

# ECCB 2014 Accepted Posters with Abstracts

## C: Pathways and molecular networks

**C01:** Inna Kuperstein, Simon Fourquet, Jean-Marie Ravel, Emmanuel Barillot and Andrei Zinovyev. A comprehensive map of programmed cell death signalling network: an analytical tool for studying regulation of different modes of cell death in human disorders

**Abstract:** Introduction: Cell death process draws special attention not only because it represents one of the fundamental mechanism in the cell, but also due to frequent perturbations in various diseases including cancer and neurodegenerative disorders. There are two major subclasses of cell death the programmed and the non-programmed cell death mechanisms. Lately it became evident that the programmed cell death (PCD) includes three major modes, apoptosis, necroptosis and autophagy. The decision about the choice of death mode is regulated by several upstream signalling pathways. In addition these three modes of programmed cell death are mutually dependent and cross-regulate each other.

Methods: The massive knowledge about cell death mechanisms mainly exists in the form of thousands of scientific papers. To understand and to study orchestration between triggering signalling to molecular mechanisms of cell death, systematic and formalized representation of this information together is needed. We used a systems biology approach to represent graphically PCD mechanisms as a seamless comprehensive map of signalling network that is accessible for computational analysis.

Results and discussion: Based on experimental data retrieved from literature by manual curation, we have constructed an integrated signalling network of PCD. The map depicts biochemical triggers and upstream signalling as death receptors; mitochondria, glucose metabolism and DNA damage in the initiation of different modes of PCD. An addition layer of cell death signalling regulation by miRNAs is described on the map as well. The molecular mechanism of each PCD mode, namely apoptosis, necroptosis and autophagy is represented in details. To facilitate readability and usage, the map is divided into functional modules that can be visualized in the context of the whole map or as individual diagrams. The map is open source platform available on the web. The navigation, curation by the community and analysis of the map is facilitated by NaviCell web-tool (<http://navicell.curie.fr>). We describe the possibilities of PCD map applications in clinical research for data visualization and analysis. We demonstrate visualization and comparison of of genome-wide data as transcriptome, proteome and phosphoproteome from neurodegenerative disorders and cancer in the context of the PCD map.

Conclusion: The network-based approach can be a framework for deciphering differences in complex molecular characteristics of PCD regulation along various diseases. A comprehensive reconstruction of PCD signalling will bring a better understanding of regulatory circuits in different human disorders and can be applied for therapeutic target identification and for treatment approaches optimisation.

**C02:** Yvonne Mayer, Karin Zimmermann, Berit Haldemann, Dido Lenze, Michael Hummels and Ulf Leser. Inclusion of miRNAs improves differential network analysis

**Abstract:** Differential network analysis (DNA) denotes a novel class of algorithms which focus on the differences in network topologies between two states of cells, such as healthy and disease or wild-type and pertubated, to identify key players in the underlying biological processes. In contrast to conventional differential analysis, DNA identifies changes in the

interplay between molecules, rather than changes in single molecules or in groups of molecules, which is especially important in complex diseases caused by mutations in transcription factors which lead to a rewiring of interactions or to wide-spread (yet often small) changes in gene regulation. To fully capture the changes in network topology, DNA should consider all cellular entities that influence the dynamics of the network under study. Here, we study the impact of including miRNA, an important class of mostly negative effectors in the regulatory machinery of mammalian cells, into human regulatory networks prior to their analysis by DNA methods. To this end, we constructed high-quality regulatory networks with and without miRNA and quantified the strength of relationships using gene expression data from four different cancer types (prostate adenocarcinoma, breast invasive carcinoma, liver hepatocellular carcinoma, lung adenocarcinoma). We then compared the ability of ten DNA algorithms to recover key players in the miRNA-free networks to those in the miRNA enriched networks. We find that the inclusion of miRNA consistently and significantly improves the performance of each tested DNA method, underlining the importance of using more comprehensive models of cells for network analysis.

**C03:** Antonio Fabregat, Joel Weiser, Steven Jupe, Phani Garapati, Oscar Forner, Pablo Porras, Robin Haw, Peter D'Eustachio, Lincoln Stein and Henning Hermjakob. Reactome: Pathway Analysis Tool Suite

**Abstract:** Reactome is a free, open-source, curated and peer-reviewed knowledgebase of biomolecular pathways. One of its goals is to provide intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic research, genome analysis, modeling, systems biology and education. Following the successful release of the new Pathway Browser web application we have continued to improve the suite of analysis tools. In response to user feedback and new features requests the tools have been significantly improved. Moreover the Pathway Browser and the interface for the analysis tools have been integrated to enhance the user experience. This presents the analysis submission interface and the results display directly within the Pathway Browser. The analysis tool suite contains an overrepresentation analysis, an expression data analysis and a species comparison tool. Results are presented in the Pathway Browser in tabular form within the details panel, as an overlay on pathway diagrams and hierarchy. Besides improving stability, performance, reliability and scalability to cope with future demands, we have also extended functionality. For example in the previous version results were only available for the top level of Reactome's pathway hierarchy while the new version provides results for all levels. Additionally a new feature has been developed, a topology-based analysis offering a score calculated by matching a submitted dataset to reactions as units of pathways. The performance has been boosted by creating a data structure to model Reactome objects and relationships which is directly placed in memory. The data structure is composed of double-linked trees to model the hierarchy, a radix tree for the identifiers used as cross-references to the target entities and a graph to model these entities and the relationships between them (also the species orthology). In order to support computational access for external tools and pipelines the functionality of the analysis tools is available via state-of-the-art web services. Results tokens generated following computational access can later be used to link to the Pathway Browser to benefit from Reactome's pathway overlay.

**C04:** Jeanne Cambefort, Guillaume Collet, Sylvain Prigent, Simon Dittami, Olivier Dameron, Thierry Tonon and Anne Siegel. AuReMe: an integrative method for Reconstruction of Metabolic networks including Automatic Gap-Filling

**Abstract:** Introduction: To understand the physiology of an organism for which its genome has been sequenced, a genome-scale metabolic network (GEM) is reconstructed. Draft GEMs may be produced either from genome annotations (Karp et al. 2010) or from orthology features (Loira et al. 2012). Nonetheless, the next steps towards a functional GEM, gap filling and validation, often involve tedious manual curation (Thiele & Palsson, 2010) and are not easy to automatize. To overcome this, we introduce AuReMe (Automatic Reconstruction of Metabolic networks) that combines several drafts GEMs, proposes biochemical reactions to fill gaps in incomplete metabolic pathways, and provides a validation score through HMM profiles. All data gathered are then visualized on a wiki website that enables automatic updates and manual modifications.

Method: The AuReMe workflow consists in 4 tools and an HMM validation. Its inputs are several draft GEMs obtained from various tools and based on several genomic and metabolic resources with identifiers from different namespaces. Firstly, AuReMe combines distinct identifiers thanks to the many cross-database references available in MetaCyc. Once each draft GEM is mapped on MetaCyc, GEMs are merged into a metabolic network using the MeMerge tool. The resulting unified draft GEM must be not functional, i.e. it does not explain how biomass is produced from the growth media. Given the composition of the growth media and a set of evidenced metabolites, including those from the biomass composition, the Meneco software (Schaub & Thiele, 2009) proposes sets of biochemical reactions to complete the unified draft GEM and make it functional. Each set of reactions restores the capability of the metabolic network to produce biomass, according to a qualitative and topological criterium. To reduce the number of added reactions and to avoid redundancies, the different sets are semantically filtered to retain only biologically relevant reactions. Finally, when enough protein sequences are available for a given biochemical reaction extracted from MetaCyc, an HMM profile is generated by the HMMER suite. The proteome of the organism of interest is then screened with each HMM profile. This analysis provides an e-value that can be considered to validate the gene-reaction associations.

Results: AuReMe has been successfully applied to reconstruct the *E. siliculosus* metabolic network (EctoGEM). The combination of draft GEMs provided a draft GEM containing 1785 reactions and 1981 metabolites. A Meneco gap-filling revealed that this draft GEM was able to produce 25 metabolites among the 50 identified by experiments. The production of all metabolites and of biomass was restored by the addition of only 55 reactions. Moreover, a Flux-Balance-Analysis confirmed the functionality of EctoGEM. In addition, the wiki, <http://ectogem.irisa.fr>, allowed experts to re-annotate 51 genes and revealed novel insights into the evolution of aromatic amino acid synthesis.

**C05:** Michaela Bayerlova, Florian Klemm, Annalen Bleckmann, Frank Kramer, Tobias Pukrop and Tim Beissbarth. Integration of breast cancer RNA-Seq data into newly constructed WNT signaling networks

**Abstract:** The WNT pathway is comprised of distinct subsets of signaling pathways with an important role in embryonic development and carcinogenesis. WNT signaling in general is highly active in the molecular basal-like subtype of breast cancer and in all breast cancer patients which later develop metastasis to the brain. Our previous results indicate that non-canonical, i.e.  $\beta$ -catenin independent, WNT signaling is of importance in this context. To further validate this finding, we generated and sequenced MCF-7 and MDA-MB-231 breast cancer cell lines with perturbation of several non-canonical WNT ligands and receptors. RNA-Seq data were aligned to the transcriptome using STAR tool and gene-level abundances were estimated by RSEM algorithm. We identified differential genes between different conditions by fitting negative binomial generalized linear models implemented in edgeR.

To integrate the results of RNA-Seq experiments into the context of prior WNT knowledge, we scanned major pathway databases (Reactome, PID, KEGG, BioCarta and Pathway Commons) for WNT pathways and other pathways linked to WNT signaling. Using the rBiopaxParser tool, we downloaded BioPAX exports of these databases and parsed 47 WNT pathways into environment of R as directed graphs. Gene Symbol IDs were mapped onto graphs nodes and non-gene nodes were filtered out. Finally, the pathway graphs were classified and merged into 4 networks representing  $\beta$ -catenin dependent WNT signaling,  $\beta$ -catenin independent WNT signaling, inhibition of  $\beta$ -catenin dependent WNT signaling and regulation of WNT signaling.

In the next step, we integrated differential genes with their fold changes from RNA-Seq experiments into the WNT networks. Through this approach we were able to extract differential sub-networks for distinct cell line perturbations and identify key signaling players. These sub-networks will be further applied to external patient data of both breast cancer primaries and brain metastases to test their prognostic power in a clinical setting.

**C07:** Sylvain Bournais, Pauline Gloaguen, Gilles Curien, Christophe Bruley, Florence Combes, Marianne Tardif, Yves Vandenbrouck, Giovanni Finazzi, Myriam Ferro and Norbert Roland. Towards the virtual chloroplast

**Abstract:** The chloroplast is a complex and integrated metabolic network that produces a high number of metabolites of industrial interest (sugars, vitamins, lipids and pigments). 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 still 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 proteins and metabolites.

Thus we decided to create a virtual chloroplast, by manually and automatically integrating all the qualitative and quantitative data currently available. It will contain a user-friendly interface allowing rapid 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 the software CellDesigner. These maps have been integrated into a web interface providing direct link with biological and bibliographical 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. Each component of a given map is linked to their description page. Furthermore, every map is connected with each-other allowing to follow a metabolite from one metabolism to another. 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 would like to use the dynamic aspect of the SVG format (Scalable Vector Graphics; possibility to create moving pictures) to visualize fluxes and quantities variations in this metabolic network. A first release of this knowledge base available to the community is expected by the end of 2014.

**C08:** Patrice Baa-Puyoulet, Augusto F. Vellozo, Jaime Huerta-Cepas, Gérard Gérard Febvay, Federica Federica Calevro, Marie-France Sagot, Hubert Charles, Toni Gabaldon and Stefano Colella. Annotating arthropods genome to study and compare their metabolism: the ArthropodaCyc collection of Cyc databases powered by CycADS

**Abstract:** Several arthropods genomes have been and are being sequenced (e.g. the i5K sequencing initiative [1]) and these data open the way to comparative studies of different species to better understand their biology. All comparative genomics studies rely heavily on the quality of genome annotations. To be useful to researchers genomics data have to be collected from various sources, updated regularly and organized in dedicated databases. During the genome annotation for the pea aphid (*Acyrtosiphon pisum*) we developed CycADS [2] (Cyc Annotation Database System), an automated annotation management system that allows the seamless integration of the latest sequence information into metabolic networks reconstruction. Specific genomic data, as well as their functional annotations obtained using different methods (such as KAAS, PRIAM, PhylomeDB, Blast2GO, MetaPhOrs, Interproscan), are collected into a SQL database and later extracted, with the possibility to apply different quality filters. CycADS allows the automatic generation of a complete set of input files to build or update BioCyc databases [3] using the Pathway Tools software. We used CycADS to create “ArthropodaCyc “ [4], a collection of arthropods metabolic network Cyc databases, which contains 23 organisms to date. Our collection of databases allows metabolic pathways visualisation, and each protein page includes information about the annotation methods used, as well as hyperlinks to genome specific resources. Comparison using interactive web functionalities, as well as user “omics” data mapping or information extraction, are also available in the BioCyc interface of ArthropodaCyc. Future plans include the addition of other sequenced genomes after their publication and/or in collaborations with arthropods genome sequencing projects. Further developments will also include the implementation of gateways to network analysis tools. [1] <http://arthropodgenomes.org/wiki/i5K>; [2] <http://www.cycadsys.org/>; [3] <http://biocyc.org/>; [4] <http://arthropodacyc.cycadsys.org/>

**C09:** Fazle Faisal and Tijana Milenkovic. Dynamic networks reveal key players in aging

**Abstract:** Motivation: Since susceptibility to diseases increases with age, studying aging gains importance. Analyses of gene expression or sequence data, which have been indispensable for investigating aging, have been limited to studying genes and their protein products in isolation, ignoring their connectivities. However, proteins function by interacting with other proteins, and this is exactly what biological networks (BNs) model. Thus, analyzing the proteins' BN topologies could contribute to understanding of aging. Current methods for analyzing systems-level BNs deal with their static representations, even though cells are dynamic. For this reason, and because different data types can give complementary biological insights, we integrate current static BNs with aging-related gene expression data to construct dynamic, age-specific BNs. Then, we apply sensitive measures of topology to the dynamic BNs to study cellular changes with age.

Results: While global BN topologies do not significantly change with age, local topologies of a number of genes do. We predict such genes as aging-related. We demonstrate credibility of our predictions by: 1) observing significant overlap between our predicted aging-related genes and "ground truth" aging-related genes; 2) observing significant overlap between functions and diseases that are enriched in our aging-related predictions and those that are enriched in "ground truth" aging-related data; 3) providing evidence that diseases which are enriched in our aging-related predictions are linked to human aging; and 4) validating our high-scoring novel predictions in the literature.

**C10:** Kaveh Pouran Yousef, Adam Streck, Heike Siebert and Max von Kleist. Analysing the c-di-GMP-dependent bistable regulation of curli fibers in *Escherichia coli* by combining boolean, continuous and stochastic modelling

**Abstract:** Background: Interactions between different components in signalling networks can indirectly be inferred using a combination of wet lab experiments. These usually measure downstream components of the network (e.g. a reporter gene assay describing the output of the signalling cascade). If the expression of the readout gene of interest obeys an all-or-nothing principle, giving rise to phenotypic multistability, then simple input and output relations are usually not given. This complicates the interpretation of experiments and network/model inference.

Aim: Here, we present a methodology aimed at the analysis of multistability in a signalling network responsible for the stress-mediated expression of biofilm in *Escherichia coli*. Specifically, we focus on investigating the role of the second messenger c-di-GMP in controlling bistable expression of curli fibers, which represent a key structural component of bacterial biofilm. The aim is to infer the structure of the underlying signalling network and to identify feasible parameterizations satisfying the observations from a set of genetic knockout experiments.

Methods: In the first step we build a logical model by making a minimal set of assumptions about the underlying biological system. We conduct model checking in order to verify that the logical network structure is compliant with the genetic knockout data. Afterwards we identify continuous kinetic models which exhibit a qualitatively equivalent multistable behaviour as the logical network by inspecting the corresponding continuous parameter space. Lastly, the candidate parameterizations are probed in terms of their dynamic properties (e.g. switching times) using stochastic simulation.

Results: We find a small set of logical models which simultaneously satisfy multistability constraints and match the expression states of various single- or multiple-gene knockout data of the curli regulation system. We then assign kinetic rate laws to the identified interactions (the wiring of the network) by incorporating the details of the underlying biochemical reactions. A parameterization for the kinetic model is then found that satisfies the constraints resulting from the logical modelling step and from experimental data.

Our final model provides a first dynamical explanation for the regulatory mechanisms behind the bistable expression of biofilm in *E. coli* and yields an insight into the noise-induced switching behaviour between the curli on/off states.

In this case study, we presented a structured workflow for simultaneous model and parameter inference in stochastic, multistable systems by a combination of different complementary modelling techniques. Further research may be required in order to devise a broadly applicable, generic methodology.

**C11:** Sun Sook Chung, Alessandro Pandini, Alessia Annibale, Anthony C. C. Coolen, Nicolas Shaun B. Thomas and Franca Fraternali. Topological analysis of protein interaction networks: the importance of loop motifs and their biological implications

**Abstract:** Protein-protein interaction networks (PPINs) have been used to identify potential novel interconnections between proteins involved in particular biological processes. However, such PPINs can be biased by the experimental methods used for detecting the interactions. They often contain false positives and negatives, which limit their usefulness [1]. In addition, some of the analytical methods used to create PPINs are flawed as we do not understand the rules governing PPINs.

This project aims to formulate fundamental principles of PPIN construction by analysing network motifs of short loops and small cyclic interactions of between 3 and 5 proteins. Network motifs are repeated patterns, e.g. feedback loops are known to be present in many biological systems [2]. In this study, we started by analysing 30 PPINs based on large-scale datasets [1] obtained by different experimental methods covering 9 species from bacteria to homo sapiens. We analysed unique features of the network, based on the number of loops and

cyclic interactions. These analyses were compared with unbiased model networks based on randomised null model analyses [3]. Beyond statistical analysis, we compared experimentally verified data of 622 human protein complexes [4]. This enabled us to investigate functional consensus, defined as a protein function which occurs in common in the same loop of defined length.

The study highlights short loops as an essential feature of PPINs in two respects. Firstly, the number of short loops can be a measure to quantify the reliability of PPINs. Most short loops share at least one biological function but some functions are enriched according to the length of the loops. We find the proteins involved in metabolic processing of mRNA and organization of cellular components are enriched in short loops. Our approach analysing short loops in PPINs has identified topological properties of sub-networks that are involved in specific cellular mechanisms. This indicates that PPINs involved in different biological processes may be structured and regulated in different ways.

1. Fernandes, L.P., et al., Protein networks reveal detection bias and species consistency when analysed by information-theoretic methods. *PloS one*, 2010. 5(8): p. e12083.
2. Carlin, L.M., et al., A targeted siRNA screen identifies regulators of Cdc42 activity at the natural killer cell immunological synapse. *Sci Signal*, 2011. 4(201): p. ra81.
3. Annibale, A., et al., Tailored graph ensembles as proxies or null models for real networks I: tools for quantifying structure. *J Phys A Math Gen*, 2009. 42(48).
4. Havugimana, P.C., et al., A census of human soluble protein complexes. *Cell*, 2012. 150(5): p. 1068-81.

**C13:** Berit Haldemann, Daniel Heinze, Michael Hinz, Claus Scheidereit and Ulf Leser. A Comprehensive Approach to the Characterization of Transcriptional Regulation of DNA Damage Responses

**Abstract:** DNA damage is an alteration of the DNA structure implicated with many diseases. It can be caused, for instance, by dysregulated metabolic processes or toxic environmental factors. Due to their potentially fatal consequences, cells harbor sophisticated endogenous response and repair mechanisms to efficiently identify and eliminate such alterations. As part of the DNA damage response, transcription factors (TF) are activated, which subsequently regulate cell fate decision programs, such as survival, cell-cycle arrest, apoptosis or senescence.

To understand better the underlying processes of transcriptional regulation upon DNA damage, we performed genome-wide time-series microarray analysis in human osteosarcoma cells (U2OS), treated with ionizing radiation. Interdependence between expression intensities of genes at consecutive time points, hinting to regulatory relationships, was analyzed using three complementary approaches. Firstly, we collected TF-TF and TF-target gene interactions from different data resources and used these to filter such relationships identified by differential expression analysis. Secondly, we scanned promoter regions of differentially regulated genes for TF binding sites using motif databases and public experimental TF binding information. Thirdly, we applied TF-network reconstruction approaches to model the underlying regulatory relationships.

We integrated the results of the three approaches to get an overall view on the transcriptional regulation processes initiated by DNA damage. Our study expands the knowledge about mechanisms behind early genotoxic stress response and serves as basis for the exploration of the crosstalk between the identified transcriptional master regulators.

**C14:** Rafael S. Costa, Nguyen Hoang Son and Susana Vinga. Comparison of cellular objectives in flux balance constraint-based models

**Abstract:** Genome-scale reconstructions are usually stoichiometric and analyzed under steady-state assumption using flux balance analysis (FBA). FBA require not only the stoichiometry of the network, but also an appropriate cellular objective function and possible additional physico-chemical constraints to predict the set of resulting flux distributions of an organism.

The most common objective assumption is to consider the optimization of the growth rate or yield. However, other objectives may be more accurate in predicting phenotypes and caution should be taken with the selection of optimization principles. So far, original publications rarely compare predicted internal flux states and measured fluxes to identify the principal cellular objective. With the recent availability of experimental high-throughput omics data it is possible to test the use of an objective function, comparing predicted with experimental fluxes and analyze on a case-by-case basis to find the right predictive objective functions and additional constraints. Since objective function selection seem to be, in general, highly depend on factors such as growth conditions and quality of the constraints, more research should be pursued for better understanding the universality of the objective function.

In this work, we explore the validity of different classes of optimality criteria and the effect of single (or combinations) standard constraints in order to improve the predictive power of intracellular flux distribution. Additionally, the simultaneous optimization (multi-objective) of two or more cellular functions was examined. These comparisons were evaluated to compare predicted fluxes to published experimental <sup>13</sup>C-labeling fluxomic datasets using two different metabolic systems with different conditions and comparison-data sets than previous evaluations (Schuetz et al. Mol Sys Biol 2007, Ow et al. AIChE 2009, Knorr et al. Bioinformatics 2007).

It was observed that by using different conditions and metabolic systems, the fidelity patterns of FBA can differ considerably. However, despite of the observed variations, several conclusions could be extracted. First, in agreement with previous studies, the single objective of maximization of biomass or ATP yield achieves the best predictive accuracy. Moreover, under this condition the flux states are also well described by the objective of minimization of reaction fluxes across the network. For the batch growth condition the most consistent optimality criteria is described by maximization of the biomass or ATP yield per flux. Secondly, the predictions obtained by flux balance analysis using additional combined standard constraints are not necessarily better than those obtained using only the single constraint.

Work partially supported by FCT, Portugal under contract PEst-OE/EME/LA0022/2011, IF/00653/2012, SFRH/BPD/80784/2011 and European Union, FP7 BacHBerry Project No. FP7- 613793.

**C15:** David Hill, Tanya Berardini, Peter D'Eustachio, Harold Drabkin, Chris Mungall and Judith A. Blake. Modeling Glycolysis in The Gene Ontology: All Roads Lead to Pyruvate

**Abstract:** The Gene Ontology (GO) is a freely-available resource that provides connections between gene products and a structured, controlled vocabulary of biological terms used to describe how the gene products function. The representation of biochemical pathways presents a challenge because different species may utilize different enzymatic activities or different substrates to carry out similar processes. Glycolysis is an example of an overall process that is well conserved among species, but varies with respect to input molecules and enzymatic activities. We describe a strategy to model glycolytic processes that accounts for variation seen in different biological contexts. To accomplish this, we factor out core, conserved processes and represent variations as subtypes. Glycolytic processes are defined by axioms that include the parent superclass process, the molecular functions that are necessary parts of the process and the input and output chemical entities. We group the conserved



processes by shared intermediate substances, such as a shared glucose-6-phosphate intermediate. For example the process ‘canonical glycolysis’ is the glycolytic pathway from glucose to pyruvate and is modeled as a subclass of ‘glycolysis through glucose-6-phosphate’ with an input of glucose and necessary glucokinase activity. ‘Canonical glycolysis’ inherits all of the molecular functions required for the execution of its superclasses and inherits the output of pyruvate from the generic superclass ‘glycolytic process’. Creating formal definitions for pathways allows use of OWL-based reasoning for further inferred classification. For example, since ‘glycolytic process’ is a ‘catabolic process’ with input carbohydrate and ‘canonical glycolysis’ is a type of ‘glycolytic process’ with input glucose, ‘canonical glycolysis’ is automatically classified as a type of ‘glucose catabolic process’. The inclusion of necessary molecular functions not only aids ontology maintenance but also guides curators in determining which genes are directly involved in a process. We use glycolysis in mouse sperm to show how curators can use the necessary functions associated with a glycolytic process to infer a subtype for annotation and how users can explore annotations to determine which isoforms of an enzyme are used in a given biological context. This work was supported by National Human Genome Research Institute (NHGRI) grants U41HG002273 to the GO Consortium and U41HG003751 to the Reactome Knowledgebase.

**C16:** Eva Strakova, Jan Bobek, Alice Zikova, Klara Novotna and Jiri Vohradsky. Inference of sigma factor controlled networks in germinating prokaryote

**Abstract:** Streptomyces have primarily been studied as producers of antibiotics and as a model organism for investigating antibiotic resistance and cell differentiation. Most of the transcriptional control in Streptomyces is governed by a large number of sigma factors; the regulators binding the promoter region of a gene and initiating transcription. We focused on a systems level inference of transcriptional networks controlling *S. coelicolor* germination. Using a numerical model of gene expression, we reconstructed the genetic networks and functional groups of genes controlled by the sigma factors.

**C17:** Stefan Kroeger, Melanie Venzke, Ria Baumgrass and Ulf Leser. Meta expression analysis of regulatory T cell experiments for gene regulatory network reconstruction

**Abstract:** 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.

**C18:** Azzurra Carlon, Barbara Di Camillo, Federica Eduati and Gianna Toffolo. A rule based modeling approach to insulin signaling pathway analysis: signal amplification and robustness

**Abstract:** Motivation: Biological systems, such as signaling pathways, act sensing input stimuli and transmitting and processing this information to provide output signals regulating different cellular activities. To understand the input-output relationship of a complex biological system, mathematical models are often used, thus providing a description of the system behavior and allowing useful analysis of its emerging properties.

In this work, we implemented a comprehensive model of the insulin signaling pathway (ISP) using Rule Based Modeling (Smith AM, 2012), able to deal with all molecular interactions typical of a signaling pathway in a more efficient and compact way, and we explored its dynamics features using the parametric sensitivity analysis (PSA).

Methods: Three different models available from the literature describing: 1) insulin receptor binding and recycle systems (Sedaghat et al., 2002); 2) PI3K-AKT pathway including translocation of transporters GLUT4 on cell membrane and TSC1/2-mTOR negative feedback (Sonntag et al., 2012); 3) RAS-MAPK pathway including activation of transcription factor ERK1/2 (Borisov et al., 2009), were integrated into a single model. Parameter values were either estimated from in-house experimental data or inferred from the literature.

For each protein within the ISP, time-varying sensitivity coefficients were computed as the ratios between the change in a biological model output and the perturbation on one or more system parameters.

Results: A detailed analysis of time-varying coefficients was conducted for proteins at the end points of the ISP model (GLUT4 and ERK1/2) and in the negative regulatory loop (mTOR) permitting to cluster the time-varying coefficients according to their time-dependent behavior. Sensitivity coefficients integrated over the prediction time as a measure of the accumulated effect were used to provide a general ranking of the most sensitive parameters and concentrations confirming the crucial role of phosphatases within ISP.

To explore the role of the main negative regulation within the ISP model, the negative feedback loop involving p70S6K-IRS1 was removed from the model. Although the strong similarity between the concentration profiles obtained from the original and the modified model, a 10-fold increase of the integrated sensitivity coefficients was observed, meaning that, without the p70S6K-IRS1 feedback loop, the system loose robustness to possible changes in the parameters. The removal of negative feedback loop involving MAPK-GRB2/SOS or of the crosstalk involving AKT-RAS gave similar results, thus confirming the fundamental role of these key network motifs in determining not only the dynamic behavior of the system but also its robustness against small biological fluctuations due to intercellular variability.

**C19:** Dmitry Ravcheev and Ines Thiele. Genomics analysis of the respiratory capacities of the human intestinal microbiota

**Abstract:** Due to the anatomical and physiological properties of the human intestine, a specific oxygen gradient builds up within this organ, which influences the intestinal microbiota. The intestinal microbiome has been intensively studied during the last years.

Nonetheless, the respiratory capacities of the gut microbiota have been investigated for only a limited number of model organisms. Here, we present the systematic analysis of the genes connected to the respiratory systems in the genomes of human gut habitants.

We studied 259 genomes of microorganisms habiting in the human gut, which belonged mostly to the phyla Firmicutes, Bacteroides, Proteobacteria, Actinobacteria, and Fusobacteria. We investigated the distribution of known respiratory terminal reductases and also predicted a number of novel enzymes with respiratory function using comparative genomics analysis, and genome context-based approaches. For instance, we predicted a novel microaerobic reductase for *Faecalibacterium prausnitzii*. Orthologs for this reductase were found in 14 additional genomes. Based on the distribution of the known and predicted aerobic reductases, 31 of the studied microbes were identified as aerobic habitants of the intestinal wall and the adjacent areas. 112 organisms were identified as either micro- or nanoaerobic. The genomes of these organisms did not contain genes for any regular aerobic reductase but we predicted the presence of high-affinity reductases, characteristic for microaerobically living bacteria. The other 116 microbes were classified as obligatory anaerobes based on their genomes.

Our analysis of the studied genomes further revealed anaerobic reductases for tetrathionate, thiosulfate, polysulfide, sulfite, adenylyl sulfate, heterodisulfides, fumarate, trimethylamine N-oxide, dimethyl sulfoxide, nitrate, nitrite, nitrogen oxide, nitrous oxide, selenate, and arsenate. No genes for respiration by chlorate, perchlorate, or metals were found. Among the enzymes for anaerobic respiration, nitrate and nitrite reductases became the most respectively. Surprisingly, 98 studied genomes lacked any of the analyzed respiratory systems.

The distribution of enzymes for nitrate and nitrite respiration allowed us to propose a cross-species interaction network through electron acceptors between different anaerobic bacteria. For examples, the genomes of *Lactobacillus* spp. contained genes that encode enzymes for nitrate respiration, whereas the *Bacteroides* spp. genomes contained only nitrite respiration genes.

We illustrate in this study that knowledge regarding the respiratory capacities of the human microbiome can substantially expand our understanding of this important microbial community. The current work represents the first ontology for respiratory enzymes in the microbiome of the human intestine. The obtained results can be useful to guide experimental and computational analysis of the human microbiome to unravel its role in human physiology and health.

**C20:** Patrick Trampert, Tim Kehl, Daniel Stöckel, Hans-Peter Lenhof, Andreas Keller and Christina Backes. GeneTrail2 - A statistical analysis tool for molecular signatures

**Abstract:** Over the past decade high-throughput methods have been used to generate a tremendous amount of ‘omics’ data. Analyzing this data is a major task in computational biology. Especially, the exploration of molecular signatures is important to study differences between biological conditions, e.g. discovering pathological mechanisms. Methods like Overrepresentation Analysis, Gene Set Enrichment Analysis, max-mean-statistics, and others are regularly applied. Many tools have been published that provide such methods. However, no gold standard has been established to evaluate the performance of existing methods and most tools turn out to be restricted to specific tasks and scenarios. The choice of appropriate tools is not always obvious and can be challenging, especially for life scientists that are typically not familiar with the underlying statistics. We present GeneTrail2, a statistical analysis tool for molecular signatures - a major upgrade of the GeneTrail web-service from Backes et al. that has been part in over 40.000 analyses in the past five years. Additional to classical microarray expression experiments we support the analysis of miRNA and SNP datasets. GeneTrail2 provides a comprehensive collection of gene set enrichment methods combined with reference categories derived from popular databases like KEGG, GO, or

Reactome for many key organisms like Homo sapiens, Mus musculus, or Arabidopsis thaliana. GeneTrail2 offers two interfaces suited for different demands: One interface that focuses on the task itself. A user only has to specify the kind of analysis to be conducted. Further steps are carried out automatically with reasonable default parameters that have been selected carefully. The other interface guides the user through the process of data analysis by an easy to use step-by-step procedure. This gives users the possibility to adapt each step of the pipeline to their very needs. Automated steps like identifier type and organism detection as well as a comprehensive documentation reduce the user's additional workload. Additionally, all implemented methods are accessible through a RESTful API that allows the inclusion of GeneTrail2 into existing scripts, pipelines, and workflow systems. Our web-service is based on state of the art Web 2.0 technologies enabling a rich user-experience with a focus on clear visualization and helpful error reporting. The Java based server is backed by a C++ library in order to ensure optimal performance. Extensions for data analysis as well as new or updated external databases can easily be added. This ensures that GeneTrail2 can provide up-to-date information and implementations of popular algorithms. Provided all above, GeneTrail2 distinguishes itself by a high usability offering users many possibilities to analyze their data. With the given range of algorithms, databases, organisms and identifiers, GeneTrail2 offers scientists a substantial statistical framework as foundation for their research.

**C21:** Jeanne Marie Onana Eloundou-Mbebi, Sabrina Kleesseen, Michaël Méret, Thomas Degenkolbe, Lothar Willmitzer and Zoran Nikoloski. Reconstruction of substrate complexes in biochemical networks from time-resolved relative compound levels

**Abstract:** Reconstruction of networks of biochemical reactions each involving multiple substrates and products together with their stoichiometries, given time-resolved profiles of the compounds, is first step in modeling and understanding the control of biochemical systems. The existing approaches largely focus on identification of (pairwise) relationships between compounds by using similarity-measures and regression-based methods, thus neglecting the stoichiometry and directed hypergraph structure of biochemical reactions. Here we present the first computational approach for identifying the reaction substrates and their stoichiometries (substrate complexes): The approach assumes mass action kinetics, largely applicable to non-enzymatic (bio)chemical systems, and employs replicated time-resolved relative levels for the considered compounds. It combines techniques from constrained regularized regression and mathematical programming to arrive at robustly identified substrate complexes. The proposed approach is validated on synthetic data from a paradigmatic network. In addition, by using metabolomics profiles obtained from a glycine hydrothermal reaction, we show that the predicted substrate complexes are in line with chemical principles, thus, shedding light on prebiotic chemistry. The approach provides the basis for incorporation of elementary biochemical principles for accurate reconstruction of large-scale time-resolved biochemical networks.

**C22:** Claire Rioualen, Quentin Da Costa, Guillaume Pinna, Annick Harel-Bellan, Emmanuelle Charafe-Jauffret, Christophe Ginestier and Ghislain Bidaut. Interactome-regulome integrative approach for genome-wide screening data analysis in a breast cancer stem cells study

**Abstract:** Background. Breast cancer (BC) is the deadliest cancer in women worldwide. It is believed that cancer stem cells (CSC) could explain its recurrence, due to their resistance to conventional treatments and their ability to generate metastases. Identifying their key signaling pathways could be a lead for the development of new efficient treatments against BC.

**Method.** A genome-wide screening of an siRNA library was performed to evaluate the effect of single-gene knockdowns on CSC population (increase or decrease). 18,500 genes were scored accordingly. In order to isolate CSC-regulating pathways, we applied a network-based integrative approach using a heavily modified version of the ITI algorithm (Interactome-Transcriptome Integration, Garcia et al 2012). ITI has been initially developed in order to perform gene-expression-based tumor classification, using cancer cells expression and a map of publicly available protein-protein interactions data.

In order to isolate pathways involved in CSC evolution, we decided to include two changes in ITI: first, we used scores from the screening instead of gene expression profiles. Second, we added a regulation map (regulome made of TF-target couples), using interactions from public databases.

In this new algorithm, the interactome and the regulome are separately investigated, using target genes as “seeds”. If the seed's score is above a given threshold it is put aside. Its neighboring nodes are aggregated recursively if they improve the subnetwork's score, calculated by averaging screening scores of genes it contains. It is then statistically validated by comparing its score to those of subnetworks generated with randomized data. Interactome and regulome subnetworks are then merged. Resulting “metasubnetworks” are believed to be regulation pathways involved in CSC survival or death.

**Results.** Our method showed promising results by confirming the implication of genes that scored high hits in the primary screening. It also revealed the potential implication of unsuspected genes or TF. RUNX1, which was already described as a master regulator of leukemic stem cells, seems to regulate breast CSC. RBM14 revealed low-hit interacting mates, which showed Gene Ontology term “intracellular estrogen receptor signaling pathway” enrichment. NCOR1 was associated with the death of CSC, and the interactome-regulome integration showed proximity with PRDX1.

**Conclusion.** This new integrative approach yielded very promising results by both confirming in vitro experiments, and unraveling new genes of interest. This method could help scientists save time and money by potentially avoiding an expensive secondary screening, and ultimately allow us to identify new targets in cancer treatment. At the era of personalized medicine, elaborating a new drug targeting CSCs could greatly improve BC clinical outcome.

**C23:** Yang Xiang, Florian Martin and Joe Whittaker. A Stochastic Penalty for Incorporating Prior Information to Improve Reconstruction of Biological Networks

**Abstract:** Identification of the interactions between molecular entities within cells is key to the understanding of the biological processes involved. Many reverse engineering methods such as ARACNE, which can infer biological network from molecular abundance data (e.g., gene expression data), have been developed. Recently, several authors show that including a prior information such as the interaction relationships already identified experimentally on the network being reverse engineered leads to a more closely representative regulatory network. However, the way by which the prior network information is utilized, is specifically designed for a given algorithm. Currently, no general way for incorporating prior network information in reverse engineering is available. Here we propose a method, SPenGM (Stochastic Penalty for Gaussian graph Model) that integrates prior network information for many current standard reverse engineering algorithms such as ARACNE and PC, in a general and consistent way. In SPenGM, an additional dataset is generated by making use of prior network information, which is then combined with the experimental dataset to form an augmented dataset. All the standard algorithms relying on the covariance matrix, without any modification, can be applied on the augmented dataset to infer a network. Several reverse engineering methods, including ARACNE, RN (relevance network), MRNET, CLR, C3, PC, GLASSO, and ZPC (Zhang), were evaluated in thousands of simulated datasets generated by

using Barabasi–Albert model and Erdos–Renyi model, with and without prior network knowledge. Our simulation results show that SPenGM significantly improves the performance (Matthews correlation coefficient) of the above algorithms. In a case study SPenGM was applied to a C57/Bl6 mouse protein expression (BALF) data to infer a network between 26 proteins. Some interesting biological hypotheses were generated.

**C24:** Noel Malod-Dognin and Natasa Przulj. L-GRAAL: Lagrangian Graphlet-based Network Aligner

**Abstract:** Understanding the patterns in molecular interactions is of foremost importance in systems biology, as it is instrumental to understanding the functioning of the cell [1]. In the case of protein-protein interaction (PPI) networks, where nodes represent proteins and edges connect proteins that interact, network alignment has been particularly successful. Given two networks, aligning them means finding a node-to-node mapping (also called an alignment) between the networks that both maximizes the number of mapped proteins (nodes) that are evolutionarily or functionally related, and maximizes the number of shared interactions (edges). Network alignment uncovers valuable information, such as evolutionarily conserved pathways and protein complexes [2], or functional orthologs [3]. The increasing number of known interactions, coupled with the computational hardness of the network alignment problem [4], call for the development of novel network alignment methods that can handle large data while producing meaningful alignments.

First, we propose a novel scoring function for guiding the network alignment process, where the homological relationships among proteins (i.e., nodes) are quantified by their sequence similarity, and the topological similarity between interactions (i.e., edges) is based on their graphlet properties (graphlets are small, connected, non-isomorphic, induced sub-graphs of a larger graph [5-6]). Then, we introduce a novel global network alignment tool that we call Lagrangian GRaphlet based ALigner (L-GRAAL), which combines our novel scoring function with a novel global network alignment solver based on Lagrangian relaxation. We compare L-GRAAL with the state-of-the-art network aligners on the largest available PPI networks from BioGRID, and observe that L-GRAAL is one of the few methods that can produce alignments for pairs of very large biological networks. Also, L-GRAAL uncovers the largest overlaps between the networks, as measured with edge-correctness and the largest connect component. These large overlaps are key for transferring annotations between networks. Furthermore, we introduce a novel method for measuring the number of functionally conserved interactions that are uncovered by network aligners and show that L-GRAAL's alignments are in better agreement with Gene Ontology [7] and KEGG [8] pathway annotations than any other network aligners. Finally, using the PPI networks of baker's yeast and human, we show that L-GRAAL's alignments can be used to predict new interactions.

[1] Ryan et al., Nature Reviews Genetics, 2013.

[2] Kelley et al., Nucleic Acids Research, 2003.

[3] Bandyopadhyay et al., Genome Research, 2006.

[4] Cook, Proceedings of the Third Annual ACM Symposium on Theory of Computing, 1971.

[5] Przulj et al., Bioinformatics, 2004.

[6] Przulj, Bioinformatics, 2007.

[7] Ashburner et al., Nature genetics, 2000.

[8] Kanehisa et al., Science & Technology Japan, 2014.

**C26:** Cedric Simillion, Heidi Lischer, Robin Liechti and Rémy Bruggmann. Avoiding the pitfalls of gene set enrichment analysis with SetRank

**Abstract:** The purpose of gene set enrichment analysis (GSEA) is to find general trends in the huge lists of genes or proteins generated by many functional genomics techniques. GSEA has become the typical first step in analysing these datasets.

We have developed SetRank, an advanced GSEA algorithm which – unlike other methods – is not dependent on arbitrary cutoff values of input gene sets, avoids several types of bias, and is able to eliminate many false positive hits. The key principle of the algorithm is that it discards gene sets that have initially been flagged as significant, if their significance is only due to the overlap with another gene set. It calculates the p-value of a gene set using the ranking of its genes in the ordered list of p-values of all genes in the input dataset.

The remaining gene sets are visualized in a gene set network which can be used to further prioritize the results.

We show how the algorithm works and demonstrate its use on a few sample datasets. SetRank and the accompanying visualization tools will be made available both as R/Bioconductor packages and through an online web interface.

**C27:** Melanie Boerries, Hauke Busch, Jie Bao, Juliana M. Nascimento, Margareta Müller, Sofia Depner and Dennis Dauscher. Global Mean First Passage Time Analysis of Secretome to Transcriptome Signaling Reveals Endothelial-Derived TNF $\alpha$  and CXCL12 as Enhancers of Lung-Tumor Cell Migration

**Abstract:** Motivation: Tumor-stroma interactions are known to enhance cancer growth and metastasis, complicating treatment and strongly impairing patient survival. However, the details by which tumor cells communicate with their surrounding tissue still remain poorly understood. To obtain a comprehensive insight into tumor-stroma interaction of lung cancer, we study migration of a non-small-cell lung tumor cell line (H838) under the influence of Human Pulmonary Artery Endothelial Cells (HPAEC) in a transfilter coculture system.

Results: Jointly analyzing the H838 transcriptome response and secretome dynamics we predicted in silico and confirmed experimentally that HPAEC-derived cytokines TNF $\alpha$  and CXCL12, also known as stromal derived factor 1, enhance migration in the tumor cell line. In detail, we retraced the cause for migration-specific gene expression by linking putative transcription factor regulation to the upstream cell surface receptors by a mean first passage time analysis of the signaling flow on the cellular protein interaction network comprising more than 130,000 interactions. The analysis predicted five receptors from the interferon, tumor necrosis factor and C-X-C motif chemokine receptor families to be causal for the enhanced tumor cell migration. Combinatorial treatment of H838 cells using predicted receptor ligands confirmed endothelial-derived TNF $\alpha$  and CXCL12 as most potent enhancers of tumor cell migration. In conclusion we demonstrated for the first time the possibility to deduce tumor-stroma communication by the combination of transcriptome with protein-protein interaction data. This corroborates the hypothesis that lung cancer metastasis are fostered by endothelial cells through inflammation-related cytokines and chemokine receptors.

**C28:** Ashutosh Malhotra and Martin Hofmann-Apitius. Exploring novelty in mechanistic models for Alzheimer's disease by assessing reliability of protein interactions.

**Abstract:** Protein function is often modulated by protein-protein interactions (PPIs), therefore, defining the partners of proteins may help to understand its activity. Particularly, for complex idiopathic diseases like Alzheimer's disease (AD), the limited efficacy of currently available treatments indicates a strong need for deciphering alternative emerging disease mechanisms offering potential for additional therapeutic target development. As a consequence, there is substantial interest in developing a rational for the determination of the

reliability of PPI networks. Motivated by the same, we developed a scoring function strategy that could help in estimating the reliability of an edge in an AD PPI network built by deriving evidences explicitly described in literature. The calculation of score accounting for literature confidence assigned to an edge in a network takes into account the following attributes.

- Number of independent evidences in literature backing up PPI
- Number of contradictory evidences
- Experiment type: In vivo (Human/Mouse/cell lines), In Vitro (Using Physical methods (Most reliable), Genetic methods or Library based methods (Least reliable).
- Factual or Hypothetical statement backing up a claim.
- Community trust of the text reporting PPI.

The core of the developed workflow encompasses empirically optimized guidelines based on multiple criteria decision making (MCDM). Using this technique we asked experts in the AD field (Clinicians, Molecular biologists, Proteomics experts) to rank multiple, usually conflicting, criteria's based on which they judge the reliability of a PPI. Proportionally, ranks supplied by the decision maker (experts) were converted into weights, which were finally used to score a particular interaction.

Assessing the confidence of AD PPIs provided incremental support for the existing disease targets or known valid targeted interactions. For instance, APP-BACE1 interaction was scored highest in the network. Conversely, this assessment also point out regions partnering high score interactions where there is a sudden decline of score. We define such suburbs in a PPI network as Knowledge Cliffs (regions where ratio of adjoining interaction scores is maximum). The major reason accountable for the low scoring interactions could be their recent discovery or an existing expert bias for established trends neglecting surging interactions. Targeting such heterogeneous Knowledge Cliff regions help us in finding new emerging candidates sticking to the "well-Known" which could be serving as upstream causal entity with a relevant downstream bio-clinical effect on the established target. We aimed to derive all such candidates as they are already embedding into a well-established functional context. Our work demonstrates, how mechanistic modelling and the identification of reliable and novel findings in a network context can enhance the chance to identify new targets causally involved in a disease.

**C29:** Darren Davis, Omer Nebil Yaveroglu, Noel Malod-Dognin, Aleksandar Stojmirovic and Natasa Przulj. Topology-Function Conservation in Protein-Protein Interaction Networks

**Abstract:** Proteins underlay the functioning of a cell, and the relationships between the functions of proteins and their wiring (i.e., topology) in protein-protein interaction network (PIN) have been extensively studied. It has been shown that proteins with similar functions tend to cluster together [1] and have similar wiring patterns in PINs [2]. Based on these observations, the functions of proteins are predicted from their topological characteristics in PINs [2, 3]. However, all of these studies assume that, for each biological function, the wiring patterns of the proteins are similar. This might be an incorrect assumption since the wiring patterns of some essential functions might be better preserved than some species-specific functions during evolution. Moreover, some functions might be carried out by proteins with different wiring patterns in the PIN, hence it may not be possible to link these functions with specific network topologies.

To improve our understanding of the topology-function (TF) relationships in PINs, we develop a statistical framework that: (1) characterises the statistically significant TF relationships in a given species, and (2) uncovers the TF relationships that are conserved between baker's yeast and human. Our framework captures the wiring patterns of proteins in a PIN by using their graphlet degrees: the number of graphlets (i.e., small, connected, non-isomorphic, induced subgraphs) that the node touches at a topologically distinct place (i.e.,



graphlet orbit) [2]. The functions of proteins are given by their Gene Ontology (GO) annotations. Using our framework, we characterise the significant TF relationships for yeast and human PINs separately. Among many species-specific GO terms that are identified to have statistically significant correlation with the proteins being wired via a particular graphlet orbit, we identify 7 biological process and 2 cellular component GO terms that have evolutionary conserved wiring. We validate the obtained TF relationships for three of these GO terms (i.e., transcription initiation, localisation and proteasome complex) by finding studies that link the mechanisms involved in these functions with the identified wiring patterns in PINs.

The species-consistent TF relationships correspond to essential functions that are carried out similarly in different species. On the other hand, species-specific TF relationships also raise interesting questions such as why they are not preserved and how they differ across species. Furthermore, by utilising our framework as a filter for topology-based function prediction algorithms, we can improve the accuracies of such algorithms. Finally, our framework is generic and can be applied to uncover consistent wiring patterns in different phenomena, including those of disease and KEGG pathway annotations of proteins.

[1] Chua et al., *Bioinformatics*, 22:13, 2006.

[2] Milenkovic et al., *Cancer Inform*, 6, 2008.

[3] Sharan et al., *Mol Syst Biol*, 3:1, 2007.

**C31:** Aristidis Vrahatis, Konstantina Dimitrakopoulou, Athanasios Tsakalidis and Anastasios Bezerianos. A time-varying method for subpathway enrichment analysis

**Abstract:** In the era of personalized prognosis and treatment, experimentalists are in need of more precise pathway-based tools that embrace the scale and complexity of genomic information in time series expression data. Current approaches continue to search for (sub)pathways containing differentially expressed genes, failing to include the network constraints that connect the genes. In addition the time-varying features of pathway interactions are underappreciated and there is still no method to relate the temporal expression changes to the subpathway interaction dynamics.

For this purpose, we developed a time-varying method for subpathway enrichment analysis, which combines Markov dynamics to analyze the dynamics of linear and nonlinear subpathways based on the expression changes of the involved genes, while it adopts a Bayesian approach to set the parameters governing the evolution of interactions in time. Our method overcomes the most inaccurate assumption of current (sub)pathway-based approaches, which model cascades of signals or biochemical processes next to one another, in time-agnostic diagrams despite the fact that in reality, these phenomena happen over time, and often at different time scales.

Using synthetic data, we show the higher performance of our time-varying method over existing state-of-the-art pathway and subpathway-based methods in capturing the evolving subpathway dynamics. Based on public real time-series data, we show how time-varying subpathway analysis offers more biological insights compared to other efficient relevant tools. Such efforts are a push forward step for unraveling more accurate disease-relevant and drug-response subpathways.

**C32:** Léonard Jaillet and Stephane Redon. Characterizing reaction pathways with an energy-driven motion planning method.

**Abstract:** Nowadays, it remains challenging to reproduce accurately with existing simulation tools the phenomena taking place at the atomic-scale in chemical reactions. One approach to characterize such reactions is to provide the minimum-energy path (MEP) [1] associated to

the reaction i.e. a description of the rearrangements and relative positions of the atoms involved from their initial position in the reactant to their final position in the product. Here, we propose a new approach based on sampling-based motion planning techniques [2] to search the MEP associated to a given chemical reaction. It extends the Transition-based RRT method [3] to the problem of chemical reactions thanks to the Brenner potential [4] that allows a consistent representation of the energies and the forces taking place along the reaction.

A bidirectional search scheme expands two trees rooted at both the initial and final conformations to explore the space of Cartesian Coordinates. An implicit Voronoi bias drives the exploration towards yet unexplored regions of the space, while a Monte Carlo-like transition test limits the search to energetically favorable regions. The balance between these two strategies is automatically achieved thanks to a self-tuning mechanism. Finally, the nudged elastic band method [5] is used to refine and locally optimize the solutions found at the first stage. Our approach is integrated within SAMSON, a software platform for adaptive modeling and simulation of nanosystems.

[1] D. Sheppard et al, Journal of Chemical Physics, 128:13, 2008.

[2] S.M. LaValle, Planning Algorithms, Part II: Motion Planning, Cambridge University Press, 2006.

[3] L. Jaillet et al., Journal of Computational Chemistry, 32:16, 2011.

[4] D. W. Brenner, Phys. Rev. B, 42:15, 1990.

[5] G. Henkelman et al, Journal of Chemical Physics, 113:22, 2000.

**C33:** David Henriques, Miguel Rocha, Julio Saez-Rodriguez and Julio Banga. Reverse engineering of logic-based models using a mixed-integer dynamic optimization approach

**Abstract:** Motivation: Systems biology models can be used to test new hypotheses formulated on the basis of previous knowledge or new experimental data, contradictory with a previously existing model. New hypotheses often come in the shape of a set of possible regulatory mechanisms. This search is usually not limited to finding a single regulation link, but rather a combination of links subject to great uncertainty or no information about the kinetic parameters.

Results: In this work, we use the CellNOpt software to generate a logic-based description of all the possible regulatory structures for a given dynamic model of a pathway, allied with mixed-integer dynamic optimization(MIDO).

CellNOpt is an open source software that provides various formalisms to build predictive logic models of signaling networks by training networks (to signaling data (typically phosphoproteomic) using different mathematical formalisms derived from a representation as a logic circuit (in this work we use differential equations). Models built with CellNOpt are useful tools to understand how signals are processed by cells and how this is altered in disease, and to predict the effect of perturbations. CellNOpt is available as a set of R packages in BioConductor, and can be used directly from Cytoscape using the CytoCopteR App. CellNOpt can import and export models and pathways using the SBML-qual format.

Our MIDO framework aims to simultaneously identify the regulatory structure (represented by binary parameters) and the real-valued parameters that are consistent with the available experimental data. The alternative to this mixed-integer approach would be to perform real-valued parameter estimation for each possible model structure, which is not tractable for models of the size presented in this work. The performance of the method presented here is illustrated with several case studies: a synthetic pathway problem of signaling regulation, a two component signal transduction pathway in bacterial homeostasis, and a signaling pathway in liver cancer cells.

Although the metaheuristic approach we present does not provide guarantees about the global

optimality of the solutions, we show, by solving synthetic problems (case studies 1 and 2), that problems of realistic size can be successfully solved with a reasonable effort. Finally in the third case study, we apply the methods to a large signaling network given real experimental data. Due to its size (109 binary variables and 135 continuous parameters) this is, from the optimization point of view, an extremely challenging problem.

**C36:** Juris Viksna, Alvis Brazma, Karlis Cerans, Dace Ruklisa and Thomas Schlitt. Hybrid systems for modeling and analysis of qualitative behaviour of gene regulatory networks

**Abstract:** We present a hybrid system based framework for modelling of gene regulation and other biomolecular networks and a method for analysis of dynamic behaviour of such models. The particular feature of the proposed framework is focusing on prediction of qualitative experimentally testable properties of the system instead of quantitative features that in practice are often hard to measure. To achieve this we introduce the notion of the hybrid system's frame, which largely encompass the qualitative aspects underlying the biomolecular network described by the hybrid system more fully. This leads to a discrete state space of the network. We propose two different methods for the analysis of this state space. The result of the analysis is set of attractors (generalizations of attractors used in Boolean models) that characterise the underlying biological system.

Whilst in general case the problem of finding 'stable behaviour' regions in state space is algorithmically undecidable, we demonstrate that our methods work for comparatively complicated gene regulatory network model of lambda-phage. For this model we are able to identify two known biological behaviours of lambda-phage: lysis and lysogeny, and also to show that there are no other stable behaviour regions in state space of lambda-phage model.

**C37:** Ricardo de Matos Simoes, Kate E Williamson and Frank Emmert-Streib. Comparative analysis of gene regulatory networks inferred from large-scale urothelial cancer RNAseq, Bead and Oligo gene expression data sets

**Abstract:** Bladder cancer is a highly heterogeneous complex disease that is difficult to study on the basis of single genes. The underlying network structure and regulatory programs on the molecular level that are driving urothelial pathogenesis are largely unknown. In this study, we compare and infer three gene regulatory networks by the application of the BC3net inference algorithm to large-scale transitional cell carcinoma gene expression data sets from Illumina RNAseq (179 samples), Illumina bead arrays (165 samples) and Affymetrix microarrays (188 samples). We provide a detailed structural and functional analysis of the networks, identify highly co-regulated genomic regions, gene families, hub genes and study the network properties of known bladder cancer specific genes and biomarkers. The bladder cancer gene regulatory networks are highly enriched by interactions that are associated with known cancer hallmarks including cell cycle, immune response, signaling, differentiation and translation. The most prevalent associated co-regulated chromosomal regions of the networks are 5q31.3, 8q24.3 and 1q23.3 which represent genomic regions known to be frequently deregulated in bladder cancer and other cancer types. Hub genes of the bladder cancer gene regulatory networks are enriched by transmembrane proteins and represent target mediators of cellular activities and signalling processes. Our results shed new light on the analysis and integration of large-scale data sets to aid in the process for the identification and development of novel diagnostical targets in bladder cancer. This is the first study to our knowledge investigating genome-scale bladder cancer networks.