Poster abstracts

(in alphabetical order, according to presenting author)

(P01) Dynamic optimization identifies optimal programmes for pathway regulation in prokaryotes

Martin Bartl[1], *Martin Kötzing*[2], *Stefan Schuster*[2], *Pu Li*[1], *Christoph Kaleta*[2]
1 Ilmenau University of Technology, Institute for Automation and Systems Engineering, Ilmenau
2 Department of Bioinformatics and Research Group Theoretical Systems Biology, University Jena, Jena

Microorganisms have to be able to quickly react to environmental challenges to survive in fluctuating environmental conditions. For example, they can quickly regulate the expression of genes encoding metabolic pathways. In this work, we study optimal programs for pathway activation that take into account limited resources for protein production. Using dynamic optimization, we show in [BAR13] that protein abundance and protein synthesis capacity are key factors that determine the optimal strategy for the activation of a metabolic pathway. These strategies reach from the simultaneous activation of all enzymes over a sequential activation of groups of enzymes to a sequential activation of individual enzymes along the pathway. In the case of pathways with large differences in protein abundance, even more complex pathway activation strategies with a delayed activation of low abundance enzymes and an accelerated activation of high abundance enzymes are optimal. We confirm the existence of these pathway activation strategies as well as their dependence on our proposed constraints for a large number of metabolic pathways in several hundred prokaryotes. In summary, our work provides a confirmed hypothesis that explains the optimality of diverse sets of different pathway activation strategies and precisely defines the physiological constraints leading to their optimality. Depending on environmental conditions, there can also be a shift in the optimal program to activate a pathway. Moreover, as the optimal abundance of proteins as well as the protein synthesis capacity of an organism change in the course of its evolutionary history, the optimal operonic organization of metabolic pathways constantly changes. Thus, protein abundance as well as protein synthesis capacity are important contributors to the often observed high evolutionary plasticity of operons.

[BAR13] Bartl, Kötzing, Schuster, Li, Kaleta; Nat. Commun. DOI: 10.1038/ncomms3243, 2013

(P02) Towards an optimal transcriptome assembly of the Naked Mole Rat

Martin Bens, Karol Szafranski, Matthias Platzer
FLI-Leibniz Institute for Age Research, Jena

Naked mole rats (NMRs) (Heterocephalus glaber) are mouse-sized subterranean rodents with an exceptional long lifespan of >30 years in captivity. During their lifespan they show no age-related decline in fertility and not the typical gradual increase in mortality. Additionally, cancer has never been observed in this species. On the basis of these attributes, investigation of NMRs offers the possibility to discover molecular mechanisms, which lead to a long and healthy lifespan[Aus09, Buf08]. Advances in RNA-seq technology and transcriptome assembly provide a cost-effective way to analyse transcript expression profiles, despite the absence of a fully sequenced genome. We present a species independent transcriptome assembly pipeline, which utilizes different tools in order to pre-process mRNA-seq reads (sickle, cutadapt[Mar11]), assemble the transcriptome (Trinity[G+11]), screen for contamination (own scripts), reduce redundancy by sequence clustering (CD-HIT[LG06],TGICL[P+03]), identify chimeric transcripts (own scripts) and annotate the assembled transcript contigs (Blast[A+90]). Additionally, scaffolding of transcript contigs based on transcript data of a closely related species is performed and descriptive assembly statistics are reported. We obtained transcriptome data of 8 different tissues from NMRs, resulting in 360 million reads in total. With the aid of our pipeline we find NMR counterparts to 80% of human genes (NCBI Homo Sapiens Annotation Release 104). Scaffolding improved the length of 12% of all annotated genes. NMR counterparts cover 97% of all Gene Ontology terms, 96% of gene families defined by HUGO Gene Nomenclature Committee and 95% of genes

annotated in GenAge[dM+09]. A comparison with the current transcript annotation of the NMR genome [K+11] will be presented.

References

[A+90] SF Altschul et al. Basic local alignment search tool. Journal of molecular biology, 215(3):403–410, 10 1990.
[Aus09] S N Austad. Comparative biology of aging. J. Gerontol. A Biol. Sci. Med. Sci., 64(2):199–201, February 2009.
[Buf08] Rochelle Buffenstein. Negligible senescence in the longest living rodent, the naked mole-rat: insights from a successfully aging species. J Comp Physiol B, 178(4):439–445, May 2008.
[dM+09] J. P. de Magalhaes et al. The Human Ageing Genomic Resources: online databases and tools for biogerontologists. Aging Cell, 8(1):65–72, Feb 2009.
[G+11] Manfred G Grabherr et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol, 29(7):644–652, July 2011.
[K+11] Eun Bae Kim et al. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature, 479(7372):223–227, November 2011.
[LG06] Weizhong Li and Adam Godzik. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics, 22(13):1658–9, Jul 2006.
[Mar11] Marcel Martin. Cutadapt removes adapter sequences from highthroughput sequencing reads. EMBnet.journal, 17(1):10–12, 2011.
[P+03] Geo Pertea et al. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics, 19(5):651–652, 2003.

(P03) The overall structure of the multi-domain amyloid precursor protein

Ina Coburger, *Sven O. Dahms*, *Manuel E. Than*
FLI-Leibniz Institute for Age Research, Jena

No abstract available

(P04) Identification of unknown metabolites using tandem MS: Improving the quality of fragmentation trees

Kai Dührkop, *Sebastian Böcker*
Chair of Bioinformatics, University Jena, Jena

No abstract available

(P05) FlipCut Supertrees reloaded: Beating Matrix Representation with Parsimony

Markus Fleischauer, *Sebastian Böcker*
Chair of Bioinformatics, University Jena, Jena

No abstract available

(P06) Quantification of liver tissue by image processing

Adrian Friebel
IZBI, University Leipzig, Leipzig

During the last years imaging technology advanced significantly, especially with regard to the amount of data which can be obtained. Different microscopy techniques, such as brightfield, fluorescent confocal or two-photon microscopy and a variety of imaging methods, such as two-dimensional whole slide scans, three-dimensional image stacks or time series made it possible to quickly acquire large amounts of image data in the order of several hundreds of gigabytes. To be able to cope with this data, and to allow for a detailed quantification of the imaged structures, it is indispensable to have efficient, flexible and easy to use image analysis tools.
We present an image processing and analysis tool that allows the quantification of structures within liver lobules, such as hepatic nuclei, the sinusoidal and bile canaliculi networks and larger hepatic vessels from three-dimensional fluorescent confocal microscopy image stacks of different magnifications. It can be used to predict the borders of liver lobules and the shape of hepatocytes. We use statistical data obtained by this tool to calibrate a three-dimensional spatio-temporal model of liver lobules, which enables us to study the mechanisms that control e.g. the regeneration processes after liver impairment.
In our contribution to the workshop we will introduce our image processing tool with an emphasis on the methods of network reconstruction and quantification. We will show how these methods can be used for the analysis of two typical network-like structures within liver lobules: the sinusoidal and the bile canaliculi network.

(P07) Structure-Function-Relationship of the APP-E1-domain

Sandra Hoefgen, Sven O. Dahms, Dirk Roeser, Manuel E. Than
FLI-Leibniz Institute for Age Research, Jena

No abstract available

(P08) AgeFactDB - The JenAge Ageing Factor Database - Towards data integration in ageing research

Rolf Hühne, Torsten Thalheim and Jürgen Sühnel
FLI-Leipniz Institute for Age Research, Jena

No abstract available

(P09) Root-games between plants: predicting tendency for cooperation along environmental gradients

Sebastian Germerodt[1], Jana Schleicher[1], Katrin Meyer[2], David Ward[3], Stefan Schuster[1], Kerstin Wiegand[2]
1Department of Bioinformatics, University Jena, Jena
2Department of Ecosystem Modelling, University Göttingen, Göttingen
3School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

The degree of root overlap between neighboring plant individuals is determined by the net-outcome of plant interactions. We developed a model based on Evolutionary Game Theory and an individual-based model (IBM) to demonstrate how a positive (facilitative) process (i.e. hydraulic lift) may change the root proliferation strategy between plants competing for belowground resources and how this influences the degree of root overlap. In our approaches we interpreted facilitative interaction as a 'cooperation' strategy. In the game theory approach, the simplified situation of two plants competing for the same region of soil is represented. The calculation of the net

payoffs includes a surplus (due to facilitative processes) if plant roots overlap, costs of root development, and competition for nutrients. We show that the payoff for cooperation is not constant and the evolutionarily stable strategies change with availability of water and nutrients. This implies that the tendency for root overlap may also change along environmental gradients. In the IBM, we evaluated how the tendency to cooperate, and thus the degree of root overlap, depends on local resource availability, the strength of the facilitative process, and the costs for root development. Under low nutrient conditions and an average to high water availability aggregations of plants with a high tendency for defection emerged in the model. At low water availability, plants with a high tendency for cooperation accumulated. In areas with high availability of resources, a mixed pattern of the two strategies emerged. In sum, environmental gradients can change the net outcome of interactions, such as cooperation and competition, and can establish local patterns of interaction modes and varying degrees of root system overlap.

(P10) Analysis of RNA-seq data for identifying flowering time regulators in vernalized and non-vernalized rapeseed

Weinholdt C.[1], Emrani N.[2], Lemnian I.[1], Jedrusik N.[2], Molina C.[2], Jung C.[2], Grosse I.[1,3]
(1) Institute of Computer Science , Martin Luther University of Halle-Wittenberg, Halle
(2) Plant Breeding Institute, Christian Albrechts University of Kiel, Kiel
(3) German Center of Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig

No abstract available

(P11) De novo genome assembly and genome-wide variation analyses of *Nothobranchius furzeri*

Philipp Koch, Bryan R. Downie, Andreas Petzold, Marco Groth, Stefan Taudien, Kathrin Reichwald, Matthias Platzer
FLI-Leipniz Institute for Age Research, Jena

No abstract available

(P12) Biodiversity Exploratory Information System BExIS

Andreas Ostrowski, Eleonora Petzold, Michael Owonibi, Birgitta König-Ries
Heinz-Nixdorf Endowed Chair for Distributed Information Systems, University Jena, Jena

No abstract available

(P13) Definitions and nomenclatures for alternative splicing events

Martin Pohl
Department of Bioinformatics, University Jena, Jena

Alternative splicing of pre-mRNAs in higher eukaryotes and several viruses is one major source of protein diversity. A widely used graphical representation of alternative splicing events shows the alignments of transcripts as boxes representing exons connected by individual links for each isoform. Early in the analysis of alternative splicing, it became clear that standardization of the nomenclature for such alternative splicing graphs is important (Zavolan and van Nimwegen, 2006). Since then, some attempts have been made without leading to a broadly used and accepted nomenclature. Here we give an overview of such notations. We revisit their suitability in terms of limitations to applicability, especially with respect to regulatory coupling of alternative splicing events. For example, there is no

reason to exclude non-adjacent mutually exclusive exons. Hence, only approaches and nomenclatures considering mutual (perhaps long-ranging) dependencies within complete genes will have a chance of success in deciphering the full splicing picture. We propose a general description to overcome identified limitations. It utilizes Boolean algebra reducing splicing data to basic information while still incorporating the full complexity of alternative splicing (Pohl et al., 2013).
Pohl, M., Bortfeldt, RH., Grützmann, K., Schuster, S., 2013. Alternative splicing of mutually exclusive exons--A review, Biosystems 114, 31-38.
Zavolan, M., van Nimwegen, E., 2006. The types and prevalence of alternative splice forms. Curr. Opin. Struct. Biol. 16, 362-367.

(P14) Metabolic Costs of Amino Acid and Protein Production in *Escherichia Coli*

C. Kaleta, S. Schäuble, U. Rinas, and S. Schuster
Research Group Theoretical Systems Biology, University Jena, Jena

No abstract available

(P15) Theoretical study of lipid accumulation in the liver – Implications for nonalcoholic fatty liver disease

J. Schleicher[1], R. Guthke[2], U. Dahmen[3], O. Dirsch[4], H.G. Holzhütter[5], S. Schuster[1]
1 Department of Bioinformatics, University Jena, Jena
2 Leibniz Institute for Natural Product Research and Infection Biology – Hans-Knöll-Institute (HKI), Jena
3 Experimental Transplantation Surgery, University Hospital Jena, Jena
4 Institute for Pathology, University Hospital Jena, Jena
5 Medical Department (Charité), University Berlin, Berlin

A hallmark of the nonalcoholic fatty liver disease (NAFLD) is the accumulation of lipids in the liver (steatosis). We developed a simple mathematical model of the lipid dynamics in hepatocytes. Our spatially homogeneous model involves fatty acid intake, lipid degradation (oxidation) and triglyceride export. It takes into account that storage of triacylglycerol within hepatocytes leads to an enlargement of cell size. Then, the swelling of hepatocytes reduces the cross-section of sinusoids and impairs hepatic microcirculation. Thus the supply with oxygen is reduced, which, in turn, impairs lipid degradation. The analysis of our model revealed a bistable behavior (two stable steady states) of the system. The first stable state is characterized by intact lipid degradation and only a low amount of stored lipids. This state may correspond to the healthy hepatocyte. The second stable state in our model is marked by a high amount of stored lipids and reduced lipid degradation caused by cell enlargement and impaired hepatic microcirculation. This state may correspond to the steatotic hepatocyte. An interesting outcome of our model is its reversibility. The system can switch from the steatotic state back to the healthy state by a reduction of fatty acid intake. This corresponds to the observation that changes of lifestyle of NAFLD patients, especially reduced caloric diet, can cure steatosis.

(P16) The JenAge Information Centre - an information hub for ageing research and systems biology

Torsten Thalheim, Rolf Hühne, Jana Schleicher, Kerstin Wagner and Jürgen Sühnel
FLI-Leipniz Institute for Age Research, Jena

No abstract available

(P17) Transcription Factor Networks and Links to Biological Processes in Liver Metabolic Diseases and NAFLD

Sebastian Vlaic[1], *Sebastian Zellmer*[2], *Rolf Gebhardt*[2], *Reinhard Guthke*[1]
1 Leibniz Institute for Natural Product Research and Infection Biology – Hans-Knöll-Institute (HKI), Jena
2 Institute of Biochemistry, Medical Faculty, University of Leipzig, Leipzig

Question
The liver is the main metabolic organ and plays an important role in the regulation of diverse biological processes. Normal liver function is impaired when its metabolic capacity is persistently overloaded by inappropriate diets. This can cause liver diseases such as non-alcoholic fatty liver disease (NAFLD) which, in the long term, advances to steatosis and steatohepatitis. Investigation of the underlying signal transduction network might reveal key aspects in the development of such diseases. We therefore inferred a dynamic transcription factor network (TFN) using time resolved data of murine primary hepatocytes that were exposed to fresh medium after an initial period of serum deprivation [Hepatology 2010, 52(6):2127– 2136].
Methods
The dataset consisting of 5 time points (0, 3, 6, 12, 24 hours after stimulation) with three replicates each was used to determine differentially expressed genes (DEG). 22 DEGs for which the protein exhibits DNA binding transcription factor activity (DETFs) were selected for modeling. Additionally, the DEGs were clustered. Over-represented functions were assigned to the 6 clusters representing functional modules, and their mean expression profiles were included into the modeling process.
Results
To infer the TFN, we developed Extended TILAR (ExTILAR), an algorithm that expands the previously published transcription factor binding site integrating LARS (TILAR) [BMC Bioinformatics. 2009 Aug 24;10:262] method and adapts its modeling concept to work with time resolved data. Detailed analysis of the inferred TFN revealed two switch-like inhibitory loops, one between Foxa1 and Nr1h4, and one between Srebf1 and Nr1h4. All three of these TFs are known to play crucial roles in the cell such as regulation of metabolic processes or proliferation. We further investigated Tgif1, a TF whose role has not been intensively studied in the context of metabolic regulation in murine primary hepatocytes. Making use of an additional Tgif1 knock-down microarray experiment, we were able to confirm an inferred negative relation between Tgif1 and Atf3.
Conclusion
We developed ExTILAR, an algorithm that infers dynamic networks from time resolved data. Due to the concept of modeling, TILAR as well as ExTILAR are able to integrate diverse prior-knowledge sources such as TFBSs and gene to gene relations obtained from databases like TRANSFAC [Nucleic Acids Res. 2006 Jan 1;34:D108-10] or literature mining programs such as PathwayStudio [Ariadne Genomics, Rockville, USA]. Its application to biological data highlights the algorithms capability to infer biologically meaningful networks and generate testable hypotheses. Therefore, this study provides both, a valuable inference tool as well as important new insights into the network of TFs affected by the exchange of culture medium, i.e. changes in nutrient supply.

(P18) Different Stimuli for Inference of Gene Regulatory Network in Rheumatoid Arthritis

Peter Kupfer[1], *Sebastian Vlaic*[1], *René Huber*[2], *Raimund W. Kinne*[3], *Reinhard Guthke*[1]
1 Research Group Systems Biology / Bioinformatics, Leibnitz Institute for Natural Product Research and Infection Biology – Hans Knoell Institute (HKI), Jena
2 Institute of Clinical Chemistry, Hannover Medical School, Hannover
3 Experimental Rheumatology Unit, Department of Orthopedics, University Hospital Jena, Waldkrankenhaus 'Rudolf Elle', Eisenberg

Since genetic and epigenetic factors are known to be involved in the pathogenesis of rheumatoid arthritis the search for key players in this disease is one of the most important challenges. For this purpose gene regulatory networks are one possibility to reveal underlying interactions for different stimuli. In this study we analyzed the cellular response of synovial fibroblasts to 4 different stimuli. We infered a gene regulatory network that is able to explain the observed data for stimulation by TNF-α, TGF-β1, IL-1 and PDGF-D simultaneously.

(P19) MDM – Secondary Metabolite Gene Cluster Prediction

Thomas Wolf, Vladimir Shelest, Ekaterina Shelest
Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute (HKI)

Co-localization and co-regulation of functionally related genes is a common genomic feature, not only in eukaryotes. Clusters of genes which belong to the same pathway are prevalent in protists, plants and fungi, but can also be found in other organisms. Especially genes involved in the biosynthesis of secondary metabolites (SMs) are often organized in clusters. Although SM gene clusters are relatively well investigated, the majority of regulatory and biochemical details remains still unknown. The analysis of SM gene clusters is an expensive and time-consuming task in the laboratory. Tools to predict gene clusters in silico already exist, however, todays predictions are rather imprecise and often based on prior knowledge from known clusters. To exactly determine the genes, which belong to a specific cluster, we are working on the so called "motif density method" (MDM). As a novelty, MDM is based on hypothesized cluster specific transcription factor binding sites (csTFBSs), which should occur frequently within the cluster, due to the former mentioned co-regulation, and less frequently outside. First, MDM extracts all promoter regions of a given genome. Secondly, it predicts putative csTFBS motifs within the promoter regions located in the proximity of cluster backbone genes, for example polyketide synthases or non-ribosomal peptide synthetases. These backbone genes form the basis of SM biosyntheses and are easy to detect in the genome. The most promising motifs are then searched in all promoter regions extracted before. The final step incorporates the calculation of a score, which is based on the variable density of motif occurrences. The set of promoters, respectively genes, with the highest score represents the SM cluster prediction provided by MDM.

(P20) A general approach for discriminative de-novo motif discovery from high-throughput data

Jan Grau[1], Stefan Posch [1], Ivo Grosse [1], and Jens Keilwagen [2]
1 Institute of Computer Science, Martin Luther University Halle-Wittenberg, Halle
2 Julius Kühn-Institut (JKI) - Federal Research Centre for Cultivated, Plants, Quedlinburg

Transcription factors are a main component of gene regulation as they bind to specific binding sites in promoters of genes and subsequently activate or repress gene expression. The de-novo discovery of transcription factor binding sites from data obtained by wet-lab experiments is still a challenging problem in bioinformatics, and has not been fully solved yet. Today, major sources of in-vivo and in-vitro data are chromatin immunoprecipitation combined with high-throughput sequencing (ChIP- seq/ChIP-exo) and protein binding microarrays (PBMs), respectively. We present Dimont, a de-novo motif discovery approach specially tailored to these high-throughput data. Dimont successfully discovers all motifs of the ChIP-seq data sets of Ma et al. [M+12]. On the data sets of Weirauch et al. [W+13], it predicts PBM intensities from probe sequence with higher accuracy than any of the approaches specifically designed for that purpose. Dimont also reports the expected motifs for several ChIP-exo data sets. Investigating differences between *in-vitro* and *in-vivo* binding, we find that for most transcription factors, the motifs discovered by Dimont are in good accordance between techniques, but we also find notable exceptions. We provide a Dimont web-server at http://galaxy.informatik.uni-halle.de and a command line application at http://www.jstacs.de/index.php/Dimont.
References
[M+12] Xiaotu Ma et al. A highly efficient and effective motif discovery method for ChIP-seq/ChIP-chip data using positional information. *Nucleic Acids Res*, 40(7):e50, 2012.

[W+13] Matthew T. Weirauch et al. Evaluation of methods for modeling transcription factor sequence specificity. Nat Biotech, 31:126-134, 2013.