel tumor-driving mutations, survival-significant patient subgroups and tumor subtypes potentially driven by either hereditary or sporadic factors. To provide genome-wide network analysis, we first generated a two-dimensional gene-patient tumor sample mutation frequency matrix where the rows and columns correspond to 9083 unique gene symbols and 334 unique tumor sample IDs respectively . [Subsequently for each gene in the matrix, we calculated the number of tumor samples with reported mutation in this gene, as well as the total number of mutation events across all tumor samples.] Using cluster analysis, we identified a sub-cluster of high-frequency mutations in 22 patients and 58 genes, forming TP53-interconnected gene network functionally associated with DNA damage/repair, apoptosis, checkpoint and cell cycle. Mutations of CHEK2, observed with the highest intensity in this network, were associated with poor therapy response and overall survival (OS) of these patients (p=8.00e-05), possibly due to detrimental effect of mutations at the nuclear localization signal. A 21-gene mutational prognostic signature significantly stratifies patients into low or relatively high-risk subgroups with 5-year OS of 37% or 6% respectively (p=7.31e-08). Further integrative analysis of these genes and high-risk subgroups revealed two distinct sub-classes of the tumors characterized by either germline mutations of genes such as CHEK2, RPS6KA2 and MLL4, or mostly somatic mutations of other genes in the signature. Our results could provide improvement in prediction and clinical management of HG-SOC, facilitate our understanding of this complex disease, guide the design of targeted therapeutics and improve screening efforts to identify women at high-risk of hereditary ovarian cancers distinct from those associated with BRCA1/2 mutations or Lynch' syndrome.

Talk

Decoding Network Dynamics in Cancer

Rune Linding

Technical University of Denmark, Lyngby, Denmark

Biological systems are composed of highly dynamic and interconnected molecular networks that drive biological decision processes. The goal of network biology is to describe, quantify and predict the information flow and functional behavior of living systems in a formal language and with an accuracy that parallels our characterization of other physical systems such as Jumbo-jets. Decades of targeted molecular and biological studies have led to numerous pathway models of developmental and disease related processes. However, so far no global models have been derived from pathways, capable of predicting cellular trajectories in time, space or disease. The development of high-throughput methodologies has further enhanced our ability to obtain quantitative genomic, proteomic and phenotypic readouts for many genes/proteins simultaneously. Here, I will discuss how it is now possible to derive network models through computational integration of systematic, large-scale, high-dimensional quantitative data sets. I will review our latest advances in methods for exploring phosphorylation networks. In particular I will discuss how the combination of quantitative mass-spectrometry, systems-genetics and computational algorithms (NetworKIN [1], NetPhorest [4] and KinomeXplorer [10]) made it possible for us to derive systems-level models of JNK and EphR signaling networks [2,3]. I shall discuss work we have done in comparative phospho-proteomics and network evolution [5-7]. Finally, I will discuss our most recent work in analyzing genomic sequencing data from NGS studies and how we have developed new powerful algorithms to predict the impact of disease mutations on cellular signaling networks [8,9]. I shall illustrate the power of these approaches in a recent study where we have identified colon cancer metastasis cell signaling networks.

References:

http://www.lindinglab.org

1. Linding et al., Cell 2007.
2. Bakal et al., Science 2008.
3. Jørgensen et al., Science 2009.
4. Miller et al., Science Signaling 2008.
5. Tan et al., Science Signaling 2009.
6. Tan et al., Science 2009.
7. Tan et al., Science 2011.
8. Creixell et al., Nature Biotechnology 2012.
9. Erler & Linding, Cell 2012.
10. Horn et al., Nature Methods 2014.

Talk and Poster 6

Using Topological Analysis to Study Metabolism of Heterotrophic Plant Cell Network

Beurton-Aimar Marie[1], Nguyen Vu Ngoc Tung[1,2] and Colombie Sophié[3]

LaBRI, Laboratoire Bordelais de Recherche en Informatique, UMR5800, University Bordeaux, France[1]
Faculty of Science and Technology, Hoa Sen University, Vietnam[2]
INRA Bordeaux, UMR1332, Fruit Biology and Pathology, France[3]

Pathway-based analysis of plant metabolism is still interested in many works and opening, for example to study the accumulation sugars and amino acids in plant cells. Tools coming from graph theory have been developed, EFMs algorithm (Schuster and Hilgetag 1994) allows finding all minimal and unique feasible pathways through a network (respecting steady state constraints) and Minimal Cut Sets (MCSs) algorithm (S. Klamt, 2005) generates the list of reactions sets which are able to stop the flux through these feasible pathways. We have defined a plant cell network containing classical pathways: glycolysis, penthose phosphate, TCA cycle, starch and sucrose pathways (73 reactions and 70 metabolites). In order to analyze the production of 5 metabolites (Starch, Fructose, Glucose, Sucrose and Glutamate) essential for the plant growing we have computed the EFMs of this network. More than 100,000 EFMs, feasible pathways, have been found. To characterize the network behaviors if the uptake of glucose, i.e. the exogenous glucose, is stopped, we have selected the EFMs which are able to produce the 5 metabolites of interest and computed their MCS to identify the key reactions for this production. A set of 8 reactions have been identified as being mandatory to keep the flux if the uptake of glucose is stopped, 3 reactions from the glycolysis, 4 from the pentose phosphate pathway and one transporter from cytosol to plastid compartment. This result points out a particular topology of the network which is not possible to identify from the analysis of each pathway individually. Combination of feasible pathways computation through the whole network and identification of reactions to stop the flux highlight the key role of some reactions. These results could lead to select those reactions as potential target for further studies by biologists.

Talk and Poster 7

Conserved cross-species network modules elucidate Th17 T cell differentiation in human and mouse

Hayssam Soueidan[3], Mohammed El-Kebir[1,2], Thomas Hume[4,5], Daniela Beisser[6], Marcus Dittrich[7], Tobias Muller[7], Guillaume Blin[5], Jaap Heringa[2], Macha Nikolski[4,5] , Lodewyk F. A. Wessels[3], Gunnar W. Klau[1]

Life Sciences, Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands[1]
Centre for Integrative Bioinformatics VU, VU University Amsterdam, The Netherlands[2]
Computational Cancer Biology, The Netherlands Cancer Institute, Amsterdam, The Netherlands[3]
Univ. Bordeaux, CBiB, Bordeaux, France[4]
Univ. Bordeaux, CNRS / LaBRI, Talence, France[5]
Genome Informatics, University of Duisburg-Essen, Germany[6]
Department of Bioinformatics, Biocenter, University of Wurzburg, Germany[7]

Motivation: Integrative network analysis methods provide robust interpretations of differential high-throughput molecular profile measurements. They are often used in a biomedical context—either to generate novel hypotheses about the underlying cellular processes or to derive biomarkers for classification and subtyping. Their results, however, are not immediately transferable to human as for ethical or practical reasons the underlying molecular profiles are often measured and validated on animal or cellular models. In particular, this is also the case in a study of the recently discovered interleukin-17 producing helper T cells (Th17), which are fundamental for anti- microbial immunity but also known to contribute to autoimmune diseases.

Results: We propose a sound mathematical model for finding active subnetwork modules that are conserved between two species. These are sets of genes, one for each species, which (i) induce a connected subnetwork in a species-specific interaction network, (ii) show overall differential behavior and (iii) contain a large number of orthologous genes. We propose a flexible notion of conservation, which turns out to be crucial for the quality of the resulting modules in terms of biological interpretability. We propose an algorithm that finds provably optimal conserved active modules in our model. We apply our algorithm to understand the mechanisms underlying Th17 T cell differentiation in both mouse and human. As a main biological result, we find that the key regulation of Th17 differentiation is conserved between human and mouse.

Availability: xHeinz, an implementation of our algorithm, as well as all input data and results are available at http://software.cwi. nl/xheinz.

Talk and Poster 8

Meta expression analysis of regulatory T cell experiments for gene regulatory network reconstruction

Stefan Kroeger[1], Melanie Venzke[2], Ria Baumgrass[2], Ulf Leser[1]

Humboldt Universität zu Berlin, Institute for Computer Science, Berlin, Germany[1]
Deutsches Rheuma-Forschungszentrum, a Leibniz Institute, Berlin, Germany[2]

Reconstruction of gene regulatory networks (GRN) from gene expression data is a promising technique for elucidating key mechanisms in living organisms. However, successful applications so far have mostly been reported for organisms with small genomes. In general, the amount of data necessary to obtain robust results grows quickly with the complexity of the networks under study. Here, we report on our efforts to study regulatory processes in murine regulatory T cells (Tregs) using expression data-based network reconstruction. Better knowledge about T cell development is important for the understanding of physiology and pathophysiology of the adaptive immune system.
Our key idea to alleviate the data acquisition bottleneck is to use large amounts of publicly available, albeit heterogeneous, datasets. We augment this primary and noisy data with specific gene sets obtained from text mining Medline abstracts and full texts using Treg-specific queries. By combining large expression data sets with a smaller set of high confidential gene sets we aim to overcome the “small n, large p” problem.
In detail, we computed a large, manually curated expression profile matrix from 36 Treg cell related experiments obtained from GEO [1]. Analysis was performed using a two-step quantile normalization to reduce batch effects. We next applied different GRN-reconstruction algorithms, namely Aracne, Genie3, CLR, MRNet and co-expression. The reconstructed networks were integrated into a weighted consensus network by aggregating the assigned edge attributes (e.g. mutual information). Subsequently, for each network edges were ranked considering attribute values as edge-weights.
An initial evaluation of the top ranked edges shows that the consensus network contains roughly same amount of gene-gene-interactions present in the respective KEGG [3] or Reactome[4] pathways or listed in String [4] or MSigDB[5] as the selected GRN-reconstruction algorithms. We see this as an encouraging first result towards methods for regulatory network reconstruction in mammals.

Keynote talk

HER2 and EGFR: at last, cancer therapy meets systems biology

Yosef Yarden

Weizmann Institute, Rehovot, Israel

Tumor-specific combinations of oncogenic mutations often free cancer cells from their reliance on growth factors. One important example comprises the epidermal growth factor receptor (EGFR) and its kin, HER2. In tumors, both EGFR and HER2 frequently display overexpression, internal deletions and point mutations. Accordingly, several monoclonal antibodies and kinase inhibitors specific to these receptors have been approved for clinical application. However, similar to the application of chemotherapeutic drugs, the efficacy of therapies specifically targeting EGFR or HER2 is limited by primary (intrinsic) and secondary (evolving) resistance. In addition, intra-tumor heterogeneity severely reduces therapeutic efficacy.

My lecture will introduce a systems biology approach to understanding drug action and evolvement of secondary resistance. I will stress the evolutionary origin of signaling networks, the structural and functional features that confer robustness to therapeutic interventions, as well as the roles played by feedback regulation.

BioNetVisA workshop poster abstracts

Poster 9

Fully Bayesian Structural Inference for Bayesian Networks

Sagnik Datta, Eric Leclerc, Ghislaine Gayraud, Frédéric Y. Bois

UTC , INERIS, Compiègne, France

Background: In systems biology, difficulties regarding inference on gene regulatory pathways' structure arise because most microarray gene expression data sets are sparse compared to the number of genes involved. Various modelling and inference paradigms have been proposed in literature to minify this difficulty. Graphical models of gene networks have been useful in that context. Bayesian networks (BNs), graphical Gaussian models (GGMs) and relevance networks (RNs) are among the few widely used graphical models used in systems biology to draw structural inference. We present here a fast software for inference BNs.

Method: We consider inference on continuously valued BNs as a Gaussian regression problem where the regressands are children node and the regressors their parent. We use a fully Bayesian approach in which the marginal likelihood of the data (the data prior predictive distribution) and prior information about the network structure are considered. Independently priors can be specified for edge probabilities, degree distribution and the frequency of specific loop motifs. We use topological sorting algorithm to check the acyclicity of the proposed graphs. The proposed graphs are obtained only by addition or removal of edges during a systematic scanning of the adjacency matrix representation of the graph. A Metropolis-Hastings algorithm is used to accept or reject the proposal graphs. The C code written is very compact, fast, uses low memory and disk storage.

Results: We checked it on simulated data sets (of 5, 10, 20 and 40 nodes) and on a microarray gene expression data set corresponding to the KEGG apoptosis pathway (with 78 genes and two control nodes representing experimental conditions). We compare the capabilities and performance of our software with several others previously published.

Poster 10

Global Analysis of coregulation for the identification of functional modules

Rim Zaag[1], Etienne Delannoy[1], Marie-Laure Martin Magniette[1,2]

Plant Genomics Research, Evry, France[1], AgroParisTech, Paris, France[2]

One of the challenges faced by genomics currently is the understanding of gene function. Genome wide analysis of gene function mostly relies on guilt by association approaches through coexpression analysis taking advantage from the availability of transcriptome data. Generally co-expression is performed by analyzing correlations between all the gene pairs from multiple microarray experiments collected from international repositories. Such approach has two drawbacks: First it leads to a local point of view about functional modules and second the dataset is composed of heterogeneous transcriptome results.

In constrast, we performed a global analysis of highly homogeneous transcriptome data extracted from CATdb (Gagnot et *al*., 2008). The whole dataset is composed of more than 18 000 genes described by 424 expression differences dealing with stress conditions. The coexpression analysis is performed through a model-based clustering method.

Without a priori knowledge, the model has guided us to divide the whole dataset in twenty types of stresses leading to the identification of gene clusters having the same pattern of response under a single stress type. However coexpressed genes are not necessarily coregulated and then are less likely to be functional partners. To find groups of coregulated genes, we integrated these coexpression studies by calculating the occurrence number in a same cluster for each gene pair. Some pairs have a coordinated transcriptional response in up to 15 different types of stress and a resampling procedure showed that a gene pair observed in the same cluster in more than 4 stresses is significant. This approach allows us to focus our study on the potential key players of stress responses. Furthermore, the resulting coregulated gene network reveals an interesting topology of highly connected substructures. Preliminary analyses of these components containing orphan genes showed that they are more homogeneous than coexpression clusters highlighting probable functional modules.

Poster 11

Contribution of graphs to the analysis of bacterial genome architecture

Porion Olivier and Bénédicte Lafay

Laboratoire AMPERE, Ecole Centrale de Lyon, Écully, France

The genome of bacteria is classically separated into the essential, stable and slow evolving chromosomes and the accessory, mobile and rapidly evolving plasmids. Recently, a new class of genomic elements, neither chromosomes nor plasmids, has been revealed, the nature and evolution of which proves difficult to pinpoint. In the context of characterizing the different replicons forming the bacterial genomes, we investigated the analytical methodologies best able to decipher and visualize the functional and evolutionary relationships between replicons. Using databases of genes involved in genetic information transmission systems (GITS) as inputs, we performed a global comparative genomics analysis on all available bacterial genomes according to the following methodology: i) identification and assessment of functional homologs from all bacterial genomes, ii) construction of clusters of proteins linking the bacterial replicons by their GITS, with regard to the shared functions and protein sequence homologies, iii) testing of several unsupervised approaches (visualization, clustering). The comparison of these methods relied on the biological result assessment using the replicon host taxonomy and structure as external criteria as well as stability measurements as internal criteria. Bipartite graphs proved most useful for the meaningful representation of the replicons according to their GITS proteins. Community detection algorithms (INFOMAP) performed best in terms of stability with reference to the bacterial taxonomy in comparison to traditional clustering methodologies. Furthermore, our study brought about a dual functional and taxonomical structuration of the replicon space. This led to results with strong biological implications. Indeed, we were able to characterize the third class of replicons relative to chromosomes and plasmids, and to propose novel defining criteria for these genomic elements. Beyond the biological relevance, our study sets the basis for further analyses (workflow improvement/enrichment, classifications...) in order to bring to light driving forces of genome evolution.

Poster 12

Omix – A Visual Modeling & Data Visualization Tool for Biochemical Networks

Peter Droste

Omix Visualization GmbH & Co. KG, Lennestadt, Germany

Effective integrative visualization solutions bringing together data from experiment and simulation are of central importance to generate insights and new hypotheses. Thus, the interactive network-centric visual analysis of data sets has become an integral part of systems biology investigations. Omix® is a tool for customizable visualization of multi-omics data in the context of biochemical networks.

In Omix, diagrams can be drawn as in common graphics software but supported by automatic graph drawing and semi-automatic network layout techniques tailored to the application field of biochemical network. Here, the focus of Omix lies in providing a best serving user guidance.

Omix allows to visualize data in an unprecedented degree of flexibility. The so called *Omix Quick Visualizer* lets the user create sophisticated visualizations and animations in quick and easy dialog-based manner. Furthermore, visualization can be created by scripting because Omix allows programmable access to the visual appearance of the diagram components. By this, users can design the data visualization process according to individual requirements.

Multiple plug-ins are available extendin Omix with data management, interactive network analysis, database access (KEGG®, BioCyc) and modeling features. Kinetic models of biochemical processes can be created with Omix in graphical manner. Furthermore, a set of plug-ins extends Omix to a comprehensive modeling tool for isotope-based metabolic flux analysis.

Omix produces images and videos in widespread file formats. Furthermore, network models can be exported to SBML®, SBGN, 13CFLUX2, Matlab®, Dymola®, and more.

By providing easy access to data visualization and modeling, Omix is well suited as pre- and post-processing framework both, for experimenters and for modelers working interactively together in a rapidly evolving research field of systems biology and life sciences.

Supplementary information: www.omix-visualization.com

Poster 13

Multiscale mathematical modelling of breast cancer invasion

Arnau Montagud, Andrei Zinovyev, Emmanuel Barillot

Institut Curie, Paris, France; Ecole des Mines ParisTech, Paris, France; INSERM U900, Paris, France

Metastasis is a process that starts with invasion of surrounding tissue by tumour cells. It requires remodelling of the extra-cellular matrix and Epithelial-to-Mesenchymal Transition (EMT), loss of cell adhesion and polarity and increased motility. Understanding invasion mechanisms is crucial to improve prognosis and develop new cancer treatment strategies, but we still lack a detailed explanation of this process. In the past years several efforts have been done in systematising different mechanisms of cell migration, also termed invasion modes, and understanding their underlying causes.

We devised a mathematical model that would incorporate information of a series of traits, cellular and environmental, that would output in a set of invasion modes. For this, the model incorporates different pathways such as apoptosis, EMT determinants, cell cycle, tumour microenvironment-cell sensing, cell motility, extracellular matrix modification, etc. The resulting influence network is being translated into a mathematical model using discrete logical modelling. The model will be ready to be tuned by observed phenotypes on existing data from experimental results on tumours, cell lines and organoids.

Any realistic and useful mathematical model of tumour invasion must be multiscale as the process of invasion involves at least three levels of details: intracellular molecular processes determining individual cellular properties; interaction between a cell and its microenvironment affecting cell state and properties; and biochemical and biophysical interactions between cells in the context of tumour microenvironment, leading to various patterns of collective cell behaviour. Present work is part of a collaborative effort to model tumour invasion in order to identify treatment strategies and to understand underlying properties of metastasis.

Poster 14

Sparse factor models for gene co-expression networks

Yuna Blum[1], Magalie Houée-Bigot[2], Sandrine Lagarrigue[3] & David Causeur[2]

Department of Medicine/Division of Cardiology, David Geffen School of Medicine, UCLA, Los Angeles, CA USA [1], Agrocampus Ouest- Applied Mathematics Department, Rennes, France [2]. Agrocampus Ouest - UMR PEGASE INRA, Agrocampus Ouest, Rennes, France [3]

Inference on gene regulatory networks from high-throughput expression data turns out to be one of the main current challenges in systems biology. Such interaction networks are very insightful for the deep understanding of biological behavior and relationships between genes in a system. In particular, a functional characterization of gene modules enables the detection of biological processes underlying complex traits as diseases. Inference on this dependence structure shall account for both the high dimension of the data and the sparsity of the interaction network.
Extending the idea introduced for differential analysis by Friguet et al (2009) [1] and Blum et al (2010) [2], we suggest to take advantage of a low-dimensional latent linear structure of dependence to improve the stability of correlation estimations. We propose an L1-regularized EM algorithm to fit a sparse factor model for correlation. We demonstrate how it helps extracting modules of genes and more generally improves the gene clustering performance in comparisons with other competitors. As an alternative, we propose to use biological knowledge as a prior to infer on the sparsity pattern. The integration of prior biological knowledge based on Gene Ontology annotation is illustrated on a real dataset in the context of a lipid metabolism study.
Our method is implemented in the R package called FANet that includes the L1-regularized estimation using a cyclic coordinate descent algorithm, and a function that automatically integrates available biological annotation in the sparse factor model algorithm.

References
[1] C. Friguet, M. Kloareg and D. Causeur. A factor model approach to multiple testing under dependence. Journal of the American Statistical Association, 104(488): 1406-1415, 2009.
[2] Y. Blum, G. Le Mignon, S. Lagarrigue and D. Causeur. A Factor Model to Analyze Heterogeneity in Gene Expression. BMC Bioinformatics, 11(1): 368, 2010.
[3] P. Langfelder, and S. Horvath. WGCNA: an R package for weighted correlation network analysis. BioMed Central Ltd, 9(1): 559, 2008.

Poster 15

Identification of microRNAs, isomiR variations, target networks and differential expression visualization from NGS datasets

Dilyan Dimitrov, Ivan Milev, Georgi Minkov, Galina Yahubyan and Vesselin Baev

Dept. Plant Phys. and Mol. Biol, University of Plovdiv, Bulgaria

MicroRNAs (miRNAs) are a class of small non-coding RNAs, which are negative regulators of gene expression in human. miRNAs frequently exhibit differences from their “reference” mature sequences, generating multiple variations that are called as “isomiRs”. In many cases both mature miRNA and corresponding isomiRs were present in the NGS libraries. IsomiRs variability can be explained by the imprecise and alternative cleavage of Dicer and Drosha during pre-miRNA hairpin processing. We have developed a tool allowing identification of miRNAs in 168 species (including human) using next-generation sequencing datasets. Moreover our software can detect and visualize isomiRs of miRNAs with higher copy number relative to their mature reference sequences indexed in miRBase. Therefore an observed specific signature overabundance of isomiRs in the human samples can imply a potential role in the aberrant regulation of protein-coding genes targeted by miRNAs within the cancer or disease state. As additional information the software generates differential expression information charts of the most dys-regulated miRNA between samples, as well as link to the miRNA targets network datasets.

Poster 16

Streaming Visualisation for Raw Mass Spectrometry Data Based on a Novel Compression Algorithm

Yan Zhang [1,2], Henning Hermjakob [3] and Andrew W. Dowsey [1,2]

[1] *Institute of Human Development, University of Manchester.*
[2] *Centre for Advanced Discovery and Experimental Therapeutics (CADET), University of Manchester and Central Manchester Foundation Trust, Manchester Academic Health Sciences Centre.*
[3] *European Bioinformatics Institute, Hinxton, Cambridge*

Mass Spectrometry (MS) has become a pervasive technique for the analysis of biological compounds. Vast amount of data are generated using high-throughput LC-MS (Liquid Chromatography–MS). To keep pace with this data explosion, great endeavours have been made in the bioinformatics field using more intricate and automated analyses. There is a danger that the process becomes more and more opaque and inaccessible to MS practitioners. It is therefore vitally important that efficient visualisation tools are available to facilitate quality control, verification, validation and interpretation of raw MS data and MS analyses.

MS data converted to the standards-compliant Proteomics Standards Initiative mzML format are organised as a contiguous list of raw spectra. It is therefore fast to recall individual spectra due to the indexing scheme and their relatively small size. However, generating a 2D overview ('virtual gel') of an LC-MS dataset requires every single datapoint to be loaded, which takes significant time and memory space. Also, 2D panning and zooming for MS data stored in those formats is inefficient. In order to tackle these issues, we leverage a novel compression algorithm developed within our seaMass framework. This transform compressor employs a complete set of separable 2D multiscale cubic B-spline basis functions as a sparse domain; this models the raw LC-MS data as the non-negative weighted sum of a collection of overlapping tensor-product B-spline basis functions. The decompression operation is simply the sum of those B-spline building blocks and can be conducted on the fly. This scheme is able to achieve lossless compression with high compression rates. By adjusting the shrinkage parameter, MS data can be de-noised and compressed with compression ratio higher than 200:1.

In MS data, peaks are cognitively most important. To organise compression coefficients both spatially and ordered by intensity, an Octree data model is employed. Octrees are most often used to represent objects in a 3D space. Retention time, mass-to-charge ratio, and quantised coefficients are selected as the three dimensions for our MS Octree. This Octree creates a hierarchy of Axis Aligned Bounding Boxes, for visualisation which is used to cull away coefficients outside of view, and retrieve coefficients in a sorted order by visual importance. Data saved in the Octree can be streamed in the order of their importance, and the visualisation iteratively refined: early approximate reconstructions will permit fast response to initiation, panning and zooming, while later reconstructions will reveal further details.

Underpinned by the compression scheme and the efficient data organisation model, a prototype system has been developed to stream MS datasets at any desired resolution. This visualisation engine is fused with detailed annotations of identification and quantification results, therefore providing a complete dissemination system for the mass spectrometrist.