SAINT-PETERSBURG STATE POLYTECHNICAL UNIVERSITY

IOFFE PHYSICAL-TECHNICAL INSTITUTE
OF THE RUSSIAN ACADEMY OF SCIENCES

SYSPATHO CONSORTIUM OF THE FP7 FRAMEWORK
PROGRAM

SYSPATHO WORKSHOP SYSTEMS BIOLOGY AND MEDICINE

Abstracts

SysPatho Workshop
St. Petersburg, Tsarskoe Selo, Russia
September 11-14, 2012

St. Petersburg, 2012

INTERNATIONAL PROGRAM COMMITTEE

Prof. Dr. Roland Eils Heidelberg University, Germany

Prof. Dr. Sergei Inge-Vechtomov St. Petersburg State University, Russia

Prof. Dr. Nikolay Kolchanov Institute of Cytology and Genetics, Russian Academy of Sciences, Russia

Prof. Dr. Lars Kaderali Dresden University of Technology, Germany
Prof. Dr. Maria Samsonova St. Petersburg State Polytechnical University, Russia

Prof. Dr. Alexander Samsonov The Ioffe Phycical Technical Institute, Russian Academy of Sciences, Russia

Prof. Dr. Inna Lavrik Otto-von-Guericke-University, Magdeburg, Germany

SPONSOR ORGANIZATIONS

SysPatho Project – New Algorithms for Host Pathogen Systems Biology

EU Seventh Framework Program

DEUTSCHES WISSENSCHAFTS- UND INNOVATIONSHAUS MOSKAU

HOST SUPPORTERS

St. Petersburg State Polytechnical University

Ioffe Physical-Technical Institute of the Russian Academy of Sciences

CONTENTS

SESSION II

SESSION III

SESSION IV

Modeling Animal Transcription Networks as Highly Connected Quantitative Continua

Mark Biggin
Lawrence *Berkeley* National Laboratory, CA, USA

In vivo, Animal transcription factors each show a quantitative continuum of DNA binding to highly overlapping sets of genomic regions that are located close to most genes. These continua span functional, quasi-functional, and non-functional DNA binding events, with factor regulatory specificities being distinguished by quantitative differences in DNA occupancy patterns. Using the Drosophila blastoderm embryo as a model, we are developing computational models that describe the biochemical mechanisms that produce these patterns of DNA binding and how combinations of transcription factors cooperate to generate spatial and temporal gene expression.

e-mail: mdbiggin@lbl.gov

Systems Biology of Cellular Rhythms

Albert Goldbeter
Unit of Theoretical Chronobiology, Faculty of Sciences
Université Libre de Bruxelles (ULB), B-1050 Brussels, Belgium

Rhythmic phenomena occur at all levels of biological organization, with periods ranging from a fraction of a second to years. Many rhythms originate from the regulatory feedback loops that control the dynamics of cellular processes. Cellular rhythms provide a prototype for the field of Systems Biology as they illustrate how an emergent property, autonomous oscillatory behavior, arises from molecular interactions in regulatory networks. Moreover, in line with the close link between Systems Biology and Computational Biology, oscillations can best be addressed by combining an experimental with a modeling approach. After providing an overview of biological rhythms, I will focus on two major examples of rhythmic behavior at the cellular level: circadian clocks and the cell cycle. First, computational models will be used to address the molecular mechanism of circadian rhythms, as well as the dynamical bases of circadian clock-related physiological disorders. Second, using a model for the network of cyclin-dependent kinases (Cdks) that drives the mammalian cell cycle I will show that the regulatory structure of the Cdk network results in its temporal self-organization in the form of sustained oscillations, leading to the sequential activation of the Cdks that brings about the orderly progression along cell cycle phases. The coupling of the cell cycle to the circadian clock results in the synchronization of these two major cellular rhythms.

REFERENCES:

1. Leloup J-C & Goldbeter A (2003) Toward a detailed computational model for the mammalian circadian clock. *Proc Natl Acad Sci USA* 100, 7051-7056.

2. Leloup J-C & Goldbeter A (2008) Modeling the circadian clock: From molecular mechanism to physiological disorders. *BioEssays* 30, 590-600.

3. Gérard C & Goldbeter A (2009) Temporal self-organization of the cyclin/Cdk network driving the mammalian cell cycle. *Proc Natl Acad Sci USA* 106, 21643-21648.

4. Gérard C & Goldbeter A (2011) A skeleton model for the network of cyclin-dependent kinases driving the mammalian cell cycle. *R. Soc. J. Interface Focus* 1, 24-35

e-mail: agoldbet@ulb.ac.be

Human brain origin and the evolution of regulatory and coding regions of protein-coding genes expressed in brain

K.V. Gunbin[1], M.P. Ponomarenko[1], D.A. Afonnikov[1,2], N.A. Kolchanov[1,2,*]
[1]Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Russia, 630090, Novosibirsk, Prospekt Lavrentyeva 10
[2]Novosibirsk State University, Russia, 630090, Novosibirsk, Pirogova Str. 2

The great enigma of contemporary evolutionary biology is how tiny changes of hominid genes could make the unique human brain? In order to answer this question we made complex evolutionary analysis of hominid (*H. sapiens*, *P. troglodytes, G. gorilla, P. pygmaeus*) genes whose human orthologs are expressed in brain. We extracted multiple alignments of protein coding genes from the EnsEmbl Rel. 65; we took human gene expression data from the Allen Brain Atlas (March 2012) and human transcription start sites (TSS) from RefSeq Rel. 52. We reconstructed ancestral sequences in each internal node of the Hominidae tree using both Bayesian (MrBayes 3.2.1) and Maximum likelihood (PAML 4.4) approaches. To predict TATA-box activity we implement our algorithm described in [1], which used an empirical equilibrium equation for three-step for TBP/TATA binding: TBP slides along DNA, TBP stops in the TATA box, the TBP/TATA-complex is stabilized due to deformations in DNA. The evolution modes of protein coding regions at each tree branch were inferred using 15 K_A/K_S calculation methods implemented in the KAKS_Calculator 2.0 program. Two global evolutionary trends for hominid genes whose human orthologs are expressed in brain were found: (1) the expression of positively selected genes and genes under weak negative selection occur mainly in the frontal and temporal cortex; (2) the relative conservation of the ancestral TATA-box organization (20-90 b.p. upstream TSS) in human genes compared with intensive TATA-box losses in primate orthologs (Fig. 1). It is of interest that the first trend elucidates the evolving of frontal cortex in hominids, while the second describes the potential mechanism of hominid brains specializations.

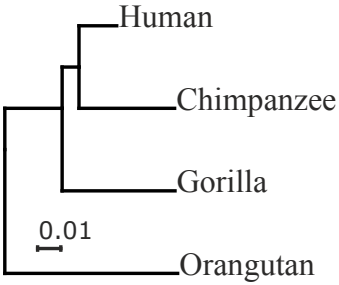

Fig. 1. Phylogram representing Hominidae evolution, branch lengths shows the excess (in %) of evolutionary cases with decrease in TBP affinity to TATA-box over cases with increase in activity.

The work supported by SB RAS (130, 39, 93), RAS (project 6.8 and program 28), and Scientific school 5278.2012.4.

REFERENCES

1. Ponomarenko P.M. et al. (2008) A step-by-step model of TBP/TATA box binding allows predicting human hereditary diseases by single nucleotide polymorphism, Dokl. Biochem. Biophys., 419: 88-92.

email: kol@bionet.nsc.ru
*Corresponding author

Evolution of regulatory interactions in bacteria

Mikhail Gelfand
A.A.Kharkevich Institute for Information Transmission Problems, RAS

Availability of numerous sequenced genomes at varying taxonomic distances, relatively large information content of transcription factor (TF)-binding motifs, and rarity of co-operative binding make bacteria an ideal model for studying evolution of regulatory networks. Such studies may be may be performed in two directions: analysis of regulation of a particular functional subsystem (e.g. a metabolic pathway) or analysis of co-evolution of TFs and their DNA motifs. Both approaches demonstrate high flexibility of bacterial regulation, involving rapid restructuring of regulatory cascades and correlated changes in TFs and their motifs. Comparison of closely related genomes demonstrates that about 40-50% of intergenic regions are subject to purifying selection, especially at TF-binding sites and promoters. Conservation of individual regulatory interactions may depend on their network context such as feed-forward loops. On the other hand, some regulatory systems are conserved at surprisingly large evolutionary distances.

e-mail: mikhail.gelfand@gmail.com>

A quantitative systems biology study on a model bacterium

Luis Serrano
Center for Genomic Regulation, Barcelona, Spain

The goal of Systems Biology is to provide a quantitative and predictive description of a living system to the extent that it can be fully simulated in a computer. We have undertaken such Endeavour using as a model the small bacterium, M. pneumoniae. use Mycoplasma pneumoniae, a human pathogenic bacterium causing atypical pneumonia as model system for our study. Containing a reduced genome with only 690 ORFs, this bacterium is an ideal organism for exhaustive quantitative and systems-wide studies, avoiding technical limitations due to exceeding sample complexity, constrained by limitations in dynamic range and resolution of current generation mass spectrometers. Available data on the transcriptome, on protein complexes, as well as on metabolic pathways facilitate the integration of the data generated for this study into an organism-wide context. Additionally, M. pneumoniae represents a relevant organism to study stochastic noise in living systems. The cells are significantly smaller than other bacteria, such as Eschericha coli (0.05 mm3 and 1 mm3, respectively) resulting in principle in an increased susceptibility to abundance fluctuations of cellular molecules. Our analysis shows that even apparently simpler organisms have a large hidden layer of complexity and that for every question we have answered we have got two new ones. We are still far away to be get a full understanding of a cell.

e-mail: luis.serrano@crg.eu

Model based analysis of interferon induced apoptosis

Irene Otero Muras* and Joerg Stelling
Department of Biosystems Science and Engineering, ETH Zürich, Universitätstrasse 6, 8092-Zürich, Switzerland

Type I interferons are used in the treatment of several types of cancer and other diseases [1]. One important aspect of the antitumor activity of type I interferons is related to their capacity to induce apoptosis. However, several of the molecular mechanisms driving interferon induced apoptosis are not yet resolved.

Extrinsic and intrinsic apoptosis pathways transduce different signals (caused by a death ligand in the case of the extrinsic pathway, and by radiation or cell damage in the case of the intrinsic pathway) leading to the activation of initiator caspases. The initiator caspases cleave in turn the effector caspases, responsible for the proteolysis of key cellular substrates. The activity of the effector caspase 3 is widely used as a readout of apoptosis.

Recent experimental work by the Schreiber group [2] identified a novel mechanism of interferon induced activation of the extrinsic apoptosis pathway that is highly dependent on the levels of initiator caspase 8 and of the inhibitor CFLAR, but independent of the death ligand. The associated mechanisms for assembly and activation of the receptor and adaptor proteins in the so called death inducing signaling complex (DISC) remain still unclear. In order to investigate the mechanisms responsible for this activation in detail, we have developed a dynamic (ODE based) mathematical model for the extrinsic apoptosis pathway based on the experimental data from [2]. Unlike previously published models where the death ligand triggers the signaling cascade [3], our model takes as inputs the measured levels of caspase 8 and other key proteins in the apoptosis pathway induced upon interferon treatment, which requires a physiologically realistic representation of the steady-state without death ligand. The model reproduces several aspects of the dynamic behavior for the activation of the caspases

and protein-protein interactions observed experimentally [2]. The model based approach is currently being used, together with the experimental analysis by the Schreiber group, to investigate hypotheses on additional mechanisms which can be critical for interferon induced cell death.

REFERENCES

1 M. Chawla-Sarkar, D. J. Lindner, Y.-F. Liu, B. R. Williams, G. C. Sen, R. H. Silverman and E. C. Borden. (2003). Apoptosis and interferons: Role of interferon-stimulated genes as mediators of apoptosis, Apoptosis, 8:237-249.

A. Apelbaum, G. Yarden, S. Warszawski, D. Harari and G. Schreiber. (2012). Type I interferon induces apoptosis by changing the delicate balance between cFLIP and caspase-8 in a death-ligand independent manner, submitted for publication.

S. Spencer and P. Sorger (2011). Measuring and modeling apoptosis in single cells, Cell, 144(6):926-939.

e-mail: joerg.stelling@bsse.ethz.ch
*Corresponding author

Reverse engineering infection diseases: towards a virtual patient

Vitor A.P. Martins dos Santos
Chair for Systems and Synthetic Biology, Wageningen University,
Dreijenplein 10, Building number 316
6703 HB Wageningen
The Netherlands

When a pathogen infects a host, essential elements interact in dozens, if not hundreds, of ways. Host-pathogen relationships are characterized by a complex interplay between host defence mechanisms and attempts to circumvent these defences by microorganisms. This complexity imposes enormous challenges which are best tackled through computational analysis. Obtaining an integrated view of the immune system and of its interactions with pathogens entails the development of models that look at the immune system in various ways.

The use of "*in silico*" methods to analyse disease mechanisms enhances understandings that come from both *in vitro* and *in vivo* research as it allows us to consider host-pathogen relationships in the context of its relevant interactions. Such analyses permit simulation of events that occur over very long time scales (eg. in latent infections) and, very importantly, allow (*in silico*) experimentation when no proper animal models are available (eg. as in Streptococcal Toxic Shock Syndrome, STSS). In the long run, its is expected that *in silico* R&D would allow reasonable simulation experiments in a computer, rapidly testing what would likely take months or years in the laboratory or clinic and thereby greatly reducing product development costs and, possibly allowing targeted drug development.

Hence, aiming at providing quantitative understanding of key events that take place upon development and progression of the host immune response to a microbial infection, a reverse-engineering research programme integrating essential features of the host-pathogen interactions is proposed. For the top of the "interaction scaffold", I will illustrate this concept with a simple mathematical model describing the relations between superantigen production, cytokine levels and clinical outcome by *Streptococcus pyogenes* infections. At the intermediate level, I will exemplify this with a systems biology analysis of the metabolic interactions among the members of a microbial community. At the lower level of the scaffold, I will shortly describe the use of genome-wide constraint–based models to explore the metabolic and transport capabilities of a potential pathogen. Experimental model validation and implications for the understand-

ing of the underlying mechanisms will be discussed for the various levels. I will present a further example on the systems biology of tuberculosis, by handling the interplay between top down and bottom up models dealing with different aspects of infection.
e-mail: vitor.martinsdossantos@wur.nl

Homeostasis and adaptation. Towards a trade-offs model

Alexei Romanyukha
Institute of Numerical mathematics RAS, Moscow, Russia

The aim of the work is the analysis of the principles of regulation and adaptation of systems maintaining the homeostasis of a multicellular organism.

Adaptation is defined as the change in the functional capacity of the physiological systems in response to changing external conditions. It is assumed that the purpose of adaptation, is the increase in the average net energy surplus, which can be used for reproduction.

The research is conducted with the help of the mathematical model describing the characteristics of the organization of the metabolic machinery cells and multicellular organism.

The model describes the adaptation of the organism to changes in the nutrition schedule and caloric intake and changes in their physical loading. The mechanism of metabolic syndrome are discussed.

e-mail: eburg@inm.ras.ru

Optimization of combined treatment (interferon, ribavirin, protease or polymerase inhibitors) for hepatitis C virus.

S. Tapilina**, *O.Demin Jr
Institute for System Biology SPb, Moscow, Russia

Nearly 3% of the world's population (130–170 million people) is chronically infected with hepatitis C virus (HCV). Currently there are several ways of treatment: interferon (IFN) therapy (including ribavirin - RBV) and new inhibitor of HCV polymerase or protease therapy. Unfortunately, some patients are non-response for IFN therapy and HCV could mutate and become resistant for inhibitors of nonstructural proteins; also virus has 6 different genotypes (and many subtypes) and efficacy of treatment is depended on virus genotype.

The main aim of our work is to optimize treatment with various combinations of IFN and RBV therapy with inhibitors of virus protease NS3 and polymerase NS5B.

Using existed mathematical models for IFN therapy and protease or polymerase inhibitor we develop mathematical model for combination therapy. The model was verified and validate against all available in vitro, in vivo and clinical data (including individual data).

We predict optimal combination of IFN, RBV and protease or polymerase inhibitors, the most effective order, doses and regimes of drugs administration for

responders or non-responders patients

patients with different virus genotype (1a, 1b, 2a etc)

patient with various rate of virus mutation

This model could be used for drug discovery and development for HCV and improvement and optimization of therapies and their combinations.

email: tapilina@insysbio.ru
*Corresponding author

Understanding molecular mechanisms of HCV patients susceptibility to IFN therapy via systems pharmacology modeling

***Sergey Smirnov*[1,*], *Neil Benson*[2], *Piet Van Der Graa*[2], *Victoria Flores*[2], *Ekaterina Goryacheva*[1], *Oleg Demin*[1]**
[1]Institute for Systems Biology SPb, Moscow, Russia
[2]Pfizer, Pharmacometrics, Global Clinical Pharmacology, Walton Oaks, KT20 7NS, UK

Approximately 2% of the world population is currently infected with hepatitis C virus (HCV). The current standard-of-care therapy is based on interferon-α (IFN-α) in combination with ribavirin for 48 weeks. This therapy is associated with significant adverse effects and successful in only ~50% of patients.

One of the key characteristics determining susceptibility of HCV patients to standard-of-care therapy is the EC50 of IFN. It has been demonstrated clinically that the EC50 of non-responder patients (0.45 μg/L) substantially exceeds that measured for responder patients (0.04 μg/L). The aim of this work was to explore using a systems pharmacology approach if this difference can be explained in terms of quantitative characteristics of signaling pathways mediating the hepatocyte response to INF.

There are two main experimentally observed phenomena contributing to EC50 variability of IFN-α treatment: desensitization and refractoriness. *Desensitization* is defined as IFN–α *independent* decrease in sensitivity of hepatocytes with respect to IFN–α which results from interaction of virus particles with signaling proteins/receptors. *Refractoriness* is defined as IFN–α *dependent* decrease in sensitivity of hepatocytes to IFN–α treatment which is based on negative feedback observed in IFN-α signaling pathway.

In the current study we have reconstructed key signaling pathways participating in IFN-based cell response and developed a mathematical model of HCV combining virus/cell dynamics with a quantitative description of IFN-dependent JAK/STAT mediated signaling pathways. Molecular mechanisms explaining both desensitization and refractoriness and, consequently, the status of HCV patient (responder *vs* non-responder) in terms of dynamic and regulatory properties of signaling and gene-regulatory pathways were identified. Our mathematical model was subsequently applied to find out possible biomarkers indicating patient status before initiating INF therapy. Taking into account the values of the biomarkers the model is also able to predict optimal dosage and administration regime strategy for each particular patient.

In conclusion, a systems pharmacology model has been developed which provides a quantitative framework for a more personalized treatment paradigm in HCV.

e-mail: smirnov@insysbio.ru
*Corresponding author

Type-I Interferon response is controlled by multi-layered stochasticity

Ulfert Rand[1], Melanie Rinas[2], Johannes Schwerk[1], Mario Köster[1], Thomas Höfer[2], and Hansjörg Hauser[1]*
[1]Department of Gene Regulation and Differentiation,
Helmholtz Centre for Infection Research, Braunschweig, Germany
[2]Division of Theoretical Systems Biology,
German Cancer Research Center (DKFZ) and BioQuant Center, Heidelberg, Germany

The interferon (IFN) system provides a powerful defense against viral infections. Upon viral infection pathogen recognition receptors activate an intracellular signaling cascade that leads to the activation of nuclear factor κB (NF-κB) and interferon regulatory factors (IRFs). Subsequent induction of the IFN-β gene and other type-I IFNs is a hallmark of the early response to infection. Upon secretion and binding of type-I IFNs to their specific receptor, the Jak/STAT signaling pathway is activated to reprogram gene expression. Here, we provide a comprehensive analysis at single-cell resolution of type-I IFN induction by virus and the cellular response to secreted IFN. Using fluorescent reporters based on bacterial artificial chromosomes (BACs) and chimeric transcription factors, we imaged the successive key steps of the IFN system leading to the establishment of an antiviral state. Our results show that cell-to-cell heterogeneity is a pervasive feature of the IFN system. This heterogeneity manifests itself not only in the virus-induced antiviral signaling and IFN gene induction but equally in the IFN-induced protective response. By comparing recently separated sister cells we show that the temporal unpredictability of IFN-β expression is largely due to cell-intrinsic noise generated both upstream and downstream of NF-κB and IRF transcription factor activation without heritable (genetic or epigenetic) variability. Furthermore, we show that the IFN-stimulated response within a (clonal) cell population exhibits bimodality at suboptimal concentrations, where the expressing subpopulation increased with extracellular IFN concentration. Consequently, we show that only the subpopulation of cells that is identified by ISG reporter gene expression coordinately expresses an antiviral gene program to protect cells against Newcastle Disease Virus (NDV) infection.

To link the stochastic single-cell dynamics to antiviral protection at the cell-population level, we developed a data-driven mathematical model and tested its predictions experimentally. The model allows to translate the all-or-nothing decisions in individual cells into predictable dynamics of IFN secreting and protected cell fractions at the population level. Our results show that a reliable antiviral response in the face of sin-

gle-cell stochasticity is achieved through the paracrine amplification of the response. This underscores the importance of understanding innate immune responses in terms of collective spatio-temporal dynamics of cellular decisions.

REFERENCES

Rand et al., Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response. Mol Systems Biol., in press

email: Hansjoerg.Hauser@helmholtz-hzi.de

SysPatho: Systems biology of Host-Pathogen Interactions

Lars Kaderali
The Institute for Medical Informatics and Biometry, Dresden, Germany

SysPatho is a research consortium funded in the 7th framework program of the European Union. It focuses on the development of novel and generally applicable mathematical methods and algorithms for systems biology. These methods and algorithms will be applied to study the complex interactions of hepatitis C virus (HCV), a human-pathogenic virus of high medical relevance, with its host at the systems level. Using a multidisciplinary, integrative approach, SysPatho aims to (a) develop methods to analyze and integrate a wide variety of data from wet lab experiments, databases and biological literature, (b) develop and apply machine learning tools to reconstruct and study intracellular interaction networks from experimental data, (c) develop new and improve existing algorithms and mathematical methods for bottom-up modelling, to fit models to data, and to analyze the dynamic behaviour of models (d) generate new experimental data to gain novel insights into hepatitis C virus host interactions, and (e) use the newly developed methods and data to model and analyze HCV-host interactions at the systems level. Guided by biological data, SysPatho focuses on the design of novel algorithms and mathematical methods for systems biology, with the aim to provide generally applicable tools to elucidate biological processes. Based on developed models and using systems analysis, SysPatho aims to elucidate virus host interactions of Hepatitis C virus at an unprecedented level. Ultimate aim is the identification of new candidate host cell target genes applicable for the design of novel antiviral drugs against hepatitis C. Targeting of host cell factors will reduce the likelihood for the development of therapy resistance and increase the chance for broad-spectrum antivirals.In the presentation, I will present the project and key results on HCV-Host Interactions obtained so far.

email: lars.kaderali@tu-dresden.de

Modeling and visualization of the spatio-temporal dynamics immune processes in lymphoid organs

Ilya Burenko* [1], **Gennady A. Bocharov**[2]
[1] Moscow State University
[2] Institute of Numerical Mathematics

We present an approach for visualization of spatio-temporal virus infection processes in the secondary lymphoid organs. Organs can change their volumes depending on some parameter, for example, viral load. In order to simulate and visualize spatio-temporal dynamics of the immune processes we developed methods that allows to control a set of parameters of mathematical model and visualize the virus and immune processes with three dimensional geometric models.

We the mathematical model of virus infection formulated with a system of delay differential equations, which describes the dynamics of the virus, interferon, infected and healthy plasmacytoid dendritic cells (pDC) and macrophages (M-) [1]. We implemented a graphical user interface using MATLAB facilities, which allows to user to specify a time interval of DDE solutions, the initial infection, as well as a set of parameters in the model equations. GUI allows to user plots graphics of the temporal dynamics for different equations. The user controls what graphics should display. We considered two approaches to implement three-dimensional geometric models: polygonal method (using Blender software) and the method utilizing the R-Functions [2]. Polygonal model is easier to implement, but the boundary of geometric region can not be parameterized. An approach based on the R-functions, allow us to define the boundaries the compartments considered in the model. Initially, we use the following sufficiently complete system of functions:

$$\begin{aligned} \vee_0 &= x_1 + x_2 + \sqrt{x_1^2 + x_2^2} \\ \wedge_0 &= x_1 + x_2 - \sqrt{x_1^2 + x_2^2} \\ \bar{x} &= -x \end{aligned}$$

Geometrics model were integrated with the solution of the mathematical model (using MATLAB). Depending on the virus concentration in the compartments, the color of the compartments boundaries changed. Overall, we developed GUI that allows to user to set the parameters of the mathematical model, observe temporal and spatio-temporal dynamics of immune processes.

Funding.

The authors acknowledge financial support from the Russian Foundation for Basic Research and the Programme of the Russian Academy of Science \Fundamental Research for Medicine».

REFERENCES

1. G. Bocharov, R. Zust and others A System Immunology Approach to PlasmacytoidDendritic Cell Function in Cytopatic Virus Infection, PloS Pathogens, July 2010

2. V. Rvachev Theory of R-functions and some of its applications. - Naukova Dumka,Kiev, 1982. e-mail: ivanov@bionet.nsc.ru

email: iburenko@gmail.com

*Corresponding author

Tracking and Registration for Live Cell Image Analysis

Karl Rohr
University of Heidelberg, BIOQUANT, IPMB, and German Cancer Research Center (DKFZ), Dept. Bioinformatics and Functional Genomics, Biomedical Computer Vision Group (BMCV), Im Neuenheimer Feld 267, 69120 Heidelberg, Germany

Tracking and registration are central tasks for analyzing live cell microscopy images. We have developed an approach for cell tracking, which in conjunction with segmentation, feature extraction, and classification, determines cell lineages as well as quantifies cell cycle durations. To correct classification errors we have introduced a state transition model which exploits the temporal context. Our approach can cope with normal as well as abnormal phenotypes and was applied to RNAi knockdown image data of HeLa cells, and more recently was extended to high-throughput screens of Neuroblastoma cells. We have also developed an image analysis approach for quantifying large scale RNAi high-throughput screens of virus infected cells (hepatitis C virus, dengue virus) to identify relevant factors for virus replication.

To analyze the spatial-temporal behavior of particles, we have developed a probabilistic tracking approach based on particle filters or Kalman filters. The approach was applied, for example, for tracking of HIV particles in multi-channel fluorescence microscopy images to study pathways of virus entry. We have also developed an image registration approach for spatial normalization of live cell microscopy images and accurate classification of particle motion. Our approach allows rigid and non-rigid registration and has been applied to 2D and 3D microscopy images to study the motion of nuclear particles and viral particles.

REFERENCES

1. Harder N, Mora-Bermúdez F, Godinez WJ, Wünsche A, Eils R, Ellenberg J, and Rohr K. (2009) Automatic Analysis of Dividing Cells in Live Cell Movies to Detect Mitotic Delays and Correlate Phenotypes in Time, *Genome Research* 19:11, 2113-2124

2. Godinez WJ, Lampe M, Wörz S, Müller B, Eils R, and Rohr K. (2009) Deterministic and Probabilistic Approaches for Tracking Virus Particles in Time-lapse Fluorescence Microscopy Image Sequences, *Medical Image Analysis* 13, 325-342

3. Chang, L., Godinez, W.J., Kim, I.H., Tektonidis, M., de Lanerolle, P., Eils, R., Rohr, K., Knipe, D.M. (2011) Herpesviral replication compartments move and coalesce at nuclear speckles to enhance export of viral late mRNA. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 108 (21), E136-E144

4. Kim I-H, Chen YC. Spector DL, Eils R, and Rohr, K. (2011) Non-Rigid Registration of 2D and 3D Dynamic Cell Nuclei Images for Improved Classification of Subcellular Particle Motion. *IEEE Transactions on Image Processing* 20:4, 1011-1022

5. Matula P., Kumar A., Wörz I., Harder N., Erfle H., Bartenschlager R., Eils R. and Rohr K. (2009) Single-cell-based image analysis of high-throughput cell array screens for quantification of viral infection. *Cytometry A*, 75A:309-318

6. Reiss S, Rebhan I, Backes P, Romero-Brey I, Erfle H, Matula P, Kaderali L, Poenisch M, Blankenburg M, Hiet MS, Longerich T, Diehl S, Ramirez F, Balla T, Rohr K, Kaul A, Bühler S, Pepperkok R, Lengauer T, Albrecht M, Eils R, Schirmacher P, Lohmann V, and Bartenschlager R. (2011) Recruitment and activation of a lipid kinase by NS5A of the hepatitis C virus is essential for the integrity of the membranous replication compartment, *Cell Host & Microbe* 9:1, 32-45

e-mail: k.rohr@dkfz.de

SESSION II
BIOINFORMATICS AND SYSTEMS BIOLOGY OF HIGH-THROUGHPUT DATA

Robust transcriptomic data analysis across individuals

Benno Schwikowski
Institut Pasteur

Transcriptomic measurements are an important tool in the elucidation of the genetic and environmental determinants of disease susceptibility, but also drug response, and altered cellular function in general across individuals. The ongoing decline in the cost of transcriptomics increasingly allows additionally the observation across multiple time points, exposing speed variability between individuals in certain biological processes. Computational analysis of this data has been difficult, because existing approaches either perform gene-by-gene analyses on a weak statistical basis, or ignore speed variability altogether.

In my talk, I will discuss a new data analysis approach designed to be robust to speed variability, and its evaluation it in a biomarker discovery scenario across four publicly available datasets. In our evaluation, existing methods perform surprisingly poorly; only the robust approach yields reproducible and biologically plausible results. Our results indicate the need to capture gene expression between potentially heterogeneous individuals at multiple time points, and highlight the importance of robust transcriptomic data analysis across individuals.

e-mail: benno@pasteur.fr

Computer analysis of 3D chromosome contacts mediated by RNA Pol II in human

Yuriy L. Orlov*, Guoliang Li[1], Kuljeet .S. Sandhu[1], Melissa J. Fullwood[1], Dmitriy A. Afonnikov, Chialin Wei[1], Oleg L. Serov, Nikolay A. Kolchanov, Yijun Ruan[1]
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
[1]Genome Institute of Singapore, Singapore

Computer analysis of higher-order chromosomal organization affecting transcription regulation in human is challenging problem of system biology. Evidence from *in situ* fluorescence studies suggest that clusters of genes actively transcribes are not random. Transcription in nucleus is concentrated within large discrete foci, assuming that genes are organized into "transcription factories" containing RNA polymerase II and other protein complexes. Chromosome Conformation Capture (3C) and similar microchip techniques along with traditional in situ techniques have demonstrated that chromatin interactions exist and regulate transcriptional and epigenetic states of genes. Several papers recently suggest modifications of high-throughput sequencing technique allowing detection of chromosome loops and interactions by whole genome sequencing, such as Hi-C, ChIA-Pet, TCC [1,2].

Genome-wide Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET) technology recently have published long-range chromatin interactions associated with RNA polymerase II in human cells and uncovered widespread promoter-centered interactions, further classified as intra-genic, extra-genic and inter-genic interactions [1]. Such interactions could be further aggregated into higher-order clusters, there proximal and distant genes are engaged through promoter-promoter interactions. Computer analysis of gene location and chromosome interacting sites revealed strong genome-wide association to transcription factor binding sites (in particular by the data from ENCODE project). In addition, analysis of chromatin interactions mediated by estrogen binding (ER-mediated interactome) confirmed associations of ER binding sites to RNA Pol II [3]. Overall, most genes with promoter-promoter interactions are highly active and could transcribe cooperatively, and that some interacting promoters could influence each other, implying combinatorial complexity of transcriptional controls. Comparative analyses of different cell lines (such as MCF-7, HeLa) imply that cell-line specific chromatin interactions could provide structural framework for the transcription. We found enrichment of ChIP-seq defined transcription factor binding sites from ENCODE project in human genome in spatial proximity

to chromatin bound contacting sites. Suggested models of chromosome 3D contacts such as fractal globule, continue discussion on structure function chromosome organization in nucleus [2]. Our study shows the three-dimensional basis of gene transcription activity and raises new bioinformatics challenges for transcription regulation modeling.

REFERENCES

1. G.Li et al. (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation, *Cell*, 148(1-2): 84-98.

2. J.R. Dixon et al. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions, *Nature*. 2012 Apr 11. (Epub ahead of print)

3. M.J. Fullwood et al. (2009) An oestrogen-receptor-alpha-bound human chromatin interactome, *Nature*, 462(7269): 58-64.

e-mail: orlov@bionet.nsc.ru

*Corresponding author

Computational study of protein-protein interaction networks

Miguel Andrade
Computational Biology and Data Mining group
Max Delbrück Center for Molecular Medicine,
Robert-Rössle-Str. 10, 13125 Berlin, Germany

Summary:

Protein-protein interaction (PPI) data is very useful to define protein function. These data is described in the literature and deposited in databases. However, the large amount and heterogeneous quality of such data complicates its use. I will present computational approaches that we are developing to facilitate the use of PPI data: the HIPPIE database of human PPIs scored according to experimental evidence (http://cbdm.mdc-berlin.de/tools/hippie/), the PESCADOR tool to extract PPI data from selections of medline abstracts (http://cbdm.mdc-berlin.de/tools/pescador/), and some applications where we used PPI data to find directed networks active in reprogramming or to study the function of polyQ in proteins.

REFERENCES:

1. Schaefer, M.H., J.F. Fontaine, A. Vinayagam, P. Porras, E.E. Wanker and M.A. Andrade-Navarro. 2012. HIPPIE: integrating protein interaction networks with experiment based quality scores. PLoS One. 7, e31826.

2. Schaefer, M.H., E.E. Wanker and M.A. Andrade-Navarro. 2012. Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks. Nucleic Acids Research. In press.

3. Barbosa-Silva, A., J.F. Fontaine, E.R. Donnard, F. Stussi, J. Miguel-Ortega and M.A. Andrade-Navarro. 2011. PESCADOR, a web-based tool to extract and study biointeractions from PubMed queries. BMC Bioinformatics. 12, 435.

4. Mah, N., Y. Wang, M.C. Liao, A. Prigione, J. Jozefczuk, B. Lichtner, K. Wolfrum, M. Haltmeier, M. Flöttmann, M. Schaefer, A. Hahn, R. Mrowka, E. Klipp, M.A. Andrade-Navarro and J. Adjaye. 2011. Molecular insights into reprogramming-initiation events mediated by the OSKM gene regulatory network. PLoS One. 6, e24351.

5. Lopes, T.J., M. Schaefer, J. Schoemaker, Y. Matsuoka, J.F. Fontaine, G. Neumann, M.A. Andrade-Navarro, Y. Kawaoka and H. Kitano. 2011. Tissue-specific subnetworks and characteristics of publicly available human protein interaction databases. Bioinformatics. 27, 2414-2421.

6. Vinayagam, A., U. Stelzl, R. Foulle, S. Plassmann, M. Zenkner, J. Timm, H.E. Assmus, M.A Andrade-Navarro, E.E. Wanker. 2011. Constructing directed protein interaction networks for investigating intracellular signal transduction. Science Signaling. 4, rs8.

e-mail: Miguel.Andrade@mdc-berlin.de

High-throughput analysis of human transpositional landscape

[1*]***Alexander Kanapin,*** [2]***Anastasia*Samsonova,*** [3]***Andrey Pisarev,*** [3]***Maria Samsonova***
[1] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK;
[2]Department of Genetics, Harvard Medical School, Boston, USA;
[3] St.Petersburg StatePolytechnicalUniversity,St.Petersburg,Russia

Transposable elements (TE) are discrete, repetitive, mobile DNA elements, which can freely move within genomes and sometimes between them and constitute about 45% to the human genome. A distinct class of TEs (retroelements) undergoing reverse transcription stage in their life cycle is of particular interest due to their abundance and harmful effect on genome and an organism as a whole. Whole genome sequencing provides excellent opportunity to study TE dynamics, their impact on genome stability and role in genetic innovation. However, it is even more important to associate above events with human diseases and reveal mechanisms, which could induce their onset.

We develop a robust computational pipeline for identification of novel TE insertions in the genome using information about members of various TE classes and whole genome sequencing data. The algorithm employs using of HMMs to generalize the knowledge about different TE families and their variability, alongside with existing assembly and mapping algorithms to identify TE insertions and discover their potential locations in the genome.

The pipeline is applied to publicly available personal genomic data as well as to the clinical data from patients with Mendelian diseases and various types of cancer. We provide TE insertion profiles to reveal pedigree- and disease- specific traits in transposition dynamics and as well as genomic loci susceptible to TE invasion in normal and pathological conditions.

e-mail: akanapin@well.ox.ac.uk
*Corresponding author

"Walking pathways" and how promoters can help to find new drugs

Alexander Kel
CSO, GeneXplain GMBH, Am Exer, 10 b, D-38302, Wolfenbüttel, Germany

Massive changes of expression of hundreds of genes as well as changes in genomic and epigenomic landscapes observed in human diseases often represent just an "echo" of relatively few causative molecular processes in the cells taking place during the transformation into the disease state, e.g. in cancer, during malignant transformation. Non-reversible structural changes in gene regulatory networks may cause transformation of the cell homeostasis switching it from the normal state to the disease state. We call such structural network changes as "walking pathways". Analysis of this phenomenon helps us to understand the mechanisms of molecular switches (e.g. between programs of cell death and programs of cell survival) and to identify perspective biomarkers and drug targets of cancer. In the current study we applied systems biology approaches to understand mechanisms of non-genotoxic carcinogenicity. EGF transgenic mouse model, as a model of sporadic liver cancer, was subjected to advanced methods of systems biology based on a combination of sequence analysis and entrained graph-topological algorithms. Promoter analysis of differentially expressed genes suggested the majority of regulated transcription factors to display specificity to either the pre-tumor or the tumor state. Many of those TFs could be confirmed by electromobility band shift assay at recognition sites of gene specific promoters and by western blotting of nuclear proteins. Subsequent search for signal transduction key nodes upstream of the identified transcription factors and their targets suggested the insulin-like growth factor pathway to render the tumor cells independent of EGF receptor activity. Notably, expression of IGF2 in addition to many components of this pathway was highly upregulated in tumors. Together, we propose a switch in autocrine signaling to foster tumor growth that was initially triggered by EGF and demonstrate the knowledge gain from promoter analysis combined with upstream key node identification. Deciphering the intimate mechanisms of the malignant transformation using systems biology approaches helping to combat cancer and other diseases is the ultimate goal of the systems medicine.

In order to automatize the causal analysis of high throughput data we have developed a geneXplain™ platform (www.genexplain.com) – an integrated systems biology platform. GeneXplain™ applies a unique upstream analysis approach based on implementation of machine learning and graph topological analysis algorithms in order to

identify causality keynodes in the network of gene regulation and signal transduction and combines it with full genome sequence analysis and chemoinformatics methods for drug discovery. The power of this approach is that we are trying to identify causal biomarkers - those which are more than just correlating with disease or treatment outcome but which are parts of the disease mechanism, which may differ in patient cohorts. Such personalized networks are analyzed in order to find key nodes - most important nodes triggering the disease. These keynodes and genes directly influenced by them are considered as promising biomarkers which can discriminate patients into cohorts from the disease mechanism point of view. In the current study, we analyzed a large scale gene expression and ChIP-seq data from a study of a breast cancer samples treated with antineoplastic agents including the novel drug compounds - RITA and Nutlin, targeting p53 and Mdm2. We analyzed promoters of downregulated pro-survival genes and identified combinations of transcription factors involved in their regulation. Topological modeling of the signal transduction network upstream of these transcription factors revealed key-nodes - potent master-regulators of the cell survival program that prevent efficient apoptosis of cancer cells. We considered these key-node proteins (e.g. PI3K subunits) as causal biomarkers as well as prospective targets for novel anticancer drug combinations. We applied a cheminformatics computer tool PASS to these targets and identified two novel prospective antineoplastic chemical compounds which were experimentally validated in a cellular assay confirming their synergistic potential in highly selective triggering of apoptosis of cancer cells.

e-mail: alexander.kel@genexplain.com
*Corresponding author

Complex molecular interactions as determinants of disease outcome and therapy response

Lodewyk F.A. Wessels
Bioinformatics and Statistics, The Netherlands Cancer Institute, Plesmanlaan 121
1066CX Amsterdam, Amsterdam, The Netherlands

Cancer arises as a result of the acquisition of DNA mutations. The exact mechanisms involved in tumor initiation, development and therapy response are still largely unclear. Two major challenges need to be addressed in order to systematically unravel these genetic interactions. First, large amounts of high quality data are necessary in order to ensure the statistical power required to uncover these genotype-to-phenotype relationships. Current advances in high-throughput biology are enabling the generation of such datasets. Second, suitable algorithms need to be developed to extract these interaction in a reliable fashion from these datasets. Here we report on two computational approaches we developed to approach this problem.

In the first approach, we developed a kernel-based, scale space approach to detect molecular interactions (co-occurrences and mutually exclusivities) from copy number data. We demonstrate the approach on 95 hematological cell lines and then apply it to a cohort of 313 breast cancer samples for which both gene expression and copy number data derived from the same tumor are available. Our analyses reveal genome-wide significant interactions that stratify the samples in distinct outcome groups. More specifically, we identify several gene-gene co-amplifications, which predominantly occur in a subgroup of breast cancer patients with particularly poor outcome. We validated these results on independent cohorts.

Second, we propose a novel computational approach based on integer programming that infers logic combinations of discrete genetic events (mutations, aberrations) that predict the observed phenotype. These logic formulas can be more complex than simple co-occurrences or mutual exclusivities. The use of a logic formalism enables the

formulation of intelligible models, which facilitates the speedy formulation of novel hypotheses and experiments. Here, we report on the application of this approach to a panel of 1000 cancer cell lines which is part of the Cancer Genome Project of the Wellcome Trust Sanger Institute. The mutation status of 66 known cancer genes has been characterized for each cell line in this panel. Additionally, the response of these cell lines to 250+ anti-cancer therapeutics have been recorded. Our models show that for most drugs, combinations of mutations explain the drug response better than single mutations. For example, of the eight BRAF inhibitors in the panel, the drug sensitivity to three of them is better explained using a logic combination of a BRAF mutation and one or more mutations in other genes, such as TET2. These results immediately suggest putative drug combination therapies. Additionally, cancer signaling pathways as annotated in e.g. the Pathway Interaction Database are employed to dramatically reduce the search space and offer a mechanistic explanation of the uncovered gene combinations.

e-mail: l.wessels@nki.nl

Formalizing and exploring top-down causation in morphogenesis and cell differentiation

L.V. Beloussov
Faculty of Biology, Lomonosov Moscow State University, Moscow 119991, Russia

One of the main features of a systems approach is the accent to the vertical (both down-top and top-down) causation in the multileveled systems to which all the biological objects belong. Among these two, top-down causation (TDC) is a modern incarnation of the old holistic tradition in developmental biology. In spite of a huge bulk of empirical data demonstrating the leading role of top-down effects in development, their proper formalization and a purposeful exploration is still rudimentary. The aim of this presentation is to join together the available but as yet scattered data related to TDC and to outline the way for further investigations.

Formalization. We argue that top-down effects in morphogenesis and cell differentiation imply non-local and geometry-dependent actions irreducible to those associated with any strictly pre-localized centers. In this respect, the concept of positional information (1) and the model of epithelial morphogenesis (2) are discussed. We suggest that TDC effects to a large extent can be reformulated in terms of parametric regulation.

Ways for exploring TDC. Fields of mechanical stresses in embryonic tissues, owing to their non-locality and geometry-dependence, are good candidates for mediating TDC. We review the modeling and experimental data illustrating their role in integrating the different levels events. In particular we show that the macroscopic deformations of amphibian embryonic tissues lead to rearrangements of expression sites of tissue-specific genes fitting the patterns of mechanical stresses (3). Together with other recent data, these results show that large scale macroscopic perturbations are able to affect the molecular level events.

REFERENCES

1. Wolpert L. (1969) Positional information and the spatial pattern of cellular differentiation. J. theor. Biol. V. 25: 1-47.

2. Belintzev B.N., Beloussov L.V., Zaraiskii A.G. (1987) Model of pattern formation in epithelial morphogenesis. J. theor. Biol. 129: 369-394.

3. Kornikova E.S., Troshina T.G., Kremnyov S.V. and Beloussov L.V. (2010). Neuro-mesodermal patterns in artificially deformed embryonic explants: a role for mechanogeometry in tissue differentiation. Devel Dynamics 239: 885-896.

morphogenesis@yandex.ru
**Corresponding author*

Dioxin in regulation of the genes involved in the cytokine synthesis by macrophages – a possible pathway underlying dioxin immunotoxicity and cancer

***D.Y. Oshchepkov*[1]*, E.A. Oshchepkova[1], E.V. Kashina [1], E.V. Antontseva [1], D.P. Furman[1,2], V.A. Mordvinov[1]**
[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia

Motivation and Aim: TCDD is the most toxic among the dioxin xenobiotics and induces a broad spectrum of biological responses, including also immunotoxicity and cancer [1]. Macrophages are key regulators of the innate immune response, as well as one of the first types of cells to respond to stress, and therefore it is important to study the action of TCDD in these cells in order to decipher the possible mechanisms of TCDD immunotoxicity and TCDD-induced cancer risk. TCDD mediates gene expression via AhR/ARNT transcription complex activation, which binds to dioxin responsive elements (DRE) in the regulatory regions of the inducible genes. TCDD acts as a stimulator of some inflammatory cytokines [2] and our subsequent analysis [3] showed that the list of stimulated cytokines is not yet complete. Also the possibility of direct as well as indirect regulation of cytokine synthesis via intrinsic transcription factors (TFs), through DREs in its regulatory regions was shown [3]. The subunits of NFκB, a key TF in the inflammatory response, are also in this list. Inflammatory conditions in turn increase the risk of cancer [4].

Methods and Algorithms: The analysis of the gene network of macrophage activation was performed; regulatory regions of the genes, involved in cytokine synthesis, were searched for DREs. The experimental verification was performed using U937 human macrophages cell line and 2nM TCDD concentration.

Results: DRE sites have been detected and confirmed by EMSA in number of macrophageal cytokine and TF gene promoters, with IL12a, IL12b, IL-4 and subunits of NFκB - Rel and RelA among others. Rt-PCR experiments confirmed TCDD–mediated modulation of the expression of these genes, ELISA experiments confirmed cytokine expression modulation on the protein level.

Conclusion: TCDD can directly mediate the gene expression of the IL12a, IL12b and IL-4 genes containing DREs in their regulatory regions, thus affecting the immune response. TCDD–mediated modulation of the NFκB subunits, in turn could be a possible pathway underlying dioxin-induced cancer risk.

Acknowledgements: This work was supported by RFBR (No. 12-04-01736).

REFERENCES:

1. P.K. Mandal. (2005) Dioxin: a review of its environmental effects and its aryl hydrocarbon receptor biology. *J Comp Physiol B*. 175: 221-230.

2. C.F. Vogel et al. (2007) Modulation of the chemokines KC and MCP-1 by 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in mice. *Arch Biochem Biophys*. 461:169-175.

3. Furman DP, et al. (2009) Promoters of the genes encoding the transcription factors regulating the cytokine gene expression in macrophages contain putative binding sites for aryl hydrocarbon receptor. *Comput Biol Chem*. 33(6):465-8.

4. Mantovani A. (2010) Molecular pathways linking inflammation and cancer. *Curr Mol Med. 10(4):369-73*.

e-mail: diman@bionet.nsc.ru

*Corresponding author

In silico verification of transcription factor binding sites in chip-seq data

V.G. Levitsky*[1], D.Y. Oshchepkov*[*1], G.V. Vasiliev[1], N.I. Ershov[1], T.I. Merkulova[1], Kulakovskiy[2,3] I.V., Makeev[3,2] V.J.

[1]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia;
[2]Engelhardt Institute of Molecular Biology RAS, Moscow, Russia;
[3]Vavilov Institute of General Genetics RAS, Moscow, Russia

Motivation and Aim: ChIP-Seq data analysis requires tools to identify transcription factor (TF) binding sites (BS). While there are many existing approaches for modeling of TFBS sequences, their effectiveness still raises many questions even for well-defined binding motifs. One example is FoxA TF, which is critical for liver development and function.

Methods and Results: Nucleotide sequences of 81 functional FoxA BS were manually retrieved from the literature. A subset of 53 BSs complying with degenerate motif TRTTTRYH [1] was used as a training set for SiteGA recognition method [2]. The "jack-knife" cross-validation test on this dataset has shown that SiteGA outperformed conventional dinucleotide position weight matrix (diPWM) by its accuracy. An improved dinucleotide version of the existing ChIPMunk algorithm [http://autosome.ru/ChIPMunk] was then applied to deduce the optimal alignment and corresponding diPWM model from 4455 sequences (peaks) having coverage no less than 15 in ChIP-Seq annotation [3]. A series of cross-validation (jack-knife) tests [2] was then performed to select the BS length of the diPWM model resulting in 28bp. To access recognition performance and select appropriate thresholds EMSA experiment were performed for the set of arbitrary selected 64 potential BSs. These BSs were extracted from ChIP-Seq annotation [3] with base coverage no less than 15 and mapped to 1kb long regions upstream transcription start sites. Identification of TFBS from initial ChIP-Seq data was performed with SiteGA and diPWM models with selected recognition thresholds.

Conclusion: ChIPMunk and SiteGA methods identified PBSs in 78.7 and 76.7% of 4455 peaks, respectively; 88.9% of peaks contained a PBS identified by at least one method; 56.5% peaks had PBSs identified by both methods and located not farther than 14bp apart. Only for 8.4% peaks missed PBSs predicted by any of two methods; these peaks may represent experimental errors, non-canonical or indirect TF-DNA interactions. Similar analysis of 4367 peaks [4] confirmed these conclusions.

Funding: This work was supported, in part, by a Dynasty Foundation Fellowship to I.K.

REFERENCES:

1. V.G. Levitsky et al. (2010) *VOGiS Herald*, 14: 685-697.
2. V.G. Levitsky et al. (2007) *BMC Bioinformatics*, 8: 481.
3. E.D. Wederell et al. (2008) *Nucl. Acids Res.,* 36: 4549–4564.
4. O. Wallerman et al. (2009) *Nucl. Acids Res.,* 37: 7498-7508.

diman@bionet.nsc.ru
*Corresponding author

Modelling Vascular Morphogenesis

Alvaro Köhn-Luque
Department for Innovative Methods of Computing
Technishe Universität Dresden, Germany

During vasculogenesis, isolated vascular cell progenitors assemble into characteristic network patterns. This type of *de novo* vascularisation is critical to many processes, ranging from embryonic development or a number of pathologies to successful tissue engineering. Despite the identification of VEGF as a crucial signalling agent, the mechanisms underlying the coalescence and patterning of vascular progenitors in these processes remain unclear.

In this lecture, I will present a number of mathematical and computational approaches to model and simulate vasculogenesis. Particular attention will be paid to describe a hybrid cell-based model that accounts for a novel patterning mechanism. The model can be used to explore the dynamics of network formation as well as the role of cell shape and cell density during embryonic vasculogenesis. I will also discuss some experimental validation efforts carried out so far as well as other related studies still in progress.

e-mail: alvarokohn@gmail.com

SESSION IV
MEDICAL SYSTEMS BIOLOGY

From Targets to Human: Engineering of Biotherapeutics

Matthias Machacek
Novartis Pharma AG, Switzerland

The application of Systems Biology to drug development is mostly associated with its use for target discovery. While target discovery is a critical step it is only the first one in a long sequence of equally hard challenges in bringing a new drug to human. Thus, focusing all the computational effort on a single development stage might not bring the desired improvement in success rate in drug development. To unfold the full potential of Systems Biology/Medicine and to have the biggest impact its use should be extended beyond target discovery to support the entire development cycle. This would include drug design, drug candidate selection, support of toxicology studies, design of first in man studies or dose finding. A few examples across the development cycle will be presented and the relevance of pharmacokinetics, pharmacodynamics, physiology, heterogeneity in disease and disease progression over time will be discussed.

e-mail: matthias.machacek@novartis.com

Application of systems pharmacology modeling in drug development

Oleg V. Demin
Institute for Systems Biology SPb, Moscow, Russia

Quantitative Systems Pharmacology (QSP) is an emerging modelling technique that combines the flexibility of systems biology and tractability of compartmental pharmacokinetic–pharmacodynamic modelling techniques [1]. QSP is applied for quantitative dynamic description of regulatory mechanisms of disease development/progression and mechanisms of drug action at intracellular/tissue/organism levels. QSP Modelling combines disease-specific description of intracellular pathways [2], cell dynamics, drug pharmacokinetics, tissue cross-talks and allows to express intracellular effects of the drug in terms of clinically measured biomarkers and end-points [3].

In my presentation the impact of QSP within drug discovery and development is considered by discussing several examples illustrating application of the modeling technique to resolve the problems arising in the field of pharmacology. In particular, QSP is applied to find out mechanism of action of anti-asthmatic drug Zileuton. On the basis of the mechanism we have developed the mechanistic explanation of complex relationships between Forced Expiratory Volume (FEV1) and dose/administration regime of Zileuton observed in clinical trials.

REFERENCES:

Agoram BM, Demin O; (2011) *Drug Discov Today*, Dec;16(23-24):1031-6.

Demin O., Goryanin I. (2008), *Kinetic Modelling in Systems Biology*. Taylor & Francis (United States), pp.360

Benson N, Cucurull-Sanchez L, Demin O, Smirnov S, van der Graaf P.; (2012) *Adv Exp Med Biol*.;736:607-15

e-mail: demin@insysbio.ru

PK properties optimization for Transglutaminase-2 inhibitor as a potential drug for celiac disease treatment

V.Sokolov *[*1], ***O.Demin Jr*** [1], ***S.Smirnov*** [1], ***L.Cucurull-Sanchez*** [2], ***C.Pichardo-Almarza*** [2], ***V.Flores*** [2], ***N.Benson*** [2], ***O.Demin*** [1]

[1] Institute for System Biology, Moscow, Russia

[2] Pfizer Global R&D, Sandwich, UK

Celiac disease is an autoimmune disorder that occurs in genetically predisposed people. It is caused by reaction to the gluten protein found in wheat that leads to villous atrophy. Nowadays there is no cure for celiac disease. The only known treatment is lifelong gluten-free diet. The main aim of our work is to predict the most effective method and time of administration and PK properties of transglutaminase-2 inhibitor as possible drug for celiac disease treatment.

Objectives:

To develop a detailed mathematical model of small intestine in order to describe spatial distribution of native and deamidated gluten peptides between lumen, lamina and blood in different parts of the intestine (duodenum, jejunum, ileum).

To develop PK model of TG2 inhibitor for the most effective and most safety PK parameters estimation.

Methods:

Circulator describes the whole organism as the system of interconnected compartments which can exchange their liquid medium (blood, lumen etc.). Transport of biological fluids between compartments is described with step-by-step algorithm and depends on only physiological properties of system (but not on metabolites concentrations). Each compartment can be described as the kinetic model that includes fixed phase (tissues) and movable phase (fluid). Whole system is described as a system of ordinary differential equations with cell numbers and concentrations as variables. Model was verified and validated against all available in vivo data.

Results:

Oral administration of TG2 inhibitor is more effective than intravenous. Optimal values for PK parameters were determined. Administration of TG2 inhibitor just before meal proved to be optimal for treatment.

TG2 distribution between lumen and lamina along with zonulin properties (“liquid” or “fixed”) doesn’t have a significant influence on TG2 inhibition by hypothetical drug, intestine cells, antibodies and cytokine levels.

Conclusions:

The developed mathematical model allows us to estimate and to predict the most effective PK properties, method and time of TG2 inhibitor administration, the influence of different conditions (TG2 distribution or zonulin properties) on drug efficacy and safety. This model could be used for drug discovery and development for autoimmune disorders in small intestine and for design of clinical trials.

sokolov@insysbio.ru
**Corresponding author*

Exploring ensemble properties of serum antibody binding: Mathematical modeling and analysis of antibody reactivity data

Michal Or-Guil
Dept. of Biology, Humboldt University Berlin
Research Center ImmunoSciences, Charité
Hessische Str. 3/4, 10115 Berlin

Antibodies are proteins used by the immune system to recognize molecular structures. The binding of antibodies to foreign molecules is a hallmark of humoral immunity. However, the antibody mixture contained in blood is of unknown complexity.

To gain insight into recognition capabilities of antibody mixtures, we formulated a minimal model of antibody binding. We found that recognition is governed by ensemble properties if the antibody mixture is random and highly diverse. These properties imply that recognition of peptides by such a mixture does not depend on the position of amino acids within the peptide sequence, and that assigning a weight to each amino acid is sufficient to describe and predict peptide recognition. Since this recognition does not depend on the exact antibody repertoire, it defines a signature of random and highly diverse antibody mixtures that is only disrupted when few antibodies dominate.

To verify our findings, we probed peptide microarrays with blood sera of healthy and infected mice and analyzed the resulting antibody binding profiles. Our results indicate that position-independent amino acid-associated weights predict peptide binding by sera of healthy individuals quite well and provide for a "signature of health". These findings have implications for both serological diagnostics and B cell epitope mapping.

e-mail: michal.orguil@gmail.com

Modeling of celiac disease immune response and therapeutic effect of potential drugs.

O.Demin Jr*[*1], S.Smirnov[1], V.Sokolov[1], L.Cucurull-Sanchez[2], C.Pichardo-Almarza[2], V.Flores[2], N.Benson[2], O.Demin[1]

[1] Institute for System Biology, Moscow, Russia

[2] Pfizer Global R&D, Sandwich, UK

Motivation:

Celiac disease is an autoimmune disorder that occurs in genetically predisposed people and is caused by reaction to gluten protein found in wheat that leads to villous atrophy. The only known treatment is lifelong gluten-free diet. The main aim of our work is to predict the efficacy of transglutaminase-2 inhibitor as possible drug for treatment of celiac disease.

Objectives:

To develop a mathematical model of innate and adaptive immune response in celiac disease, including influence of different cells and proteins on villous area, deamidation of gluten peptides and antibody synthesis.

Use this model to predict the efficacy of transglutaminase-2 inhibitor and other possible drugs for treatment of celiac disease.

Methods:

A mathematical model was developed integrating all known *in vitro*, *in vivo* and clinical data about relevant cells and proteins taking part in immune response in celiac disease. This model includes the following components: (i) innate immune response, (ii) deamidation of gluten peptides by transglutaminase-2, (iii) adaptive immune response.

Results:

Transglutaminase-2 inhibitor treatment leads to decreasing of antibodies level only in 2-3 times, so it remains higher than healthy people antibodies level.

Transglutaminase-2 inhibitor treatment doesn't lead to significant increasing of villous area.

The most effective possible drug for treatment of celiac disease is gluten peptide analogs that bind to APC instead of immunogenic gluten peptides, but don't activate it.

Conclusions:

The developed mathematical model of immune response in celiac disease allows to predict efficacy of transglutaminase-2 inhibitors and other possible drugs for celiac

disease treatment. This model could be used for drug discovery and development for autoimmune small intestine disorders and for design of clinical trials.

e-mail: demin_jr@insysbio.ru
*Corresponding author

Mechanistic modeling approach relating human gut microbial community to physiologically-relevant biomarkers

K. Peskov, Yu. Kosinsky*
Modeling & Simulation, Novartis Pharma LLC, Moscow, Russia

The adult human gut houses a microbial community which contains a large number of bacterial species. It is well-known now that the actual composition of this community has a significant influence on human vital functions and may be an important determinant of various pathologies (*e.g.*, obesity, inflammation). However, the mechanisms controlling the assembly of gut microbiota and its relationship with host human tissues remain poorly understood. This paper represents a first attempt in developing an integrated quantitative understanding of factors relating gut microbiota to measures of physiologically-significant blood plasma biomarkers.

Based on the results of [1], showing that the human gut microbial community is typically formed by two bacterial phyla (Firmicutes and Bacteroidetes), we developed various sub-models describing generalized metabolic peculiarities of these bacteria, their sensitivity to various nutrients, and processing of endpoint metabolites such as short-chain fatty acids. Individual sub-models were integrated to provide a unified model of microbiota relationships with host tissues. Model predictions were verified against experimental data from the literature, on qualitative and quantitative gut microbial composition, biochemical characterization of particular bacteria, and results of gnotobiotic mice colonization by various microbial cultures.

All individual sub-models provided adequate descriptions of isolated interactions. The integrated model provided good descriptions of literature-reported changes in butyrate, acetate and propionate in response to different bacterial composition (in accordance to the data published in [2]). It was shown that different steady-state ratios of short-chain fatty acids produced by one or another microbial composition can be considered as risk factors for obesity.

A mechanistic model of the relationship between human gut microbial community and host tissues was developed. The model can be used to evaluate the potential effect of various compositions of microbial community to the steady state ratios of short-chain fatty acids and *in silico* testing of possible therapies related to interventions and changes in gut microbial composition.

REFERENCES:

1. Turnbaugh P. *et al.* (2009) A core gut microbiome in obese and lean twins, Nature, 457: 480-484.
2. Mahowald M. *et al.* (2009) Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla, PNAS, 106: 5859-5864.

email: yuri.kosinsky@novartis.com
*Corresponding author

Therapy of chronic myeloid leukaemia: simulation studies of different treatment combinations and patient-specific risk estimation

Ingmar Glauche
Institute for Medical Informatics and Biometry (IMB),
Faculty of Medicine Carl Gustav Carus, Dresden University of Technology

Newly diagnosed patients with chronic myeloid leukaemia (CML) are currently treated with tyrosine kinase inhibitors (TKIs) such as Imatinib, Nilotinib or Dasatinib leading to a biphasic but heterogeneous decline of BCR-ABL transcript levels. However, incomplete eradication of residual disease is a general problem of long-term TKI therapy, and the question remains when a therapy can safely be stopped. Activation of mouse haematopoietic stem cells (HSCs) by Interferon-α (IFNα) stimulated the discussion of whether a combination treatment leads to accelerated eradication of the CML clone.

We base our simulation approaches on a mathematical model describing human CML as a competition phenomenon between normal and malignant cells. We amend this model to incorporate the description of IFNα activity and simulate different scenarios for potential treatment combinations.

Our model predicts the long-term response to Imatinib based on the observed BCR-ABL transcript levels for individual patients. We derive a simple predictor to assess the individual risk of molecular relapse upon therapy discontinuation. In a second step we demonstrate that the overall sensitivity of CML stem cells to IFNα activation is a crucial determinant for the benefit of a potential combination therapy. We furthermore show that pulsed IFNα together with continuous TKI administration is the most promising strategy for a combination treatment in which the therapeutic benefit prevails adverse side effects.

Our modelling approach is a highly beneficial tool to quantitatively address the competition between normal and leukaemic haematopoiesis in treated CML patients. We derive testable predictions for different experimental settings that are suggested prior to the clinical implementation of different therapeutic strategies.

e-mail: ingmar.glauche@tu-dresden.de

Google goes cancer: Improving outcome prediction for cancer patients by networkbased ranking of marker genes

Roy Janine*[1], *Winter Christof*[1], *Isik Zerrin* [1], *and Schroeder Michael*[1]
Biotechnology Center, Dresden University of Technology, Tatzberg 4749, 01307 Dresden, Germany

Disease progression in cancer can vary substantially between patients. Yet most patients receive the same treatment. Recently, there has been much work on predicting disease progression and other outcome variables from gene expression to personalize treatment options. Despite first diagnostic kits on the market, found marker genes often show limited prediction accuracy, limited reproducibility, and unclear biological relevance. In order to solve these problems, we developed a novel outcome prediction algorithm *NetRank* to identify genes prognostic for outcome using both expression data and network information. Our approach adapts the random surfer model of Google's PageRank algorithm to rank genes according to their prognostic relevance. We applied the algorithm to gene expression profiles obtained from 30 pancreas cancer patients, and identified seven candidate marker genes [1]. The accuracy of NetRank was assessed by using support vector machine classifiers and Monte Carlo crossvalidation scheme. Compared to genes found with state of the art methods, NetRank improves the outcome prediction accuracy by 7%. When experimentally validating the prognostic value of our seven candidate markers on an independent set of 412 pancreatic cancer samples, we achieved an accuracy superior to established clinical prognostic factors. Besides, we systematically evaluated the prognostic power of networks and NetRank for signature identification on 25 published cancer datasets. Our results indicate that NetRank algorithm performs better than classic feature selection methods such as Pearson correlation, ttest and fold change. In addition, reproducibility of signatures created by NetRank increases significantly between different datasets of the same cancer type. As a conclusion, network based gene expression analysis leads to more detailed understanding of cancerrelated processes and in future signatures predicted by NetRank could be applied in daily clinical routine.

REFERENCES

1. Winter C. et al. (2012) Google goes cancer: Improving outcome prediction for cancer patients by networkbased ranking of marker genes, *PLOS Computational Biology*, in publication.

e-mail: ms@biotec.tudresden.de
*Corresponding author

Dissecting cancer heterogeneity

Florian Markowetz
Cancer Research UK, Cambridge Research Institute,
Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, UK

I will talk about computational methods to address heterogeneity of breast and ovarian cancer at different levels:

(I) At the *population* level, cancer risk is distributed unevenly. I will discuss recent results on understanding the function of FGFR2, the top breast cancer GWAS hit.

(II) At the *patient* level, different cancer sites can differ significantly in the genomic aberrations they carry. I will discuss how to quantify intra-patient heterogeneity and how to leverage it to infer driver mutations.

(III) At the *sample* level we often find cancer cells mixed with immune cells, stromal cells and others. This mixture of cells leads to a mixture of signals when DNA, RNA, or proteins are measured in these samples. I will present an automated and quantitative approach to estimate cell mixtures and devolute molecular signals.

e-mail: florian.markowetz@cancer.org.uk

SESSION V
SIGNALING AND CELLULAR REGULATION

Combinatorial and Dynamic Control Logic within pathogen-responsive Gene Regulatory Networks

Alexander Hoffman*
UC San Diego, San Diego Systems Biology Center

Studies of the cellular responses to pathogens have identified several primary response transcription factors (TFs), that are activated in stimulus-specific combinations and temporal profiles. To understand how dynamic activities may combine to produce the regulatory logic of pathogen-responsive gene expression, we utilized mathematical modeling to guide the analysis of a multi-dimensional expression dataset of 714 transcripts induced by bacterial endotoxin. We found that gene clusters are controlled by signal-responsive TFs either singly or sequentially in OR gates, but that further specification of expression programs is mediated by constitutive and signal-responsive mRNA halflife control. The results reveal that mRNA halflife control is responsible for decoding not only stimulus-responsive TF dynamics, but also pathway combinatorics. We surmise that predictive models of gene regulatory networks (GRNs) cannot be based on chromatin-associated events alone but must include non-nuclear control mechanisms as well. Our work begins to delineate how intra-cellular combinatorial and dynamic signals that encode information about the extra-cellular stimulus are decoded through specific mechanisms within GRNs.

e-mail: ahoffmann@ucsd.edu

Systems-biology approaches to understanding network functions of the human brain

Simon B. Eickhoff
Cognitive Neuroscience Group, Department of Clinical Neuroscience and Medical Psychology, Heinrich-Heine University Düsseldorf
Brain Network Modelling Group, Institute for Neuroscience and Medicine (INM-1), Research Center Jülich

Connecting different aspects of brain connectivity. Understanding interactions and network-functions of the human brain have become an important focus in systems neuroscience research. There is, however, no such thing as «the» connectivity between two brain regions. Rather, brain connectivity and interactions in cortical networks may be described along various dimensions using different imaging methods and analysis approaches. Most commonly, however, three fundamental concepts of connectivity are distinguished. Structural or anatomical connectivity describes the presence of fibers connecting two regions, functional connectivity denotes the temporal synchronization of measured activity and effective connectivity uses models to ascribe causal influences between different areas. This talk first juxtaposes the different concepts and their strength and weaknesses from a theoretical and practical perspective. The main focus of this presentation will then rest on studies aiming at a direct comparison or joint analysis of different aspects of brain connectivity. While these have in several instances demonstrated converging evidence for the presence or strength of several connections, they also highlight that each approach reflects unique aspects of brain networks and the interactions within them. In summary, present data on the relationship of different connectivity measures thus strongly indicates that these may represent complementary measures and that an important perspective but also major challenge lies in their integration for a more comprehensive characterization of cortical networks.

e-mail: S.Eickhoff@fz-juelich.de; Simon.Eickhoff@uni-duesseldorf.de

The microRNA determine the early stage dynamics of the regulation network

***M. A. Duk*[1], *A.M. Samsonov**, *M.G. Samsonova*[1]**
The A.F.Ioffe Physical Technical Institute of the Russian Academy of Sciences,
St.Petersburg, 194021 Russia
[1]State Polytechnical University, St.Petersburg, 195251 Russia

One of the regulatory mechanisms of gene expression extensively studied in the last years involves miRNA, see, e.g., [1,2]. These small regulatory molecules bind a recognition sequence of the target protein-coding matrix RNAs (mRNAs) and preclude them from translation. Recent studies showed that miRNAs participate in buffering of genetic noise in the regulatory systems and in the reduction of the phenotypic variability. It is realized now that this function of miRNA can be explained only via understanding its role in the regulatory network comprising miRNA interacting with transcription factors (TFs). One of the modern interaction models was introduced in [3] in the form of the coupled kinetic o.d.e. to describe either coherent or incoherent network. In the *coherent* network both pathways from the TF to the target protein exhibit the same action (repression or activation of the target expression), while in the *incoherent* one the two pathways have opposite actions. The system was solved in [3] in the steady state approximation ($\partial/\partial t = 0$) in both deterministic and stochastic statements. The transcription rates of the miRNA gene and of the target gene were assumed to be the Hill functions of the number of TFs (q), while the translation rate of the target gene - a repressive Hill function of the number of miRNAs (s). Evidently, one and the same stationary solution to the system can fit to solutions, which are completely different at the early stage of development, and some of them may *not* even have any biological sense, that cannot be verified in the algebraic version of the o.d.e.

We obtained the *exact solutions to the coupled o.d.e.* under some biologically relevant restrictions and found the detailed dynamics of both coherent and incoherent networks. The remarkable variations of concentration at early stage were shown, invisible in the steady state solutions. We have found that depending on the initial conditions several concentrations, being acceptable at steady state, may be negative at early stage, which is inconsistent with biology. Exact solutions found to the coupled o.d.e. are useful to reconstruct a proper parameter set for modelling in order to avoid meaningless concentration values at early stage in both deterministic and stochastic statements.

The support by the RFBR grant no. 11-01-00573 and the State scientific educational contract no.14.740.11.0166 is gratefully acknowledged.

REFERENCES

1. N.Bushati,S.M.Cohen. (2007). MicroRNA functions. *Annu. Rev. Cell Dev. Biol*. 23:175–205.

2. D.P.Bartel. (2009). microRNAs: Target recognition and regulatory functions. *Cell*, 136:215-233.

3. Osella M, Bosia C, Cora D, Caselle M. (2011) The Role of Incoherent MicroRNA-Mediated Feedforward Loops in Noise Buffering, *PLoS Comput. Biology*, 7:e1001101.

e-mail: samsonov@math.ioffe.ru

*Corresponding author

Towards understanding of apoptosis regulation by systems biology

Kolja Schleich[1,2,*], Uwe Warnken[3,*], Nicolai Fricker[1,2], Martina Schnölzer[3], Peter H. Krammer[1], Inna N. Lavrik[1,2,4,5]

[1]Division of Immunogenetics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany
[2]Bioquant, 69120 Heidelberg, Germany
[3]Functional Proteome Analysis, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany
[4]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany
[5]Department of Translational Inflammation Research, Institute of Experimental Internal Medicine, Otto von Guericke University, Magdeburg, Germany

In our studies we apply systems biology to investigate apoptosis regulation and development of drugs that interfere with diseases associated with defects in apoptosis. Our interest is especially connected to CD95/Fas-induced apoptosis which is one of the most important apoptotic pathways. The CD95 (Fas/APO-1) death-inducing signaling complex (DISC) is essential for the initiation of CD95-mediated apoptotic and non-apoptotic responses. In our recent studies we have investigated a stoichiometry of the CD95 DISC using quantitative Western blots, mass spectrometry, and mathematical modeling. Mathematical modeling of the CD95 DISC was based on agent-based modeling. The agent-based modeling was more advantageous in this system compared to Ordinary Differential Equations (ODEs) that were mostly applied before for modeling the CD95 DISC and apoptotic networks. This study elucidated the new insights into apoptosis regulation and dynamics of the CD95 DISC. The implications of our findings will be further discussed.

e-mail: lavrik@mail.med.uni-magdeburg.de
*Corresponding author

Using data-driven models of feedback regulation in EGFR signal transduction to predict optimal targeted therapy in colon cancer

Nils Bluethgen
Humboldt University zu Berlin, Institute for Theoretical Biology, Invalidenstrasse 43 10115 Berlin

Alterations in the EGFR signalling networks are common in most solid tumours, and targeted therapies exist to modulate the activity of these pathways. However, success of these therapies is limited, and it remains unclear which molecule to target under which mutational pattern. Thus, molecular markers are urgently needed to stratify patients. To address this questions mechanistically, we developed and applied a combined experimental and computational approach to systematically quantify signal transduction downstream of the EGFR in a panel of colon cancer cell lines. Experimentally, we performed defined perturbations (using ligands to receptors and experimental drugs alone and in combination), and measured signaling status by a high-sample-throughput proteomics platform. We used this data to estimate parameters for mathematical models to quantitatively describe the signaling pathways by an approach similar to modular response analysis. We developed an analytical identifiability analysis, which automatically reduces the model such that we can reliably estimate the parameters and compare these between different cell lines. We used this approach to identify mechanisms that counteract drug action, such as feedback loops. By using next generation sequencing, we characterize the panel of cell lines and correlate the quantitative traits of signal transduction with genotype. This allows to predict successful intervention depending on mutation. Using phenotypic assays, we confirm our predictions.

e-mail: nils.bluethgen@charite.de

Architecture and Dynamics of the Mitochondrial Reticulum in Healthy Systems and in Age-Related Disorders

Valerii M. Sukhorukov [1,2*], ***Daniel Dikov*** [3,4], ***Andreas S. Reichert*** [3,4], ***Michael Meyer-Hermann*** [1,5]
[1] Helmholtz Centre for Infection Research, Braunschweig, Germany;
[2] Frankfurt Institute for Advanced Studies, Frankfurt am Main, Germany;
[3] Frankfurt Institute for Molecular Life Sciences, Frankfurt am Main, Germany;
[4] Goethe University of Frankfurt am Main, Frankfurt am Main, Germany;
[5] Technical University Braunschweig, Germany

A reticular three-dimensional network of mitochondria spans intracellular volume outside the nucleus. Despite considerable advances in understanding of the mitochondrial physiology and biochemistry, neither quantitative characterization of its architecture, nor rationale behind remarkable structural flexibility of this organelle is readily available. Using a combination of computational, mathematical and experimental techniques, we investigate mitochondrial large-scale organization and dynamics. Analysis of the reticulum recreated in silico indicates that a multitude of its basic biological properties may stem from the network's vicinity to a structural phase transition point. Cellular quality control mechanisms counteract the damaging influence of reactive oxygen and nitrogen species generated by mitochondria. Efficiency of their operation is currently the in hotspot of numerous experimental studies because of its importance for the long-term cellular survival. Explicit representation of the chondriome within our agent-based model provides detailed insights into the homeostasis of this organelle. It is used for pinpointing the key processes governing efficient turnover of mitochondrial material and denotes the progression of selective but stochastic autophagy due to successful isolation of the least efficient network components. The system is then applied for the investigation of irregularities in mitochondrial dynamics accompanying cellular aging and the resulting late-onset common disorders like the Parkinson disease, type II diabetes and cancer.

e-mail: info@helmholtz-hzi.de
*Corresponding author

Specificity and flexibility of RAS-ERK signalling by slow and fast loops

Ovidiu Radulescu
DIMNP
Universite' de Montpellier 2
Place Euge`ne Bataillon
34095 Montpellier Cedex 5
France

Flexibility and specificity are important properties of signalling cascades, enabling reliable processing of a multitude of signals and eliciting different phenotypes and cell fate decisions by shared pathways. Feed-forward and feed-back control can generate complex dynamics of signalling networks and were used elsewhere to explain pathway flexibility and specificity of cell fate decision. In this paper we discuss the consequences of multiple time scales on the response of two signalling modules, the RAS switch and the MAPK cascade, coupled by both transcriptional and post-translational interactions. Using this model we reconcile several datasets concerning flexible cell fate response in various human cell types and elucidate the observed complex dynamical behavior of these pathways. In particular, we discuss the dual response of K-RAS dependent malignant cell lines to stimulation by fibroblast growth factor 2 (FGF2), a bona fide oncogenic factor.

e-mail: ovidiu.radulescu@univ-montp2.fr

Designing biocomputing devices based on RNA-RNA interactions

Ilka Maria Axmann
Institute for Theoretical Biology (ITB), Charité-Universitätsmedizin Berlin, Germany

In the field of synthetic biology, newly designed molecular devices enable manipulation of biological systems. Cells with new properties can be derived and tested for biotechnological and medical applications. Further, the redesign of cellular systems offers an effective strategy to get a deeper understanding of biological processes in general. We aim at programming cellular networks using newly engineered devices that transmit and process information within cells in an innovative way, namely via RNA-RNA interactions. RNA has three major advantages: (1) RNA turnover is fast thus efficient computing is possible with RNA networks, (2) RNA folding and RNA-RNA interactions can be well predicted, thus, a large number of novel devices can be build, (3) RNA production is energetically cheap, therefore the host cell is not affected by computing. For example, we designed a genetic circuit in E. coli based on RNA-RNA-interactions that exhibits XOR-gate behaviour in terms of digital Boolean circuits. This logic operator compares different inputs and gives an output signal only if inputs are unequal. In case inputs are equal the output signal is zero. By combining experimental and in silico approaches, we can systematically analyse dynamics and quantitative properties of regulatory RNA-based cellular signalling.

e-mail: ilka.axmann@biologie.hu-berlin.de

Modeling and analysis of dynamics of the gene networks: automatic generation and storage in a new database

***Akberdin I.R.*[1*], Kazantsev F.V.[1], Ree N.A.[1], Timonov V.S.[3,4], Oshchepkova E.A.[1], Ratushny A.V.[5], Khlebodarova T.M.[1], Fadeev S.I.[2,3], N.A. Kolchanov[1,3] and Likhoshvai V.A.[1,3]**

[1]-Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
[2]-Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
[3]-Novosibirsk State University, Novosibirsk, Russia
[4]-Siberian State University of Telecommunications and Information Sciences, Novosibirsk, Russia
[5]-Institute for Systems Biology, Seattle, USA

Mathematical models of molecular-genetic systems are based on the information about the structural and functional organization of gene networks and their dynamic properties that disseminated over hundreds and thousands of scientific papers. The problem arises of data comparison and analysis of non-uniformed experimental data, analysis of cause-and-effect relations between molecular structure, dynamics and phenotypic features of molecular-genetic system, and software development for automatic generation of mathematical models, storage of creating models in the database and their numerical analysis.

In the context of solving some of the above mentioned problems we have developed an integrated computer system and models database (http://modelsgroup.bionet.nsc.ru/MGSmodelsDB/) that do not only render automatically the process of mathematical models reconstruction based on the structural and functional organization of gene networks but also implements original approaches and algorithms to modeling and studying molecular-genetic systems.

The current version of the database includes models of enzymatic reactions and gene expression regulatory processes of *de novo* nucleotide metabolism and anaerobic respiration in *Escherichia coli*. The database contains 110 elementary models, each of which represents an enzymatic reaction or regulatory process rate function. Model parameters were obtained from published data or fitted to the available experimental data. MGSmodelsDB enables to search, select and automatically generate more complex models in SBML and other formats from a subset of elementary models.

The examples of the system usage are demonstrated on a modeling of some gene regulatory networks.

Acknowledgements

Funding: This work was partially supported by the RFBR [10-01-00717-a], by the "PATHOSYS" grant [260429] of the Seventh Framework Programme, by the Interdisciplinary Integration Projects of the Siberian Branch of the Russian Academy of Sciences [80]. This study was also partially supported by the Scientific school № 5278.2012.4 and Programs of the Presidium RAS "Molecular and cell biology" (6.8) as well as "Biological diversity" (30.29).

e-mail: akberdin@bionet.nsc.ru
*Corresponding author

Comprehensive analysis and computational modeling of microtubule regulation in neuromorphogenesis

Julia Arens[1,2], Tomáš Mazel, Anja Biesemann and Leif Dehmelt[1,2*]
[1]Fachbereich Chemische Biologie, Dortmund University of Technology
[2]Department of Systemic Cell Biology, Max Planck Institute of Molekular Physiology
D-44221 Dortmund, Germany

The first morphological change after neuronal differentiation is the initiation of neurites, thin cell protrusions, which later differentiate into axons and dendrites. During neurite initiation, microtubules are arranged into parallel arrays within the neurite shaft. Microtubule-related genes are candidates for mediating this process. We performed a morphological screen to evaluate the role of 408 microtubule-related genes during neuromorphogenesis. SiRNA-mediated knockdown of several genes leads to dose-dependent appearance of distinct phenotypes, including alterations in neurite length, neurite diameter and neuronal differentiation efficiency. Phenotypic classification revealed specific dynein complex subunits involved in neurite initiation, cell cycle regulators which influence neuronal cell body size and neurite diameter, and novel roles for microtubule plus end tracking proteins in the regulation of neurite outgrowth. Based on our analyses, we built a stochastic computational model to simulate cytoskeletal interactions during neuromorphogenesis.

e-mail:leif.dehmelt@mpi-dortmund.mpg.de
*Corresponding author

The Virtual Liver Network - an overview

***Johannes Bausch**, Adriano M. Henney*
Virtual Liver Network, University Heidelberg,Im Neuenheimer Feld 327
69120 Heidelberg Germany

The liver is the central metabolic organ in human physiology, with functions that are fundamentally important to the detoxification of drugs, the maintenance of homeostasis of numerous blood metabolites and the production of mediators of the acute phase response. Liver toxicity, whether actual or implied is the reason for the failure of a significant proportion of many promising novel medicines that consequently never reach the market, and diseases such as atherosclerosis, diabetes and fatty liver, that are a major burden on current health resources, are directly linked to functional and structural disorders of the liver.

The German Virtual Liver Network is one of the most exciting and ambitious modelling projects in the field of systems biology and systems medicine. This major multidisciplinary research programme (funded for five years with 44 Million Euro from the German Federal Ministry of Eduction and Research) is aimed at developing a whole-organ model of the human liver, representing its central physiological functions under normal and pathological conditions. The model will be composed of a larger battery of interconnected sub-models representing liver anatomy and physiology, integrating processes across hierarchical levels in space, time and structural organisation.

e-mail: johannes.bausch@virtual-liver.de
*Corresponding author

Systems Biology of Lung Cancer – Linking Dynamic Properties of Signaling Networks to Migratory Behavior of Lung Carcinoma Cells in Monolayer Culture.

***Sofia Depner**, Lars Velten, Ruth Merkle, Aastha Mathur, Franziska Matthaeus and Ursula Klingmüller*
Group Systems Biology of Signal Transduction, German Cancer Research Center (DKFZ) Im Neuenheimer Feld 280
69120 Heidelberg, Germany

Lung cancer, with its predominant form non-small cell lung cancer (NSCLC), is the leading cause of cancer related-deaths world-wide. The key medical problem in lung cancer is early systemic spread of tumors independent of tumor size. Activating mutations or overexpression of receptor tyrosine kinases such as hepatocyte growth

factor (HGF) receptor (c-Met) are frequently observed. To unravel key regulatory mechanisms promoting early spread and to establish a quantitative link of information processing through signaling networks, target gene induction and lung cancer cell migration, we examined in monolayer cultures of selected NSCLC adenocarinoma cell lines HGF mediated signal transduction, and induction of target genes, as well as cellular decisions such as proliferation, migration and apoptosis at the cell population level. The quantitative data obtained by quantitative immunoblotting, mass spectrometry, quantitative real time PCR and NGSeq was used to establish detailed mechanistic models of signaling pathways that will be linked to gene regulatory network models and correlated to cellular responses.

e-mail: s.depner@dkfz.de
*Corresponding author

Genetic based method for discrimination of lipoproteins level in citizens of North-West region of Russia

***Glotov A.S.*[1,2*], Glotov O.S.[1,2], Vashukova E.S.[1,2], Marochkina E.U.[2], Tarkovskaya I.V.[2], Makchrova I.A[2], Pugacheva I.V[2], Kurilov R.V.[2], Aseev M.V.[1,2], Ivashchenko T.E.[1,2], V.S.Baranov[1].**
[1]Laboratory of prenatal diagnostic, The D.O.Ott Research Institute of Obstetrics and Gynecology, Saint-Petersburg, 199034 Russia
[2]Genetic Laboratory, Ltd BioGlot, Saint-Petersburg, 197341 Russia

Prediction of human QTLs on the basis of genetic testing is an important goal of genomic fingerprinting, population and medical genetics. GWAS studies in several population detected genes polymorphisms associated with QTL such as a human height, weight and level of lipoproteins. The aim of this study was to develop a method for assessing of low and high density lipid levels based on genetic testing in citizens of North-West Region of Russia. For candidate locus we selected genes which are associated with lipids level in previous studies, repeated in independent GWAS studies and associated with changing of proteins level. Polymorphisms are functional, located in promoters or splicing regions of gene. Frequency of minor alleles is 10-40%. The polymorphisms of *CETP* (rs3764261), *LPL* (rs328), *LIPC* (rs1800588), *ABCA1 (*rs3890182), *LCAT* (rs255049), *APOA1* (rs5128), *APOB* (rs6754295), *FAD (*rs174570), *PLTP* (rs378114), *SCARB1* (rs5888), *MC4R (*rs12970134), *ANGPTL3 (*rs10889353), *GCKR (*rs1260326), *AKR1D1 (*rs3735023), *CEL (*rs2013751), *CYP7A1 (*rs3808607), *CILP2 (*rs16996148), *ABCG5 (*rs6756629), *NPC1L1 (*rs17655652), *LIPE (*rs34845087), *PPARG (*rs10865710),

*NEGR1 (*rs2568958), *FTO (*rs1421085), *LEPR (*rs1137101), *GHR (*rs4410646), *ADCY3 (*rs2033655), *TNFA (*rs1800629), *LPIN1 (*rs2716609), *NRXN3 (*rs10146997), *TFAP2B (*rs987237), *SH2B1 (*rs7498665), *GHRL (*rs2075356), *ESR1 (*rs9340799), *MSRA (*rs545854) were studied (N=350). DNA was extracted from the blood cells. Then the polymorphisms of 34 genes were determined by PCR-RFLP method. A lipoprotein level was detected by Elecsys-2010. To construct the predictions model of lipid levels have used linear regression. We constructed several prediction models for high and low density lipoproteins levels. Pprediction model of HDL levels included such parameters as sex, *LIPC*, *SCARB1, PPARG, TNFA, ESR1, MSRA, TFAP2B, CEL* (multiple R-squared=0.37, $p=1.53*10^{-6}$). Common prediction model of LDL levels had low R-squared (0.04). Therefore, we used prediction models for men and women separately. For man it were be *CEL, APOA1, LIPC, PPARG1, NEGR1, SH2B1, FTO, MC4R, GHR, GHRL, LPIN1, NRXN3* genes (R-squared=0.72), for women - *CILP, APOB, SCARB1, LIPE, GHRL* genes (R-squared=0.17). We believe that using of regression model of HDL and LDL level would become the convenient tool for medical genetic research.

The work was supported by grant of Russian Ministry of Education and Science (GK №16.512.11.2035).

e-mail:anglotov@mail.ru
* Corresponding author

Novel insights into epidermal tissue repair using 3D tissue cultures

Niels Grabe
National Center for Tumor Diseases, University Hospital Heidelberg

Human epidermal wound healing provides an ideal case study for tissue systems biology. Until now controversial results and mechanisms have been proposed to understand how human epidermal wounds close in 3D by integrating cell migration, differentiation and proliferation. Especially collective cell migration is assumed to play a role, but it remains still unclear how it serves to close the wound. We established a novel medium-throughput wound healing assay with organotypic 3D skin cultures and performed a series of analyses on over hundred in vitro tissue models using immunohistochemistry and whole slide imaging and cytokine multiplex assays. Our results show that the currently discussed models of 3D "tongue" extension during

epidermal wound closure cannot be hold. Instead our results led us to a novel "shield extension" model. Our results clarify a long-stand question by showing for the first time a consistent experimental and theoretical model for epidermal wound closure in 3D. Moreover, it appears that epidermal wound closure is not simply an example of collective cell migration, but instead is a complex regulated 3D phenomenon. The results especially show the limits of 2D monolayer wound closure assays and ask at least partly for re-assessment of those results in an experimental 3D setting in the case of human epidermis.

e-mail: niels.grabe@bioquant.uni-heidelberg.de

Model of gap gene expression in *Drosophila* embryo based on telegraph equation

Igor A. Gula**, Alexander M. Samsonov
The Ioffe Physical-Technical Institute of the Russian Academy of Sciences,
St. Petersburg, 194021 Russia

By convention a mathematical model based on parabolic reaction-diffusion equation has been used for description of gene expression process in the *Drosophila* embryo at the early stages of its development. According to this model equation any perturbation of protein concentration function will propagate immediately to any point of a living system even with exponentially small values [1], [2]. This unpleasant feature breaks the causality principle formally and is called "the diffusion paradox". We propose a refined model, which is based on the telegraph equation and allows one to eliminate the diffusion paradox and describe the gene expression correctly. Application to the gap gene expression description is discussed.

REFERENCES:

1. D.D. Joseph, L. Preziosi. (1989) Heat waves. *Rev. Mod. Phys*., 61: 41–73.
2. S.L. Sobolev. (1997) Local non-equilibrium transport models. *Phys. Usp.*, 40: 1043–1053

ACKNOWLEDGEMENTS

The support by the RFBR grant no. 11-01-00573, the FP7 project HEALTH-F5-2010-260429 and the State contract no. 14.740.11.0166 is gratefully acknowledged.

e-mail: gula@mail.ioffe.ru
*Corresponding author

Method for derivation of connectionist gene circuit equations from a sequence-level theory of transcriptional control

***Vitaly V. Gursky*[1]*, *John Reinitz*[2]**
[1]Ioffe Physical Technical Institute of the Russian Academy of Sciences, St. Petersburg, Russia
[2]University of Chicago, Chicago, U.S.A.

The connectionist equations provide a general framework for modeling gene regulation at the mesoscopic level of details [1]. A major shortcoming of this approach is that its parameters quantifying gene regulatory interactions have no explicit link to the genomic DNA sequence. We propose a method to calculate this link and derive the connectionist equations by an explicit approximation from a theory of transcriptional control parameterized by DNA sequence together with a small number of phenomenological parameters describing protein-protein interactions [2]. We apply this method in the context of a previous application of the sequence level model to control of the *Drosophila even-skipped* gene by the proximal 1.7 kb of its control region in early embryos [3]. The results provide an insight into the general limits of applicability for the connectionist equations for gene regulation.

Acknowledgement: The study was supported by NIH Grant RR07801, CRDF GAP award RUB1–1578, RFBR grants 10–01–00627a, 11–01–00573a, and 11–04–01162a, the State contract N 14.740.11.0166 from the Russian Ministry of Education and Science, and FP7 grant SysPatho HEALTH-F5-2010-260429.

REFERENCES

1. E.Mjolsness et al. (1991) A connectionist model of development, *J. Theor. Biol.*, 152: 429-453.
2. J.Reinitz et al. (2003) Transcriptional control in Drosophila, *ComPlexUs*, 1: 54-64.
3. H. Janssens et al. (2006) Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even-skipped gene, *Nature Genetics*, 38: 1159-1165.

e-mail: gursky@math.ioffe.ru
*Corresponding author

Integration of protein-protein interactions and RNAI screens for reconstruction of signaling networks

S. Hashemikhabir, T. Can*
Department of Computer Engineering, Middle East Technical University, 06800, Ankara, Turkey

There are three major approaches for signaling network reconstruction: protein-protein interaction (PPI) based methods, RNAi based methods, and hybrid methods that use poth

PPI and RNAi data. PPI based methods usually focus on finding a cluster of proteins in a given network. Most of the methods in this category identify a series of genes known as Seed or Core genes. RNAi methods, use only RNAi screen data. Researchers develop and apply machine learning approaches to reconstruct signaling networks from perturbation data. Due to the limitations in learning and optimization methods, these approaches are not applicable for reconstructing large networks. InfluenceFlow by Singh et al. [1] is one of the few methods that use the both RNAi and PPI data. Ruth et al. [2] propose an approach for explaining the knockdown effects by the use of PPI data. Both methods use one of the source of information (PPI or RNAi) as true source of information, and try to reconcile the differences by filtering the other source. In this study, we propose a new model by extending our earlier work in this area [2] via integration of high confidence PPI with high scoring RNAi observations by treating both sources of information equally and try to reconstruct the signaling network by selectively augmenting the core network with less confident PPI interactions and low scoring RNAi observation. Our goal is to develop efficient heuristics to scale the network construction problems to large networks that may contain hundreds of components. Our other contribution is that we avoid binary interpretation of RNAi data and explain the observed fold change via the concept of "betweenness centrality" in graphs.

REFERENCES

1. Singh, R. (2011). Algorithms for the Analysis of Protein Interaction Networks. Ph.D. thesis, MIT. Chapter 6: Influence Flow: Integration of PPI and RNAi Data.

2. Ruths, D., Tseng, J.-T., Nakhleh, L., and Ram, P. T. (2007). De novo signaling pathway predictions based on protein-protein interaction, targeted therapy and protein microarray analysis. In RECOMB-SAT, pages 108–118, Berlin, Heidelberg. Springer-Verlag.

3. Hashemikhabir, S., Ayaz, E.S., Kavurucu, Y., Can, T., and Kahveci, T. (2012). Large scale signaling network topology reconstruction using protein-protein interactions and RNAi data. Submitted to IEEE-ACM Transactions on Computational Biology and Bioinformatics.

e-mail: tcan@metu.edu.tr
*Corresponding author

Automatic Segmentation of Liver from Computed Tomography Examinations

R. Hassanpour* and O. Gazi
Cankaya University, Ankara - Turkey

Computed Tomography (CT) examinations are widely used for evaluating the internal organs of the human body. A complete CT scan of the abdomen, encompassing

multiple organs and other tissues may contain hundreds of images. The radiologists' job of studying all of these images is complex, requires extensive training and is prone to error. It would be helpful to have a method to segment critical organs, such as the liver, to assist diagnosing abnormalities and measuring volumes. However, computer-based automatic segmentation of the liver is challenging because of the wide range of variations in the shapes of the liver and the similarity of attenuation of contiguous and nearby structures. In this paper, we propose a fully automatic liver segmentation method from computed tomography images. Our proposed method consists of two stages. In the first stage, texture features of the organs are extracted and an initial approximation of liver area is made by combining them with the statistical location information obtained from an ATLAS. In the second stage, the internal region of the initial estimated area is used to generate a set of seed points for a region-growing algorithm using zero level sets. We are using gradient information to create a speed image to control the growing rate of the level set after applying an edge-preserving smoothing filter. The algorithm proceeds in three dimensions after performing the segmentation in the first slice by including the segmented area of the previous slice in the initial area estimation process which improves the accuracy of our method. Histogram analysis is used to estimate the initial points and threshold value ranges automatically.

e-mail: reza@cankaya.edu.tr
*Corresponding author

Reconstruction of the associative genetic networks based on integration of automated text-mining methods and protein-ligand interactions prediction

***Ivanisenko T.V.*[1,2], *Demenkov P.S.*[1,2], *Ivanisenko V.A.*[*1,2]**
[1]Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia.
PBSoft LLC, Novosibirsk, Russia

Motivation and Aim: Genetic network is a group of coordinated functioning genes. The associative genetic network is a genetic network reconstructed *in silico* with the use of formalized rules for identification for identification of interactions between molecular-genetic objects from the texts of scientific publications and databases. Manual analysis of such data by experts has high degree of accuracy but is very time-consuming therefore it makes timely the task of development of interactive tools for analysis

of a full-text articles which can be used for reconstruction of the associative genetic networks. The aim of this work was a development of the integrated computer system for the reconstruction of associative genetic networks based on interactive analysis of scientific literature and prediction of protein ligand interactions.

Methods and Algorithms: Automated extraction of information about molecular-genetic, genetic-genetic, metabolite-genetic and other types of interactions in text was performed using the text-mining methods we developed. For text-mining, we also used previously developed by us thesaurus with the names of proteins, genes, metabolites, diseases, microRNAs, biological pathways, cells and organisms. A web-based interface allows text-mining module to work in a real-time mode with textual files (in a *.pdf or *.txt format) uploaded by experts and to demonstrate results of data extraction in user-friendly way as well. The integrated system for prediction of protein-ligand interactions, based on recognition of functional sites in a tertiary structure of proteins feature, allows to get new previously unknown interactions.

Results: An integrated computer system for reconstruction of associative genetic networks based on interactive analysis of scientific literature and prediction of protein-ligand interactions (ITMSys) was developed. The ITMSys software is equipped with tools for reconstruction, visualization of genetic networks, prediction of new interactions and web-based interface. The ITMSys ensures analysis of full-text articles connected with genetic networks.

Conclusion: Developed system can be used by experts for the acceleration of genetic networks reconstruction process, as well as in the other fields of science related to automated analysis of the scientific texts.

Acknowledgements: Work is supported in part by Russian Ministry of Education and Science № 07.514.11.4003 and 14.740.11.0001 .

e-mail: itv@bionet.nsc.ru
*Corresponding author

Associative network discovery system (ANDSYSTEM): automated literature mining tool for extracting relationships between diseases, pathways, proteins, genes, micrornas and metabolites

Ivanisenko V.A.*[*1,2], Demenkov P.S.[1,2], Ivanisenko T.V.[1,2]*
[1]Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia.
[2]PBSoft Llc, Novosibirsk, Russia.

Motivation and Aim: Work with scientific literature as well as factual databases is required for research in every area of knowledge. In order to enable formulation of problems and hypotheses scientists are studying the existing pool of scientific knowledge. The size of this pool is immense and expands exponentially. Nearly 40% of such data are biomedical in nature. For example, the Pubmed database contains over 20 million of scientific abstracts, and their number increases annually by 1 million per year. Reading of all of them, even if it will take only 3 minutes on each, would take more than 200 years. In this way development of tools for automated literature analysis (text-mining) becomes a timely task.

Methods and algorithms: The text-mining algorithms implemented in the ANDSystem are based on semantic patterns. The improved method for semantic analysis of biological texts employs a link grammar parser combined with semantic patterns was developed. Also we have developed original methods for construction of pathways and cells vocabularies. The ANDSystem is provided with methods for automated reconstruction of associative gene networks, which describe semantic relationships between molecular-genetic objects (proteins, genes, metabolites and others), biological processes, and diseases.

Results: The ANDSystem was developed for the purpose of scanning literature for extracting relationships between diseases, pathways, proteins, genes, microRNAs and metabolites. The ANDSystem incorporates utilities for automated extraction of knowledge from Pubmed published scientific texts and analysis of factographic databases. The ANDCell database contains information on molecular-genetic events retrieved from texts and databases. The ANDVisio is a user's interface to the ANDCell database stored on the remote server. It provides graphic visualization, editing and search features as well as possibilities to save an associative gene networks in different formats resulting from user's request. The ANDVisio is provided with various tools supporting

filtering by object types, relationships between objects and information sources; graph layout; search of the shortest pathway; cycles in graphs.

Conclusion: The ANDSystem can assist in the interpretation of complex multifactorial experimental data. In particular, the ANDSystem was used for the analysis of proteomic experimental data. For example, with the ANDSystem was reconstructed and analyzed networks of molecular and genetic interactions of proteins of Helicobacter pylori, differentially expressed in different strains isolated from patients with chronic gastritis and gastric tumors. On the example of the establishment of interactions between human proteins and proteins of hepatitis C virus was shown a high accuracy of the method of knowledge extraction from texts.

Avaibility: The ANDSystem is available by request to developers.

Acknowledgements: Work is supported in part by 7th Framework Programme (FP7) project "New Algorithms for Host Pathogen System Biology" SYSPATHO №260429.

e-mail: salix@bionet.nsc.ru
* Corresponding author

A stochastic model for suppression of subgenomic hepatitis C virus replication in Huh-7 cells

N.V. Ivanisenko[1,2], E.L. Mishchenko[1], I.R. Akberdin[1], P. S. Demenkov[1], V.A. Likhoshvai[1,2], M.G. Samsonova[3], D. Clausznitzer[4], L. Kaderali[4], N.A. Kolchanov[1,2], and V.A. Ivanisenko[1,5*]
[1] The Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia; [2] The Novosibirsk State University, Novosibirsk, Russia; [3] St.Petersburg State Polytechnical University, St.Petersburg, Russia; [4] Institute for Medical Informatics and Biometry, Medical Faculty Carl Gustav Carus, University of Technology Dresden, Dresden, Germany; [5] PB-soft Llc, Novosibirsk, Russia

Mathematical simulation of the molecular events underlying viral reproduction in the cell is a promising approach in the search of new drug targets and in gaining insights into the complexities of drug action on hepatitis C virus (HCV). Early we have proposed a mathematical model which was designed for examination of the effects of anti-HCV drugs. However this model described the kinetics of viral RNA suppression by various inhibitors during a short time, not over some days [1].

Here, we have developed for the first time a mathematical model for subgenomic HCV replicon replication in Huh-7 cells in the presence of the HCV NS3 protease inhibitors. Our stochastic model describes the experimental kinetics of the viral RNA suppression during an observation time of up to 15 days. The model takes into ac-

count the fact that mutant drug-resistant replicons preexist in a treatment-naïve replicon population [2], also the stochastic nature of the replication-degradation of viral RNA. When disregarding the stochastic of the replication-degradation of viral RNA, the model is still able to describe the experimental kinetics during the initial few days action of the inhibitors when cellular concentration of viral RNA is rather high (≥ 20 RNA molecules/cell).

In the model we have used a simplified scheme of subgenomic HCV replicon replication in Huh-7 cells in the presence of inhibitors, also a minimal number of parameters. The novelties were vesicles of two types, one producing mutant RNAs, the other RNA of wild-type into the cytoplasm. Computational analysis of the model demonstrates that the stochastic nature of the subgenomic HCV replicon replication and the drug-resistant mutant replicons in cells is necessary to explain the experimentally observed biphasic reduction of viral RNA in the presence of the inhibitors.

REFERENCES:

1. E.L. Mishchenko et al. (2007) Mathematical model for suppression of subgenomic hepatitis C virus RNA replication in cell culture, J.Bioinform. Comput. Biol., 5: 593-609.

2. M. Robinson et al. (2011) Preexisting drug-resistance mutations reveal unique barriers to resistance for distinct antivirals, Proc. Natl. Acad. Sci. USA., 108: 10290-10295.

e-mail: salix@bionet.nsc.ru
*Corresponding author

Improved differential evolution entirely parallel method

K.N. Kozlov*[1], A.M. Samsonov[2] and M.G. Samsonova[1]
[1]St.Petersburg State Polytechnical University, St.Petersburg 195251, Russia
[2]A.F. Ioffe Physico-technical Institute of the RAS, St.Petersburg, 194021, Russia

Currently, the design of efficient algorithms and systems to solve the inverse problem of mathematical modeling continues to be a challenge due to large volume and heterogeneity of biomedical data, as well as high computational complexity of biomedical applications. It is well established, that an efficient optimization method should not only be fast and scalable across modern high performance architectures, but also reliable and robust.

In this work we describe the improvements to the Differential Evolution Entirely Parallel (DEEP) method developed in the work [1]. The Differential Evolution, introduced in 1995 by Storn and Price, considers the population, that is divided into

branches, one per computational node [2]. The DEEP method takes into account the individual age, that is defined as the number of iterations the individual survived without changes, the improved method additionally: (I) allows several oldest individuals at (k + 1)th branch to be overwritten by the same number of best ones from kth branch and (II) implements a new selection rule for Differential Evolution that uses several different objective functions in offspring evaluation. The offspring replaces its parent if the value of the quality functional for the offspring set of parameters is less than that for the parental one. The additional objective functions are checked in the opposite case. The offspring replaces its parent if the value of some objective function is better and the randomly selected value is less than the predefined parameter for this function.

We compared the performance of ImDEEP with original method and the state of the art optimization techniques such as Evolutionary Strategy (ES). The numerical results shows clear that ImDEEP outperforms the predecessors and is at least 2 times faster than ES.

We are very thankful to V. Gursky for many valuable discussions. The support of the study by the State Contract № 14.740.11.0166, RFBR Grants 11-04-01162, 10-01-00627, 11-01-00573, EU-FP7 SYSPATHO №260429 is gratefully acknowledged.

REFERENCES:

1. Konstantin Kozlov, Alexander Samsonov (2011). DEEP – Differential Evolution Entirely Parallel Method for Gene Regulatory Networks, Journal of Supercomputing, Volume 57, pp.172-178.

2. Storn R., Price K., (1995), Differential Evolution – A Simple and Efficient Heuristic

e-mail: kozlov@spbcas.ru

*Corresponding author

Modeling of gap gene expression in Drosophila Kruppel mutants

Konstantin Kozlov, Svetlana Surkova, Ekaterina Myasnikova, Maria Samsonova*

Dept of Computational Biology/Center for Advanced Studies/St.Petersburg State Polytechnical University, St.Petersburg,Russia

The segmentation gene network in Drosophila embryo solves the fundamental problem of embryonic patterning: how to establish a periodic pattern of gene expression, which determines both the positions and the identities of body segments. The gap gene network constitutes the first zygotic regulatory tier in this process. Here we have applied the systems-level approach to investigate the regulatory effect of gap gene Kruppel (Kr) on segmentation gene expression. We acquired a large dataset on the

expression of gap genes in Kr null mutants and demonstrated that the expression levels of these genes are significantly reduced in the second half of cycle 14A. To explain this novel biological result we applied the gene circuit method which extracts regulatory information from spatial gene expression data. Previous attempts to use this formalism to correctly and quantitatively reproduce gap gene expression in mutants for a trunk gap gene failed, therefore here we constructed a revised model and showed that it correctly reproduces the expression patterns of gap genes in Kr null mutants. We found that the remarkable alteration of gap gene expression patterns in Kr mutants can be explained by the dynamic decrease of activating effect of Cad on a target gene and exclusion of Kr gene from the complex network of gap gene interactions, that makes it possible for other interactions, in particular, between hb and gt, to come into effect. The successful modeling of the quantitative aspects of gap gene expression in mutant for the trunk gap gene Kr is a significant achievement of this work. This result also clearly indicates that the oversimplified representation of transcriptional regulation in the previous models is one of the reasons for unsuccessful attempts of mutant simulations.

e-mail: samson@spbcas.ru
*Corresponding author

Detecting Non-Uniform Clusters in Large Scale Interaction Graphs

Nissan Levtov[1], Sandeep S Amberkar[2], Lars Kaderali[2] and Zeev Volkovich[1]*
[1]Department of Software Engineering, Ort Braude College, Karmiel, Israel
[2]University of Heidelber, Bioquant, Heidelberg, Germany

In this paper we propose a new graph clustering algorithm for the problem of revealing the inherent cluster structure in interaction graphs. We show that if the inherent clusters are highly non-uniform in their sizes, then even for very simple cases many current clustering techniques fail to detect the correct clusters. We deal with the cluster size variability by relying on the reasonable assumption that nodes that belong to the same cluster share common characteristics, including a resemblance in their node degrees. We can therefore reduce the search space by looking for small clusters only among the low degree nodes while larger clusters are searched for among the higher degree nodes. We suggest two different approaches for decomposing the search space into several degree dependant strips. The first approach is using the core-decomposi-

tion of the graph. The second approach is by fixing the strip size and partitioning the sorted array of degrees. Within each strip we cluster the nodes using a structural dissimilarity distance function. This dissimilarity function satisfies the triangle inequality which enables a natural embedding of the graph in an n-dimensional metric space. We cluster the embedded points with the farthest-point-first algorithm known to attain a 2- approximation ratio for the k-center problem with metric distances. Since the true number of clusters is not known in advance, we estimate the number of clusters, k, by using the «elbow criterion» on the target optimization function. The fixed approximation ratio of the clustering algorithm ensures a robust estimation.We evaluate the performance of our algorithm on a synthetic network and on a biological network of yeast protein interaction data and compare the results to several known algorithms. We have also developed an implementation of the algorithm as a plugin for Cytoscape.

e-mail: zeev@actcom.co.il
*Corresponding author

Role of the Mad2 dimerization interface in supporting the Spindle Assembly Checkpoint independently from kinetochores

Luca Mariani[1+], Elena Chiroli[1,+], Luigi Nezi[2], Heiko Muller[3], Simonetta Piatti[4], Andrea Musacchio[5,6] and Andrea Ciliberto[1*]

[+] These Authors contributed equally to the work.

[1] IFOM - The FIRC Institute of Molecular Oncology, Via Adamello 16, Milan, Italy;

[2] Luigi Nezi, Department of Genomic Medicine, MD Anderson Cancer Center, Houston, USA; [3]- Istituto Italiano di Tecnologia, Center for Genomic Science of IIT@SEMM c/o IFOM-IEO-CAMPUS, Via Adamello 16, 20139 Milan, Italy; [4] Centre de Recherche de Biochimie macromoléculaire, 1919 route de Mende, 34293 Montpellier Cedex 5, France; [5] MPI für molekulare Physiologie, Otto-Hahn-Straße 11, D-44227 Dortmund, Germany;

[6] Centre for medical biotechnology, Faculty of Biology, University of Duisburg-Essen, Universitätsstr. 45141 Essen, Germany.

The Spindle Assembly Checkpoint (SAC) arrests cells when kinetochores are unattached to spindle microtubules. The signaling pathway is initiated at the kinetochores by one SAC component, Mad2, which catalyzes the initial steps of the cascade via the conformational dimerization of its open and closed conformers. Away from kinetochores the dimerization surface of Mad2 has been proposed, based on data *in vitro*, to either interact with SAC activators or inactivators, and thus to contribute to SAC activation or silencing. Here, we analyze its role *in vivo*.

To analyze the putative pathway downstream of the kinetochores, we used two complementary approaches: we activated the SAC ectopically and independently from kinetochores, and we separated genetically the kinetochore-dependent and independent pools of Mad2. We found that the dimerization surface is required also downstream of kinetochores to mount a checkpoint response.

Our results show that away from kinetochores the dimerization surface is required for stabilizing the end-product of the pathway, the Mitotic Checkpoint Complex. Surprisingly, downstream of kinetochores the surface does not mediate Mad2 dimerization. Instead, our results are consistent with a role of Mad3 as the main interactor of Mad2 via the dimerization surface.

e-mail: andrea.ciliberto@ifom.eu
*Corresponding author

Identifiability analysis and predictive power of the gene circuit model

Ekaterina Myasnikova
St.Petersburg State Polytechnical University, St.Petersburg, Russia

For the comprehensive study of modeling results it is necessary to know how reliable the parameter estimates are, that constitutes the problem of identifiability analysis. The detection of parameter non-identifiabilities is closely connected with the study of predictive properties of the model.

A typical example of predictive model is a gene circuit model [1] that dynamically reconstitutes the set of interactions within the genetic network. In an ideal world the model is fitted to wild type (WT) data and the solution obtained by zeroing of parameters related to a target gene should show a good fit to data null mutant for this gene. At the same time the correct prediction of the model behavior at fixed values of some parameters is only possible if these parameters are identifiable.

In this study we introduce a criterion of predictive power of the model that is based on measures of model sensitivity to parameters and apply it to the gene circuit model presented in [2] that describes the dynamics of segmentation gene expression in *Drosophila melanogaster*. Two types of measures are considered: the first one reveals the biologically substantiated low sensitivity of the model to changes of parameters

that are responsible for correct reconstruction of expression patterns in mutants, while the second one takes into account their correlation with the other parameters.

It is shown that the model solution obtained by fitting to gene expression data in WT and Kr^- mutants demonstrates much higher predictive power with respect to gene expression patterns in null mutants for *Kr* than those only fitted to WT data. The method makes it possible to predict the possibility to correctly reproduce the expression of other genes in mutants. Besides, the requirement of high values of the measures introduced in this work may be used as a necessary condition in the search of model solutions to provide the high predictive power.

REFERENCES

1. J. Jaeger, S.Surkova, M.Blagov, H. Janssens, D.Kosman, K. Kozlov, Manu, E.Myasnikova, C. E. Vanario-Alonso, M.Samsonova, D.Sharp, and J.Reinitz, Dynamic control of positional information in the early Drosophila embryo. *Nature*, 430:368-371, 2004.

2. K.Kozlov, S.Surkova, E.Myasnikova, J.Reinitz, M.Samsonova (2012) Modeling of gap gene expression in *Drosophila* Kruppel mutants, PLoS Computational Biology, in Press.

e-mail: myasnikova@spbcas.ru

Multiscale modelling of trail therapy applied to multicellular spheroids

Holger Perfahl*[*1], Agathe Bronikowski[2], Malgorzata Doszczak[2], Peter Scheurich[2], Matthias Reuss[3]
[1]Institute of Systems Theory and Automatic Control, University of Stuttgart, Stuttgart, Germany
[2] Institute of Cell Biology and Immunology, University of Stuttgart, Stuttgart, Germany
[3] Center Systems Biology, University of Stuttgart, Stuttgart, Germany

We present an off-lattice multiscale model of tumour growth that takes into account mechanical cell-cell interactions, the distribution of diffusible substances and a subcellular apoptosis pathway. The model is applied to describe the growth of a tumour spheroid that is cultivated in a nutrient enriched hanging drop. We used in vitro experiments to estimate the tumour's growth rate as well as the location and shape of the necrotic core. Cell death is initiated if the local nutrient concentration is below a given death-threshold. This enables us to reproduce the growth behaviour of the multicellular spheroid in silico. The tumour spheroid growth model is then applied to study the effects of a TRAIL therapy under in vitro conditions. We include a TRAIL-induced apoptosis pathway model (Hasenauer et al. 2009) in each cell. The apoptosis model accounts for the observed heterogeneity of cellular responses to TRAIL therapy by

sampling one parameter from a bimodal parameter distribution. The dynamics of the apoptosis pathway is influenced by the extracellular TRAIL concentration that is described by a reaction diffusion equation. TRAIL induced tumour cell death is initiated if the intracellular concentration of the active caspase 3 exceeds a death-threshold. In our simulations, we aim to answer the question of how the interplay of the cell heterogeneity together with the spatial heterogeneity influences the drug's efficacy.

REFERENCES

1. J. Hasenauer, S. Waldherr, N. Radde, M. Doszczak, P. Scheurich, F. Allgöwer. A maximum likelihood estimator for parameter distributions in heterogeneous cell populations. Procedia Computer Science 1 (2010) 1649-1657

e-mail: holger.perfahl@ist.uni-stuttgart.de
*Corresponding author

Heuristic optimization method for inverse problem solving

Andrey S. Pisarev
St.Petersburg State Polytechnical University, St.Petersburg 195251, Russia

Important advances in modeling of pattern formation phenomenon were achieved over the last years. Further improvement of quality of obtained solutions is an important task. Here we address this problem by applying an additional generalized information on temporal dynamics of gene expression in the whole embryo

A classical approach to find parameter values of pattern formation models in Drosophila is by fitting a model solution to experimental gene expression patterns, represented by reference data for each segmentation gene, each time point and each spatial coordinate of the A-P axis. Such a data is called as one-dimensional (1D) integrated data. To obtain the generalized information on temporal gene expression dynamics during Drosophila early development we calculated the numerical values of integrals, defined in the region [a, b] of the A-P axis of one-dimensional integrated data [2] for maternal, gap and pair-rule segmentation genes at time classes 2-8 of cleavage cycle 14A.

We applied a method of decision trees [3] for development of an additional criterion for heuristic analysis of simulations. To do this the classical approach was reformulated as multiobjective optimization problem in the form:

$$f(X^*) = extr_{X \in C}\{f(X) + \sum_i \alpha_i S_i(X)\} \quad (1)$$

$$S_i(X) = \sigma_\phi(g_i(X) \quad (2)$$

, i=1,..,m,

where $f(X\ )$ - root mean square criterion, which describes the deviation of model solutions from spatial-temporal gene expression experimental data,

$\sigma_\phi(g_i(X)$ - penalty function and $g_i(X)$ - heuristic criterion.

The application of the method to simulatation of pair-rule gene expression dynamics. allowed us to improve the value of functional by factor of two and correctly predict the expression pattern of hairy (h) gene in eve mutant embryos.

The support of the study by the State Contract № 14.740.11.0166, RFBR Grant 11-04-01162 is gratefully acknowledged.

REFERENCES:

1. Jaeger J., Surkova S., Blagov M., Janssens H., Kosman D., Kozlov K., Manu, Myasnikova E., Carlos E. Vanario-Alonso, Samsonova M., Sharp D., Reinitz J. (2004a). Dynamic control of positional information in the early Drosophila embryo. Nature, 430: 368-371. PMID: 15254541.

2. Pisarev A., Poustelnikova E., Samsonova M., Reinitz J. (2009). FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. Nucl. Acids Res., 37: D560 - D566.

3. Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

e-mail: pisarev@spbcas.ru

Different models of diffusion, related to the equation of the Jeffreys type

Sergey A.Rukolaine*, Alexander M. Samsonov

The Ioffe Physical Technical Institute of the Russian Academy of Sciences, St.Petersburg, Russia

The classic conventional diffusion equation, based on Fick's law, is widely used for approximate macroscopic description of diffusion (transport of matter). The diffusion equation gives an appropriate and accurate model for di?usion phenomena in weakly inhomogeneous media and/or processes, when relaxation time is short compared to the characteristic time scale. Otherwise the description of diffusion phenomena by the diffusion equation may fail.

Many biological media, e. g., cellular cytoplasm, are strongly inhomogeneous and/or highly viscoelastic (non-Newtonian), therefore, diffusion in them is not Fickian and its description by the diffusion equation fails [1, 2].

The equation of the Jeffreys type was proposed for description of transport in rheological(non-Newtonian) media [3]. This equation generalizes the diffusion equation, it takes into account the nonlocality of transport both in time and space (time delay and inertia of molecules). In this paper we consider and investigate different models of transport, related to the equation of the Jeffreys type. We show that these models are qualitatively different, and describe different modes of transport (diffusion). We show also that some of the models have unusual properties.

REFERENCES

1. A.D. Lander, et al. (2002) Do Morphogen Gradients Arise by Diffusion? *Dev. Cell*, 2:785–796.

2. R.DeLotto, et al. (2007) Nucleocytoplasmic shuttling mediates the dynamic mainte-nance of nuclear Dorsal levels during Drosophila embryogenesis. *Development*, 134:4233–4241.

3. D.D. Joseph, L.Preziosi. (1989) Heat waves. *Rev. Mod. Phys*., 61: 41–73.

4. M.Coppey, et al. (2007) Modeling the bicoid gradient: Diffusion and reversible nucleartrapping of a stable protein. *Dev. Biol*., 312: 623–630.

e-mail: rukol@ammp.ioffe.ru

*Corresponding author

Mathematical modeling of metformin influence on oncogenesis in SHR mice

Tatiana E. Sannikova [1*], ***Vladimir N. Anisimov***[2]

[1]Institute of Numerical Mathematics RAS, Moscow, Russia

[2]Department of Carcinogenesis and Oncogerontology, NN Petrov Research Institute of Oncology, St.Petersburg, Russia

There is now abundant evidence pointing to the relation between the rate of tumor development and the state of the energy budget. Caloric excess and obesity favor the tumor growth and in contrast the caloric restriction slows it down [1,2]. The beguanide metformin postpones spontaneous carcinogenesis in mice by suppressing the transformative and hyperproliferative processes [3,4]. The purpose of the current work is to investigate the influence of metformin on homeostasis maintenance. We suggest the mathematical model describing the relationship between the fidelity of DNA replication and the rate of tissue regeneration. The system of differential equations depicts the processes of energy reserving and consumption on homeostasis and tissue regeneration. The fidelity of DNA replication is supposed to be determined by the rate of tissue regeneration. Improving the accuracy of DNA replication is accompanied by increasing in energy consumption on regeneration and maintenance of the reserved power in

tissues. It leads to decrease in fitness. By means of the model the observations from [4] are quantitatively simulated. We demonstrate that changes in the rate of carcinogenesis following the energy budget shrinking may be a result of organism adaptation.

REFERENCES

1. I.M. Lee et al. (2001) Epidemiologic data on the relationships of caloric intake, energy balance, and weight gain over the life span with longevity and morbidity, J Gerontol A Biol Sci Med Sci, 56: 7-19.
2. E. E. Calle et al. (2003) Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med 348:1625-38.
3. V.N. Anisimov (2010) Metformin for aging and cancer prevention, AGING , 2 (11):760-774
4. V.N. Anisimov et al. (20011) If started early in life, metformin treatment increases life span and postpones tumors in female SHR mice, AGING, 3(2) :148-57.

e-mail: tatiana@inm.ras.ru
Corresponding author

FDR is the d-risk

Simushkin D. S., Volodin I. N.
Kazan Federal University, Kazan, Russian Federation

In the last decade and a half statistical methods based on the FDR are actively developing (see, for example, [1], where an extensive bibliography is presented). The use of FDR in practice is possible only within the Bayesian paradigm. A similar approach to the calculation of quantity measures of statistical procedures called d-posterior, was proposed in the 80th years of the XX century [2]. Under this approach, the value of FDR is equal to d-risk of a first kind – the probability of the validity of the hypothesis, provided that the hypothesis was rejected. Various methods were developed for constructing optimal statistical procedures.

In this report, example discussed in the monograph of Efron (chapter 2, a microarray example) is analyzed from the perspective of the d-posterior approach. To identify the genes responsible for cancer, applied statistical test for which d-risk 1st kind (FDR) less then a level α and minimizes the d-risk of the 2nd kind – the rate of a genes responsible for cancer, if they has been recognized as not influencing disease (FNR in the terminology of Efron). The model of Efron (in terms of setting a priori distribution) has been expanded to adequately describe the data sample – the significance test of chi-square has a p-value equal 0.8. We show that the prior distribution for the mean test statistic can be described by a hierarchical Bayesian model. This distribution is concentrated at 0 with probability (1-p) and coincides with the normal (0, t^2)

distribution with probability p. The parameter estimates p, t are found by the method of minimum chi-square.

Note that, in contrast to Efron test, our test can be constructed for any level α. Although the critical constants of both tests for the level α =0.05 is almost identical, the values of the 2nd kind of the d-risks are substantially different.

REFERENCES

1. B. Efron. (2010) *Large-Scale Inference*, Cambridge.

2. S.V. Simushkin. (1983) The empirical d-posterior approach to the problem of guaranteedness of statistical inference. *Soviet Mathematics (Izvestiya VUZ. Matematika)*, 1983, 27:11, 47–67.

email: SimDim555@gmail.com, Igor.Volodin@ksu.ru

MODELING of aggregation of TAU proteins: system pharmacology approach.

Alexander A. Stepanov**, Oleg O. Demin, Oleg V. Demin
Institute for Systems Biology SPb, Moscow, Russia

Increasing neurodegenerative disease prevalence threatens general welfare and prompts scientists to discover new drugs and treatments for the pathologies. Tauopathies are neurodegenerative disorders encompassing more than ten pathologies characterized by deposition of the abnormal tau protein. Tau aggregation is one of the hallmarks of tauopathies including Alzheimer's disease.

Tau proteins are neurospecific microtubule-associated proteins that modulate the stability of axonal microtubules. Modeling of tau behavior in normal conditions and during tauopathy is a challenge that could be addressed using systems biology and system pharmacology modeling approaches.

We construct a mathematical model implemented *in silico* in the form of ordinary differential equations to describe tau aggregation. A current version of the model is common for all tauopathies. We incorporate some mutations of key players leading to certain tauopathies and simulate their behavior. This model enables identification of critical intervention points to impact of potential drugs, thereby, arresting or reversing the tauopathy.

We investigate sensitivity of the model parameters to evaluate the significance of the model entities for tau aggregation.

Our analysis highlights the multifactorial nature of the disease and provides insight into pathological mechanisms of tau-mediated aggregation diseases including Alzheimer's.

E-mail: stepanov@insysbio.ru
*Corresponding author

Quantitative characterization of the segmentation gene expression in Drosophila gap mutants

***Svetlana Surkova*[1]*, *Elena Golubkova*[2], *Manu*[3], *Lena Panok*[3],**
Lyudmila Mamon[2], John Reinitz[3] and Maria Samsonova[1]
[1]St. Petersburg State Polytechnical University, Russia
[2]St. Petersburg State University, Russia
[2]The University of Chicago, Chicago, USA

During segment determination in *Drosophila*, gap genes are the first to establish discrete territories of gene expression based on the interpretation of the regulatory input from long-range protein gradients. Here we study the response of segmentation system to mutations in two gap genes, *Kr* and *kni* and analyze both single mutants and double *Kr;kni* mutants. Our data is quantitative with cellular resolution in space. We first characterize segmentation gene expression in null mutants and *Kr*+/*Kr*- heterozygotes in terms of average or typical behavior. We analyze the dynamic changes in position and amplitude of expression domains in relation to wild type. Expression within gap gene domains in mutants is lower than in wild type but also has a tendency to decrease over time. As in wild type embryos, the posterior expression domains in mutants dynamically shift their positions to anterior during cycle 14A.

The massive central issue in studying the effect of mutations in segmentation genes is the estimation of individual variability in gene expression in individual mutant embryos. Mutations often produce variable phenotypes, and it is well known that the cuticular phenotype of *Kr* mutants is variable. We sought to understand the molecular basis of this effect. We detect that prior to gastrulation expression patterns of pair-rule genes are still variable in amplitude and the location of these variable regions suggests that this is the cause of variability in cuticular phenotypes. Moreover, our analysis demonstrates that in single *Kr* and *kni* mutants, unlike in wild type, the variability in positioning of the posterior Hb domain and Eve stripe 7 is not decreased or filtered with time. The posterior Gt domain in *Kr* mutants is highly variable at early times,

but this variability decreases when it shifts in the anterior direction to the position of neighboring Kni domain. In double *Kr;kni* mutants the positional variability does not decrease over time both in anterior and posterior body regions.

e-mail: surkova@spbcas.ru
* Corresponding author

Novel method of CNV analysis in FcγR locus and its application to immune-related diseases

***A.Zhernakova*[1,2,*] , L.Franke[1], A.Ioan[2], G.Trynka[1], K. Kok[1], K.Fransen[3], R.Weersma[3], S. Rantapää-Dahlqvist[4], P. Gregersen[5], C.Wijmenga[1], T.Huizinga[2], R.Toes[2].**
[1] Genetics Department, University Medical Center Groningen, Groningen, The Netherlands,
[2] Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands,
[3] Department of Gastroenterology, University Medical Center Groningen, Groningen, The Netherlands
[4] Department of Public Health and Clinical Medicine/Rheumatology, Umea University, Umea, Sweeden
[5] Feinstein Institute for Medical Research, Manhasset, New York, USA

Genetic variants near the FC-gamma receptor (FcγR) locus are associated with several immune-related diseases. However, most FcγR genes are located in complex regions of segmental duplications (SD) and they are therefore not well covered by the genotyping platforms. To be able to identify copy-number variants (CNVs) in this locus, we developed a method to analyse CNVs using principal component analysis of the raw intensity values of single nucleotide polymorphisms (SNPs) genotyped on the Immunochip platform. This platform includes 1,159 SNPs in the SD block of FcγR genes; of these only 140 (12%) passed our quality control for SNP analysis, however intensity values of all SNPs are informative for the CNVs estimation. We identified several CNV loci in the FcγR block. We confirmed our results via an independent method – arrayCGH genotyping – and observed a perfect correlation in CNV estimation between both methods. Next, we applied our method to case-control cohorts of rheumatoid arthritis (RA), celiac disease and inflammatory bowel disease (in total, 15231 individuals). We found no associations between CNVs with these diseases ($p > 0.05$), however individuals with RA more often had a rare complete deletion of *FCGR3B*. Additionally, by performing functional studies we observed a correlation between the number of *FCGR3A* gene copies and FCGR3 (CD16) expression on T-cells.

Conclusion: We have developed a method to accurately estimate CNVs based on SNP intensity data that can be extended to other phenotypes and to other SD loci in

the human genome. We have also established the functional effect of the number of copies of FCGR3A gene on the CD16 expression on T-cells.

email: a.zhernakova@umcg.nl

The power of meta-analysis of RNA-seq datasets for eQTL identification

D. V. Zhernakova*[*1] H. J. Westra[1], R. Jansen[2], E. de Geus[3], D. Boomsma[3], B. Penninx[3], C. Wijmenga[1], G.B. van Ommen[4], P. A. C. 't Hoen[4,5], L. Franke[1,5]

[1] Department of Genetics, University of Groningen, University Medical Centre Groningen, Netherlands

[2] Department of Psychiatry, Netherlands Study of Depression and Anxiety, VU University Medical Center, Amsterdam, Netherlands.

[3] Department of Biological psychology, Netherlands Twin Registry, VU University, Amsterdam, Netherlands

[4] Department of Human Genetics, Leiden University Medical Center, Leiden. Netherlands

[5] Contributed equally

Many genetic variants affect gene expression levels, although the exact mechanism through which this works is still mostly unclear. Previously, numerous expression quantitative trait locus (eQTL) mapping studies have been performed using microarray data. Recently, through next generation sequencing gene expression level quantitation has become possible (RNA-seq) and it has shown to be a very powerful approach of quantifying the transcriptome.

However, various RNA-seq strategies have been proposed, but it is still unclear what the best strategy for eQTL mapping is and how to combine eQTL data that has been generated by different technologies.

Here, we used three different types of RNA-seq data: paired-end RNA-seq (56 samples), single-end RNA-seq (64 samples) and deepSAGE data (94 samples). eQTL mapping on each dataset yielded 1287 unique genes for single-end RNA-seq, 601 unique genes for paired-end RNA-seq and 1188 unique genes for deepSAGE data. We show that a meta-analysis on different types of data can be performed to increase statistical power, permitting us to identify significant associations to 3241 unique genes. We compared the eQTLs that had been identified using RNA-seq and array based data and observed a concordance of 95% in allelic directions, indicating highly consistent results.

Our study indicates that different types of RNA-seq datasets can be well combined and that meta-analysis of RNA-seq is a logical step forward to gain better insight into the genetic regulation of gene expression variation.

e-mail: dashazhernakova@gmail.com
*Corresponding author

Use of mechanism-based mathematical modeling for biological target selection and antibody drug specifications in pharma research

***K. Zhudenkov*[1], *J. Yu*[2]**
[1] Modeling & Simulation, Novartis Pharma LLC, Moscow, Russia; [2] Novartis Institutes for Biomedical Research, Cambridge, Massachusetts, USA

The process of discovering and developing a new medicine does not start with a drug candidate, but with the selection of an appropriate human biological target for which there is some evidence of involvement in a significant disease pathophysiology. Based on the context of the disease state, a decision must be made as to the treatment modality of the desired therapeutic. One then needs to decide how to engage and modulate the biological target of interest. Should one go for a new, small molecular weight chemical entity, or a biotherapeutic such as a monoclonal antibody? Monoclonal antibodies, whether they are generated from animals or from in vitro systems, typically have better safety profiles and a higher probability of success to reach the market, as compared to more traditional low molecular weight pharmaceuticals [1]. Additionally, much historical data exists about the behavior of monoclonal antibodies in both the presence and absence of various targets. However, even with these advantages, several considerations must be still made. For example, what is the impact of target location, expression level and turnover? What binding affinity should be specified to allow for binding to the desired target and modulation of disease? Does one need several cycles of (expensive) antibody affinity maturation to achieve a picomolar dissociation constant (KD), or would nanomolar be sufficient?

In this work, an application of mechanism-based modeling is presented with its utility from the early Pharma Discovery stage of biological target selection and compound design specifications against such a target [2]. It is a monoclonal antibody against a cytomegalovirus glycoprotein complex example, which involves an antibody binding

model and a viral load model. The model was used as part of a feasibility analysis prior to antibody generation, setting the specifications for the affinity needed to achieve a desired level of clinical efficacy. From its application, we demonstrate that mechanism-based PK/PD binding models are useful for predicting human response to biologics compounds. In particular, such models have the ability to integrate preclinical and clinical, *in vitro* and *in vivo* information and facilitate rational decision making during various stages of drug discovery and translational research.

REFERENCES:

1. Hansel T. *et al*. Nat Rev Drug Discovery. 2010;9:325–38.
2. Jing Yu et al. The AAPS Journal (2011) DOI: 10.1208/s12248-011-9256-y

email: kirill.zhudenkov@novartis.com
* Corresponding author

INDEX

Научное издание

Труды международного симпозиума «SysPatho -Системная биология и медицина»,

на английском языке

Proceedings of the SysPatho workshop «Systems Biology and Medicine»