

# **Workshop on Computational Models in Biology and Medicine 2018**

Joint workshop of the GMDS/IBS-DR working groups  
"Statistical Methods in Bioinformatics"  
and "Mathematical Models in Medicine and Biology"

**March 8th-9th, 2018**

**University of Regensburg, Germany**

## Workshop outline

This workshop intends to bring together researchers from different research areas such as bioinformatics, biostatistics and systems biology, who are interested in modelling and analysis of biological systems or in the development of statistical methods with applications in biology and medicine.

## Keynote speakers

- Carsten Marr, Institute of Computational Biology, München
- Nico Pfeifer, Methods in Medical Informatics, Tübingen

## Workshop venue

The workshop will be hosted in the lecture hall H52 at the WNGB building of the University of Regensburg. The lecture hall is easily accessible by public transport. For details, please check the [University's web page](#).

## Organization

The workshop is jointly organized by the GMDS/IBS working groups "Statistical Methods in Bioinformatics" (Klaus Jung, University of Veterinary Medicine Hannover; Holger Fröhlich, University of Bonn) and "Mathematical Models in Medicine and Biology" (Markus Scholz, University of Leipzig; Ingmar Glauche, University of Dresden), as well as Rainer Spang (University of Regensburg) who is also the local organizer.

## Contact and local organization

Prof. Dr. Klaus Jung  
Institute for Animal Breeding and Genetics  
University of Veterinary Medicine Hannover  
Bünteweg 17p, 30559 Hannover, Germany  
Tel.: +49 511 953-8878  
E-mail: [klaus.jung@tiho-hannover.de](mailto:klaus.jung@tiho-hannover.de)

Prof. Dr. Rainer Spang  
Institute for Functional Genomics  
University of Regensburg  
Am BioPark 9, 93053 Regensburg, Germany  
Tel.: +49 941 943 5053  
E-mail: [rainer.spang@klinik.uni-regensburg.de](mailto:rainer.spang@klinik.uni-regensburg.de)

## Support

The workshop is funded by the "Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS)" and the "Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR)".

## Program

*Thursday, March 8, 2018*

### **12:45-13:00 Workshop Opening**

#### **Session 1 (Chair: Klaus Jung) – Signaling and disease**

13:00-13:45 Keynote 1 Nico Pfeifer

*Analyzing the interplay between HIV-1 and the immune system*

14:45-14:10 Marco Blickensdorf

*Comparative agent-based modeling of A. fumigatus infections in mice and men*

14:10-14:35 Ananya Rastogi

*Modelling of Cooperativity Mechanisms in T cell Killing*

14:35-15:00 Anastasios Siokis

*An agent-based model for the F-actin driven localization of TCR, LFA-1 and CD28 in the immunological synapse*

### **15:00-16:00 Coffee break & poster session**

#### **Session 2 (Chair: Ingmar Glauche) – Modeling hematopoiesis and cancer**

16:00-16:45 Keynote 1 Carsten Marr

*Quantifying cellular dynamics of stem cell decisions*

16:45-17:10 Yuri Kheifetz

*Integral individualized model of hematopoiesis explains dynamics of thrombocytes, granulocytes and lymphocytes under different chemotherapeutic and supportive treatments scenarios*

17:10-17:35 Xin Lai

*Modelling repression of E2F1 by cooperative microRNA pairs in the context of anticancer chemotherapy resistance*

17:35-18:00 Jens Przybilla

*Computational modeling of DNA de-methylation treatment of Acute Myeloid Leukemia*

### **19:30 Social event**

**Friday, March 9, 2018**

**Session 3 (Chair: Holger Fröhlich) – Machine learning applications**

08:30-08:55 Thomas Gerlach

*Predicting comorbidities of epilepsy patients using big data from Electronic Health Records combined with biomedical knowledge*

08:55-09:20 Klaus Jung

*Assessing false positive findings in a viral detection pipeline using high-throughput sequencing data*

09:20-09:45 Franziska Görtler

*Loss-Function Learning for Digital Tissue Deconvolution*

09:45-10:10 Moritz Hess

*A deep learning approach for uncovering lung cancer immunome patterns*

**10:10-11:00 Coffee break**

**Session 4 (Chair: Markus Scholz) – Open topics**

11:00-11:25 Nicole Radde

*From heterogeneous data of biological systems to quantitative predictive models*

11:25-11:50 Thomas Zerjatke

*Quantitative Cell Cycle Analysis Based on an Endogenous All-in-One Reporter for Cell Tracking and Classification*

11:50-12:15 Volker Schmid

*Quantitative analysis of the 3D nuclear landscape*

12:15-12:40 Florian Auer

*NDExR and Cytoscape: Interactive and automated visualization of biological networks using R*

**12:40-13:00 Workshop closing**

## Abstract of Keynote Talks

### Quantifying cellular dynamics of stem cell decisions

Carsten Marr

Helmholtz-Zentrum München

carsten.marr@helmholtz-muenchen.de

Time-lapse microscopy is a powerful method to continuously monitor single cells. Combined with appropriate markers, it allows for the time-resolved observation of protein expression. Such information is necessary to characterize state transitions, differentiation dynamics and gene expression models. However, to infer differentiation events or regulatory interactions robustly, bioimage computing and tailored data analysis is required.

I will first present bioimage informatics tools for the processing of time-lapse data. Uneven illumination and variable background levels in microscopy images affect subsequent analysis but can be corrected with a computational method based on low rank and sparse decomposition. Single-cell information can be retrieved from images using tracking, segmentation and quantification. These tools generate lineage trees that contain genealogical information but also the dynamics of morphological changes, cell speed, and expression of fluorescent proteins.

In the second part, I will present statistical models to infer cellular properties from single-cell data. In lineage trees of differentiating blood stem cells, stem cell decisions happen before an observable state change, inducing correlations in sister cells. Using these correlations and a stochastic model for a differentiation process, we find differentiation events to happen much earlier than anticipated and identify transcription factor expression that is inconsistent with the current toggle switch paradigm. To predict differentiation prospectively, we use a deep neural network trained on image patches from brightfield microscopy and cellular movement. Based on these features, we can detect lineage choice in blood stem cells up to three generations before conventional molecular lineage markers are observable.

## Analyzing the interplay between HIV-1 and the immune system

Nico Pfeifer

University of Tübingen

pfeifer@informatik.uni-tuebingen.de

With about 2 million new HIV infections per year, the human immunodeficiency virus is still one of the major health hazards world-wide. Even though antiretroviral therapy has improved over the last two decades, a universal vaccine against HIV would still have very large impact. Due to the high mutation rate of the virus there exists a very large variability between different strains world-wide. Therefore, standard antibodies isolated from HIV-infected patients usually neutralize only a small part of the different global variants. Recently, Walker *et al.* were able to isolate very potent and broadly neutralizing antibodies (bNAbs) from patients with exceptional neutralization responses [1]. Since then, several very potent broadly neutralizing antibodies have been isolated [2] showing that a universal vaccine that stimulates creation of such antibodies could become a reality in the future.

Additional to antibody responses human leukocyte antigen class I (HLA)-restricted CD8<sup>+</sup> T lymphocyte (CTL) responses are crucial to HIV-1 control. Although HIV can evade these responses, the longer-term impact of viral escape mutants remains unclear, as these variants can also reduce intrinsic viral fitness. To address this, we developed a metric to determine the degree of HIV adaptation to an HLA profile. This metric is based on a graphical model learned using our newly developed method for learning multistate phylogenies.

We demonstrated that transmission of viruses that are pre-adapted to the HLA molecules expressed in the recipient is associated with impaired immunogenicity, elevated viral load and accelerated CD4<sup>+</sup> T cell decline [5]. Furthermore, the extent of pre-adaptation among circulating viruses explains much of the variation in outcomes attributed to the expression of certain HLA alleles. Thus, viral pre-adaptation exploits 'holes' in the immune response. Accounting for these holes may be key for vaccine strategies seeking to elicit functional responses from viral variants, and to HIV cure strategies that require broad CTL responses to achieve successful eradication of HIV reservoirs.

In the bNAbs field we developed methods accounting for many sources of confounding. Those were applied in recent bNAbs trials, leading to high-impact results [3,4,6]. Furthermore, we showed how to build well-performing models for bNAbs resistance prediction [7] that are interpretable and can be employed to perform very interesting large scale bNAbs resistance studies.

1. L. M. Walker, S. K. Phogat, P.-Y. Chan-Hui, D. Wagner, P. Phung, J. L. Goss, T. Wrin, M. D. Simek, S. Fling, J. L. Mitcham, J. K. Lehrman, F. H. Priddy, O. A. Olsen, S. M. Frey, P. W. Hammond, P. G. P. Investigators,

- S. Kaminsky, T. Zamb, M. Moyle, W. C. Koff, P. Pognard, and D. R. Burton. Broad and Potent Neutralizing Antibodies from an African Donor Reveal a New HIV-1 Vaccine Target. *Science*, 326(5950):285–289, 2009.
2. F. Klein, H. Mouquet, P. Dosenovic, J. F. Scheid, L. Scharf, and M. C. Nussenzweig. Antibodies in HIV-1 vaccine development and therapy. *Science (New York, N.Y.)*, 341(6151):1199–204, 2013.
  3. M. Caskey, T. Schoofs, H. Gruell, A. Settler, T. Karagounis, E. F. Kreider, B. Murrell, N. Pfeifer, L. Nogueira, T. Y. Oliveira, G. H. Learn, Y. Z. Cohen, C. Lehmann, D. Gillor, I. Shimeliovich, C. Unson-O’Brien, D. Weiland, A. Robles, T. Kümmerle, C. Wyen, R. Levin, M. Witmer-Pack, K. Eren, C. Ignacio, S. Kiss, A. P. West, Jr, H. Mouquet, B. S. Zingman, R. M. Gulick, T. Keler, P. J. Bjorkman, M. S. Seaman, B. H. Hahn, G. Fätkenheuer, S. J. Schlesinger, M. C. Nussenzweig, F. Klein Antibody 10-1074 suppresses viremia in HIV-1-infected individuals *Nature Medicine* 2017, 23, pp 185–191.
  4. J. F. Scheid\*, Joshua A. Horwitz\*, Y. Bar-On, E. F. Kreider, C.-L. Lu, J. C. C. Lorenzi, A. Feldmann, M. Braunschweig, L. Nogueira, T. Oliveira, I. Shimeliovich, R. Patel, L. Burke, Y. Z. Cohen, S. Hadrihan, A. Settler, M. Witmer-Pack, A. P. West, Jr., B. Juelg, T. Keler, T. Hawthorne, B. Zingman, R. M. Gulick, N. Pfeifer, G. H. Learn, M. S. Seaman, P. J. Bjorkman, F. Klein, S. J. Schlesinger, B. D. Walker, B. H. Hahn, M. C. Nussenzweig, and M. Caskey HIV-1 antibody 3BNC117 suppresses viral rebound in humans during treatment interruption *Nature* 2016, 535, pp 556–560: equal contribution
  5. J. M. Carlson+, V. Y. Du\*, N. Pfeifer\*, A. Bansal, V. Y.F. Tan, K. Power, C. J. Brumme, A. Kreimer, C. E. DeZiel, N. Fusi, M. Schaefer, M. A. Brockman, J. Gilmour, M. A. Price, W. Kilembe, R. Haubrich, M. John, S. Mallal, R. Shapiro, J. Frater, P. R. Harrigan, T. Ndung’u, S. Allen, D. Heckerman, J. Sidney, T. M. Allen, P. J.R. Goulder, Z. L. Brumme, E. Hunter+, P. A. Goepfert+ Impact of Pre-adapted HIV Transmission *Nature Medicine* 2016, 22, pp 606-613 + These authors jointly supervised this work \* These authors contributed equally to this work Free access via the Springer Nature sharing initiative: pdf Check out the video explaining the main findings: MSR webpage
  6. T. Schoofs\*, F. Klein\*, M. Braunschweig\*, K. F. Kreider, A. Feldmann, L. Nogueira, T. Oliveira, J. C. C. Lorenzi, E. H. Parrish, G. H. Learn, A. P. West, P. J. Bjorkman, S. J. Schlesinger, M. S. Seaman, J. Czartoski, M. J. McElrath, N. Pfeifer, B. H. Hahn, M. Caskey, and M. C. Nussenzweig HIV-1 therapy with monoclonal antibody 3BNC117 elicits host immune responses against HIV-1 *Science* 2016, 352 (6288), pp 997-1001 \*: equal contribution
  7. A. Hake and N. Pfeifer Prediction of HIV-1 sensitivity to broadly neutralizing antibodies shows a trend towards resistance over time. *PLOS Computational Biology* 2017, 13(10):e1005789

## Abstracts of Talks

### **NDExR and Cytoscape: Interactive and automated visualization of biological networks using R**

Florian Auer, Frank Kramer, Tim Beißbarth

University Medical Center Göttingen

florian.auer@med.uni-goettingen.de

Network models form a simple and flexible way of representing diverse associations within complex systems, and its applications are well-established in a wide range of fields in biology. Within a common bioinformatics workflow data integration, network analysis and visualization accompany each other, and comprise fundamental challenges of combining various tools. Implementation of fully automated pipelines enhances the intricacy of such tasks furthermore.

Using standard technologies, we demonstrate a course from data acquisition to the finished visualization, and options to achieve the individual sub tasks. Thereby the network data exchange (NDEx) platform and the Cytoscape project, and appendant R packages, form the core components. We use our R package ndexr to retrieve networks from the public NDEx platform, and also to store the results for later collaboration and publication. Cytoscape is one of the most popular open-source software tools for the visual exploration of biomedical networks. Beside the graphical interface, the latest release offers a RESTful interface and R packages providing access to it.

Along an exemplary bioinformatics workflow, we demonstrate how the single steps can be performed not only interactively, but also in a fully automated manner. Each step can be done using Cytoscape or R, or a combination of both: Controlling Cytoscape remotely from within R. Starting at an interactive analysis we move towards automation, illustrating the interchangeability and flexibility of the different approaches.

## Comparative agent-based modeling of *A. fumigatus* infections in mice and men

Marco Blickensdorf (Speaker), Sandra Timme, Marc Thilo Figge

Leibniz-Institut für Naturstoff-Forschung und Infektionsbiologie – Hans-Knöll-Institut (HKI) Jena

marco.blickensdorf@hki-jena.de

The concept of systems biology constitutes a powerful tool to investigate biological systems. Thereby, wet-lab and dry-lab studies mutually support and complement each other. However, systems biology of infection often faces several problems on both sides of the systems biology cycle: First, since experiments can only be conducted in animal models the transferability of results to the human system is difficult. Second, even in animal experiments infection dynamics cannot be captured in all sites of infection, such as the lung. However, virtual infection modeling provides the possibility to overcome the aforementioned limitations by integration of all available experimental data and thereby drives the research in systems biology of infection.

In our current work we use virtual infection modeling to investigate *Aspergillus fumigatus* lung infections. *A. fumigatus* is an environmental wide spread fungus that is opportunistic to humans and can cause severe infections in immunocompromised patients. Its spores, also called conidia, may reach the lower respiratory tract of the lung and, if not efficiently attacked by the immune system, cause invasive pulmonary aspergillosis with high mortality rates of 30%-90% making it a relevant target for research. Due to its complex interactions with the host immune system and its ability to adopt different morphologies many levels of pathogenicity have to be considered for development of effective therapy.

Many mammalian species have been used for experimental research on *A. fumigatus* infection. Besides rats, rabbits and guinea pigs mice models have been used most extensively. However, during experiment usually much higher infection dosages compared to the natural inhalation dosage are used in order to provoke an adequate immune response. Therefore, not only the different systems with respect to the animal model but also differences in infection doses need to be considered for knowledge transfer to the human system. Surprisingly little is known about the comparability and transferability of mouse infection models in wet-lab and natural *A. fumigatus* infections in human. Therefore, we attempt to shed light on the infection with *A. fumigatus* in man and mice with respect to their different lung morphologies and infection doses in natural environment and experiment.

In recent studies we implemented an agent-based virtual infection model for the simulation of early *A. fumigatus* infections. This model reconstructs a human alveolus in three-dimensional continuous space, represented by a three-quarter sphere containing epithelial cells, alveolar macrophages (AM) as representatives of the immune system and the *A. fumigatus* conidium.

In a first publication we could show that random walk migration of AM is insufficient to detect the conidium in a reasonable time frame. Furthermore, we studied the impact of various migration parameters on the success of conidium detection [1]. Finally, we proposed the existence of a chemotactic signal that guides AM to the position of the conidium. In a second study we extended our virtual infection model by a mechanism for chemokine secretion and diffusion, where the conidium-associated epithelium secretes a virtual chemokine and a gradient is build up assisting the migration of alveolar macrophages [2]. We scanned for a broad range of parameters to gain understanding in how the parameters of chemotaxis impact on the efficiency of infection clearance. We could show that the ratio of diffusion constant and secretion rate of the chemokine needs to be high in order to establish a gradient that facilitates AM to find the conidium before onset of germination. In a third study we applied evolutionary games on graphs that allowed for qualitative predictions on the immune response

in different infection doses with *A. fumigatus* in the human lung [3]. Thereby, we considered the complement system, alveolar macrophages and the recruitment of neutrophils and could resolve the different roles of AM in different *A. fumigatus* infection doses.

Our current work aims to compare *A. fumigatus* infections in man and mice *in silico* considering natural and experimental infection dosages. Therefore, the previously established agent-based virtual infection model was used and adopted to the morphometry of mouse alveoli. This enabled us to generate quantitative predictions on the influence of morphological factors as well as dose-dependent effects during *A. fumigatus* infection.

Publications:

1. Pollmächer, J. & Figge, M. T. Agent-based model of human alveoli predicts chemotactic signaling by epithelial cells during early *Aspergillus fumigatus* infection. *PLoS One* 9, e111630 (2014).
2. Pollmächer, J. & Figge, M. T. Deciphering chemokine properties by a hybrid agent-based model of *Aspergillus fumigatus* infection in human alveoli. *Front. Microbiol.* 6, 503 (2015).
3. Pollmächer, J. et al. Deciphering the counterplay of *Aspergillus fumigatus* infection and host inflammation by evolutionary games on graphs. *Sci. Rep.* 6, 27807 (2016).

**Predicting comorbidities of epilepsy patients using big data from Electronic Health Records combined with biomedical knowledge**

Thomas Gerlach

University Bonn, Bonn-Aachen International Center for Information Technology (B-IT)

t.gerlach1@gmx.de

Epilepsy is a complex brain disorder characterized by repetitive seizure events. Epilepsy patients often suffer from various and severe physical and psychological comorbidities. While general comorbidity prevalence and incidences can be estimated from epidemiological data, such an approach does not take into account that actual patient specific risks can depend on various individual factors, including medication. This motivates to develop a machine learning approach for predicting individual comorbidities. To address these needs we used Big Data from electronic health records (~100 Million raw observations), which provide a time resolved view on an individual's disease and medication history. A specific contribution of this work is an integration of these data with information from 14 biomedical sources (DisGeNET, TTD, KEGG, Wiki Pathways, DrugBank, SIDER, Gene Ontology, Human Protein Atlas, ...) to capture putative biological effects of observed diseases and applied medications. In consequence we extracted >165,000 features describing the longitudinal patient journey of >10,000 adult epilepsy patients. We used maximum-relevance-minimum-redundancy feature selection in combination with Random Survival Forests (RSF) for predicting the risk of 9 major comorbidities after first epilepsy diagnosis with high cross-validated C-indices of 76 - 89% and analyzed the influence of medications on the risk to develop specific comorbidities. Altogether we see our work as a first step towards earlier detection and better prevention of common comorbidities of epilepsy patients.

## Loss-Function Learning for Digital Tissue Deconvolution

Franziska Görtler

Institute of Functional Genomics, Statistical Bioinformatics, Universität Regensburg

Franziska.Goertler@klinik.uni-r.de

The gene expression profile of a tissue averages the expression profiles of all cells in this tissue. Digital tissue deconvolution (DTD) addresses the following inverse problem: Given the expression profile  $y$  of a tissue, what is the cellular composition  $c$  of that tissue?

If  $X$  is a matrix whose columns are reference profiles of individual cell types, the composition  $c$  can be computed by minimizing  $L(y-Xc)$  for a given loss function  $L$ . Current methods use predefined all-purpose loss functions. They successfully quantify the dominating cells of a tissue, while often falling short in detecting small cell populations.

Here we use training data to learn the loss function  $L$  along with the composition  $c$ . This allows us to adapt to application-specific requirements such as focusing on small cell populations or distinguishing phenotypically similar cell populations. Our method quantifies large cell fractions as accurately as existing methods and significantly improves the detection of small cell populations and the distinction of similar cell types.

## **A deep learning approach for uncovering lung cancer immunome patterns**

Moritz Hess

Institut für Medizinische Informatik, Statistik und Epidemiologie

[hess@imbi.uni-freiburg.de](mailto:hess@imbi.uni-freiburg.de)

Tumor immune cell infiltration is a well known factor related to survival of cancer patients. This has led to deconvolution approaches that can quantify immune cell proportions for each individual. What is missing, is an approach for modeling joint patterns of different immune cell types. We adapt a deep learning approach, deep Boltzmann machines (DBMs), for modeling immune cell gene expression patterns in lung adenocarcinoma. Specifically, a partially partitioned training approach for dealing with a relatively large number of genes. We also propose a sampling-based approach that smooths the original data according to a trained DBM and can be used for visualization and clustering. The identified clusters can subsequently be judged with respect to association with clinical characteristics, such as tumor stage, providing an external criterion for selecting DBM network architecture and tuning parameters for training. We show that the hidden nodes of the trained networks cannot only be linked to clinical characteristics but also to specific genes, which are the visible nodes of the network. We find that hidden nodes that are linked to tumor stage and survival represent expression of T-cell and mast cell genes among others, probably reflecting specific immune cell infiltration patterns. Thus, DBMs, trained and selected by the proposed approach, might provide a useful tool for extracting immune cell gene expression patterns. In the case of lung adenocarcinomas, these patterns are linked to survival as well as other patient characteristics, which could be useful for uncovering the underlying biology.

## Assessing false positive findings in a viral detection pipeline using high-throughput sequencing data

Jochen Kruppa<sup>1,2</sup>, Klaus Jung<sup>1</sup>

<sup>1</sup>Institute for Animal Breeding and Genetics, University of Veterinary Medicine Hannover, Hannover

<sup>2</sup>Institute of Medical Biometrics and Clinical Epidemiology, Charité - Universitätsmedizin Berlin, Berlin

klaus.jung@tiho-hannover.de

**Abstract.** The analysis of high-throughput data has become more and more important in genomic research. Here, we want to present the development of a virus detection pipeline for the analysis of high-throughput data in infection research. The approach is using high-throughput data generated by next generation sequencing (NGS), which has become the state of the art for the analysis of genomic samples. Importantly, the viral detection pipeline is designed for revealing viral sequences in reference free host organisms. Normally, the raw reads are filtered from the host reads by mapping the reads to the host reference. Afterwards larger assemblies, contigs, can be build. This is not always possible in the case of infectious zoonoses research because the host reference is often not available.

Here, we present a full bioinformatic pipeline for the NGS-based virus detection: The mapping of the reads to an artificial genome consisting of 2.4 million viral sequences, the translation of DNA sequencing reads into amino acid sequences, which are then mapped to an artificial genome consisting of 3.3 million amino acid sequences, and the visualization of mapping results in an assembly plot. Due to many multiple mapped reads many false positive findings can occur. Therefore, we estimate beforehand the detection error rates using a decoy sequence database [1]. The decoy database allows to compare different DNA mapper by the false positive rates of the mapped reads. (<https://github.com/jkruppa/viralDetectTools>)

### References

1. Reidegeld, K. A., et al. (2008). An easy-to-use Decoy Database Builder software tool, implementing different decoy strategies for false discovery rate calculation in automated MS/MS protein identifications. *Proteomics*, 8(6), 1129-1137.

## **Integral individualized model of hematopoiesis explains dynamics of thrombocytes, granulocytes and lymphocytes under different chemotherapeutic and supportive treatments scenarios**

Yuri Kheifetz

IMISE (INSTITUT FÜR MEDIZINISCHE INFORMATIK, STATISTIK UND EPIDEMIOLOGIE)

yuri.kheifetz@gmail.com

### Objectives

Decreased platelet and leukocyte counts, called respectively thrombocytopenia and leukopenia, are major dose-limiting side effects of dose-intensive anti-cancer chemotherapies. However, standard courses of many chemotherapies result in considerable variability in drug-induced blood cells dynamics. Additional complexity results from different supportive treatments such as administration of growth factors and transfusion of stem cells or platelets. Numerous studies imply that maturations of different blood cell lines are interdependent and influenced from stem-cell-niches supporting osteoblasts. A major challenge of individualized medicine is to consider all relevant factors for optimal risk management using individualized modeling. To solve this task we revised biomathematical models of average human thrombopoiesis and granulopoiesis under chemotherapy or growth factor treatments (Scholz et al. 2004, 2010, 2013), combined them together as well as with model of osteoblasts/osteoclasts dynamics of other group (Komarova et al. 2003) and our novel model of lymphopoiesis.

### Methods

We performed bio-mechanistic modelling of the dynamics of bone marrow hematopoietic and mature circulating cells by ordinary differential equations. Amplifications, death rates and transition times of the system are regulated by biologically motivated feedback loops. We introduced quiescent state for stem as well as for progenitor cells. Activation of progenitor cells is mediated by interactions of growth factors, quiescent cells and osteoblasts. Attached pharmacokinetic and –dynamic models consider injections of growth factors as well as of cytotoxic drugs. Short-range treatment effects influence proliferating blood-cell precursors, while both chemotherapy and G-CSF induce a long-term depletion of osteoblasts reducing the supporting capacity of the bone marrow. Our novel lymphopoiesis model describes short- and long-living lymphocytes, circulating between blood and peripheral compartments and originating from hematopoietic stem cells after differentiation through few empirical compartments. We fitted 24 individual and 59 population parameters using simultaneously data from 11 studies measuring 19 different biological outcomes (cell counts of platelets, neutrophils, lymphocytes, leukocytes, megakaryocytes of different ploidy, osteoblasts, banded and segmented granulocytes; concentration of granulocyte-colony-stimulating factor (G-CSF), thrombopoietin (TPO) and prednisone). These 11 studies contained either individual or averaged data on hematopoiesis under five different chemotherapy regimens and stimulatory treatments by TPO, filgrastim, pegylated filgrastim (synthetic variants of G-CSF) and prednisone. We successfully applied our novel parameters estimation methodology for such complex case earlier during a versatile fitting of our individualized thrombopoiesis model.

### Results & Conclusions

We succeeded to model major blood cell lines' development and dynamics perturbed by a wide spectrum of chemotherapies as well as supportive treatments. Several new biological insights have been described. We modelled a well-known complex negative synergism between G-CSF and TPO competing on a choice between granulopoietic and thrombopoietic differentiation alternatives of progenitor cells. According to several independent studies, we upgraded our model by direct stimulating effect of G-CSF and TPO on stem as well as early progenitor cells. We have found that multi-cyclic chemotherapy significantly reduces transit times for megakaryocytes and platelets. The long-term decrease in aver-

age platelets level during multi-cyclic chemotherapy was attributed to interactions between osteoblasts, quiescent and active progenitor cells compartments. These slow changes are responsible for strong intra-individual variability of platelets' nadirs and consequently of chemotoxicity through treatment cycles. Our model described successfully few regimens of high-doses chemotherapy accompanied by bone-marrow transplantant. We achieved an optimal tradeoff between goodness of fit and overfitting for most of the patients. Heterogeneity between patients can be traced back to heterogeneity of several model parameters. The predictive potential of the model was successfully proved for thrombopoiesis and we will exploit for other cell lines the in the near future.

**Modelling repression of E2F1 by cooperative microRNA pairs in the context of anticancer chemotherapy resistance**

Xin Lai

University Hospital of Erlangen

xin.lai@uk-erlangen.de

**Abstract:** High rates of lethal outcome in tumour metastasis are associated with the acquisition of invasiveness and chemoresistance. Several clinical studies indicate that E2F1 overexpression across high-grade tumours culminates in unfavourable prognosis and chemoresistance in patients. Thus, fine-tuning the expression of E2F1 could be a promising approach for treating patients showing chemoresistance.

We integrated bioinformatics, structural and kinetic modelling, and experiments to study cooperative regulation of E2F1 by microRNA (miRNA) pairs in the context of anticancer chemotherapy resistance. We showed that an enhanced E2F1 repression efficiency can be achieved in chemoresistant tumour cells through two cooperating miRNAs. Sequence and structural information were used to identify potential miRNA pairs that can form tertiary structures with E2F1 mRNA. We then employed molecular dynamics simulations to show that among the identified triplexes, miR-205-5p and miR-342-3p can form the most stable triplex with E2F1 mRNA. A mathematical model simulating the E2F1 regulation by the cooperative miRNAs predicted enhanced E2F1 repression, a feature that was verified by in vitro experiments. Finally, we integrated this cooperative miRNA regulation into a more comprehensive network to account for E2F1-related chemoresistance in tumour cells. The network model simulations and experimental data indicate the ability of enhanced expression of both miR-205-5p and miR-342-3p to decrease tumour chemoresistance by cooperatively repressing E2F1. Our results suggest that pairs of cooperating miRNAs could be used as potential RNA therapeutics to reduce E2F1-related chemoresistance.

## Computational modeling of DNA de-methylation treatment of Acute Myeloid Leukemia

Jens Przybilla

Leipzig University/ Institute for Medical Informatics, Statistics and Epidemiology

przybilla@izbi.uni-leipzig.de

In Acute Myeloid Leukemia (AML) the number of associated gene mutations is low but some of them are related to epigenetic modifiers. One frequently observed epigenetic disturbance is DNA hypermethylation of gene promoters which often correlates with a block of differentiation. This is also related to modifications in gene expression. Treatment of AML patients with DNA methyltransferase (DNMT) inhibitors results in global hypomethylation of genes and thereby, can lead to a reactivation of the differentiation capability. Unfortunately, after termination of treatment both hypermethylation and differentiation block return in many cases.

Here, we apply for the first time a computational model of epigenetic regulation of transcription in order i) to provide a mechanistic understanding of the DNA (de-) methylation process in AML and ii) to improve the DNA de-methylation treatment strategies. The model considers a cell population of about 100 individual cells. Each cell contains an artificial transcription factor network that is regulated by itself and as an additional layer by epigenetic regulatory factors. These factors are DNA methylation of promoters, the activating histone modification H3K4me3 and the repressing one H3K27me3. By computational simulations, we analyze promoter hypermethylation scenarios referring to DNMT dysfunction, decreased H3K4me3 and increased H3K27me3 modification activity and accelerated cell proliferation. We quantify differences between these scenarios with respect to gene repression and activation. Moreover, we compare the scenarios regarding their response to DNMT inhibitor treatment alone and in combination with inhibitors of H3K27me3 histone methyltransferases and of H3K4me3 histone demethylases.

We find that the different hypermethylation scenarios respond specifically to therapy, suggesting that failure of remission originate in patient specific deregulation. We observe that inappropriate demethylation therapy can result even in enforced deregulation. As an example, our results suggest that application of high DNMT inhibitor concentration can induce unwanted global gene activation if hypermethylation originates in increased H3K27me3 modification. Our results underline the importance of a personalized therapy requiring knowledge about the patient-specific mechanism of epigenetic deregulation.

## From heterogeneous data of biological systems to quantitative predictive models

Nicole Radde

Institute for Systems Theory and Automatic Control, University of Stuttgart, Pfaffenwaldring 9, 70569 Stuttgart, Germany

Nicole.Radde@ist.uni-stuttgart.de

Experimental techniques to monitor cellular processes on a molecular level have rapidly developed and improved in the last decades, providing large amounts of data on different scales and of different nature. The integration of these data into computational models is today a major challenge in systems biology. Appropriate methods for data pre-processing, model-based experiment design, model calibration and uncertainty quantification are necessary on the way to build predictive models and to improve the quality of in silico experiments.

Here I will exemplarily show some of these challenges on two particular projects within my research group. In the first example, we use published data on the mitogen-activated protein kinase (MAPK) signaling pathway in PC12 cell lines in order to investigate context-dependent responses of this pathway to different growth factors [1]. We use statistical approaches for model calibration, which allow consistent uncertainty quantification from noise in experimental data to variances in model predictions for any quantity of interest. Our study reveals a new mechanism termed quasi-bistability that might play a role in cellular decision processes.

The second project was done in collaboration with partners from Cell Biology. Using single-cell time-lapse microscopy data on the spindle assembly checkpoint efficiency for different fission yeast strains [2], we built a finite mixture modeling framework that is able to integrate data from various experiments and to handle right- and interval-censored data. Here we were able to generate biologically insightful hypotheses about the appearance of subpopulation structures under different experimental conditions [3]. Moreover, the integration of censored data into models constitutes several problems and challenges that are also interesting from a theoretical viewpoint, and we review some of them in the presentation.

### References

- [1] Jensch A, Thomaseth C, and Radde N (2017). Sampling-based Bayesian approaches reveal the importance of quasi-bistable behavior in cellular decision processes on the example of the MAPK signaling pathway in PC-12 cell lines. *BMC Syst Biol* 11:11.
- [2] Heinrich S, Geissen E-M, Kamenz J, Trautmann S, Widmer C, Drewe P, Hauf S (2013). Determinants of robustness in spindle assembly checkpoint signaling. *Nat Cell Biol* 15, 1328–1339.
- [3] Geissen E-M, Hasenauer J, Heinrich S, Hauf S, Theis FJ, Radde N (2016). MEMO: multi-experiment mixture model analysis of censored data. *Bioinformatics* 32(16):2464-72.

## **Modelling of Cooperativity Mechanisms in T cell Killing**

Ananya Rastogi

Department of Systems Immunology and Braunschweig Integrated Centre of Systems Biology, Helmholtz Centre for Infection Research, Braunschweig, Germany

ananyarastogi13@gmail.com

Cytotoxic T cells interact with infected cells at the site of infection and are able to kill them, thereby limiting viral propagation. In 2016, Halle et al monitored the *in vivo* killing of infected cells by CTLs using 2-photon microscopy. It revealed that the death fate of infected cells required multiple contacts with CTLs and was highly dependent on the number of contacts. This suggested the existence of cooperativity, from an unknown mechanism yet, on the CTL killing process.

We developed a three dimensional agent based model to simulate and study the possible mechanisms of cooperation. By a new method of analysis; we could show that the probability of death was linearly dependent on the number of CTL contacts, and we could therefore discard the null hypothesis where CTLs have a constant probability of killing at each encounter. The question is now to understand by which mechanisms this increased probability of killing occurs. We are now considering hypotheses on different levels. These include cooperativity to attract T cells to the infected cells so that quick successive contacts can kill the infected cells. At the interaction level, we are exploring how signal integration at the CTL side (accumulated TCR signalling) or at the infected cell side (accumulated damage) could contribute to the observed cooperativity.

For each hypothesis, I will present which readouts can be used to confront them with experimental results; such as the killing patterns and the per capita killing rate (PKCR). Finally, I will show how the modelling approach can suggest some experiments that can further our knowledge about this intriguing apparent cooperativity.

## **Quantitative analysis of the 3D nuclear landscape**

Volker Schmid

LMU Munich, Department of Statistics

volker.schmid@lmu.de

Recent advancements of super-resolved fluorescence microscopy have revolutionized microscopic studies of cells, including the exceedingly complex structural organization of cell nuclei in space and time. In this paper we describe and discuss tools for (semi-) automated, quantitative 3D analyses of the spatial nuclear organization. These tools allow the quantitative assessment of highly resolved different chromatin compaction levels in individual cell nuclei, which reflect functionally different regions or sub-compartments of the 3D nuclear landscape, and measurements of absolute distances between sites of different chromatin compaction. In addition, these tools allow 3D mapping of specific DNA/RNA sequences and nuclear proteins relative to the 3D chromatin compaction maps and comparisons of multiple cell nuclei. The tools are available in the free and open source R packages `nucim` and `bioimage-tools`. We discuss the use of masks for the segmentation of nuclei and the use of DNA stains, such as DAPI, as a proxy for local differences in chromatin compaction. We further discuss the limitations of 3D maps of the nuclear landscape as well as problems of the biological interpretation of such data.

## **An agent-based model for the F-actin driven localization of TCR, LFA-1 and CD28 in the immunological synapse**

Anastasios Siokis (1), Philippe Robert (1), Philippos Demetriou (3), Michael Dustin (3), Michael Meyer-Hermann(1, 2)

(1) Department of Systems Immunology and Braunschweig Integrated Centre of Systems Biology, Helmholtz Centre for Infection Research, Braunschweig, Germany.

(2) Institute of Biochemistry, Biotechnology and Bioinformatics, Technische Universität Braunschweig, Braunschweig, Germany.

(3) Nuffield Department of Orthopedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Headington, United Kingdom.

Antigen recognition by T cells is a key step of every adaptive immune response. In the cell-cell junction between T and Antigen Presenting cells, known as immunological synapse (IS), signaling complexes, integrins, and costimulatory molecules like CD28 exhibit a specific pattern, essential for T cell activation and fate decision. Despite extensive knowledge on which molecules and signaling pathways participate in T cell activation, the mechanisms that regulate the spatial organization of these molecules during IS formation are poorly understood.

To gain insights into these mechanisms we developed an agent based model for the IS formation. The in silico experiments simulate the dynamics of the critical surface molecules of the two interacting cells (T and Antigen presenting). The model is calibrated based on experimental high resolution microscopy imaging results and shows that F-actin driven centripetal flow is crucial for the formation of the characteristic IS pattern. An emerging LFA-1 gradient in the periphery of the contact region towards the center, impacts on the IS formation and affects molecular localization, also observed experimentally. The characteristic CD28-CD80 annular structure around the cSMAC only emerges under an optimal CD28 actin coupling strength that induces centripetal motion. The model has deciphered the effects of actin coupling in complex experimental set-ups which are difficult to interpret.

The presented model shows that functional properties of the IS can be extracted from imaging data, and that actin forces are a major player in the formation of a proper synapse. The model is a cutting edge basis to predict the effect of potential therapeutics targeting actin-related pathways, and to decipher the strength of new mechanisms for molecular transport in the IS.

## **Quantitative Cell Cycle Analysis Based on an Endogenous All-in-One Reporter for Cell Tracking and Classification**

Thomas Zerjatke,<sup>1</sup> Igor A. Gak,<sup>2</sup> Dilyana Kirova,<sup>2</sup> Jörg Mansfeld<sup>2</sup> , and Ingmar Glauche<sup>1</sup>

<sup>1</sup>Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, 01307 Dresden, Germany

<sup>2</sup>Cell Cycle, Biotechnology Center, Technische Universität Dresden, 01307 Dresden, Germany

Thomas.Zerjatke@tu-dresden.de

Cell fate decisions, such as reprogramming, differentiation, and cell cycle exit, are tightly linked to cell cycle kinetics. Depending on their stage in the cell cycle, cells respond differently to internal or external cues, such as growth factors or differentiation stimuli. In recent years, long-term live cell imaging has provided extensive insights into the dynamics of key driving factors of cell cycle regulation at a single-cell level. However, so far a system of up to four fluorescent reporter constructs was needed to distinguish all cell cycle phases, thus leaving little room for simultaneously imaging additional target proteins.

Here, we present fluorescently tagged endogenous proliferating cell nuclear antigen (PCNA) as an all-in-one cell cycle reporter in long-term live cell imaging. We established an image analysis pipeline that allows segmenting and tracking single cells based on nuclear PCNA expression in proliferating cells. Furthermore, based on the kinetics of PCNA intensity and its spatial distribution we are able to classify all cell cycle stages.

This now gives us the possibility to simultaneously quantify the dynamic expression of a higher number of cell cycle related proteins and thus study their role in decision-making upon cell cycle progression and potential correlations between them. Combining the all-in-one reporter with labelled endogenous cyclin D1 and p21 as prime examples of cell-cycle-regulated fate determinants, we show how cell cycle and quantitative protein dynamics can be simultaneously extracted to gain insights into G1 phase regulation.

## Abstracts of Posters

### Data Driven Computational Modeling of Hematopoiesis and Myelodysplastic Syndrome

Lisa Bast<sup>1,2,\*</sup>, Michele Kyncl<sup>3,\*</sup>, Robert Oostendorp<sup>3</sup>, Katharina Götze<sup>3</sup>, and Carsten Marr<sup>1</sup>

<sup>1</sup>Institute of Computational Biology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany.

<sup>2</sup>Department of Mathematics, Chair of Mathematical Modeling of Biological Systems, Technische Universität München, Garching, Germany.

<sup>3</sup>Department of Medicine III, Technical University of Munich, Klinikum rechts der Isar, Munich, Germany.

[lisa.bast@tum.de](mailto:lisa.bast@tum.de)

Myelodysplastic Syndrome (MDS) describes a group of diverse bone marrow disorders, in which an abnormal growth of immature myeloid cells prevents the production of healthy blood. Although MDS blood stem cells have been reported to proliferate less effectively, for about 30% of diagnosed patients this ineffective hematopoiesis leads to acute myeloid leukemia (AML), characterized by a pathologically high abundance of immature white blood cells. In our project we seek to identify the differences of healthy hematopoiesis and blood cell production in the MDS case. A fixed amount of hematopoietic stem cells are isolated from the bone marrow of patients and healthy controls respectively, cultured in vitro and observed by fluorescence-activated cell sorting (FACS) at different time points. The resulting cell counts are used to fit a mathematical model describing the evolution of hematopoietic cell types with ordinary differential equations. Multistart optimization leads to parameter estimates, which uncover potential differences of dysfunctions in the MDS case in comparison to healthy hematopoiesis. Integrating the estimated rates into a comprehensive mathematical model of human hematopoiesis might eventually help to shed light on the progression of MDS towards AML.

\* Equal contributions

## **Compositional modeling of core splicing regulation improves prediction of SNP effects on splicing**

Jun Cheng, Nguyen Thi Yen Duong, Žiga Avsec, Julien Gagneur

Technical University of Munich

chengju@in.tum.de

Genetic variants disrupting RNA splicing is a frequent cause of genetic disorders. Various sequence-based models have been developed to understand the sequence determinants of alternative splicing, and to predict effects of genetic variations on splicing. These models are, however, limited to certain types of alternative splicing, and only accept variants that appear in restricted genomic regions. Here, we propose a sequence-based deep learning model that overcomes these limitations and predicts effects of genetic variations on splicing, including SNPs and indels, from any location in a gene. To this end, we followed a transfer learning approach. We first trained sequence-based building-block models modeling effects of 5' splice site, 3' splice site, intronic and exonic sequence on split read counts from a massive parallel splicing reporter assay (Rosenberg et al., 2015). A beta-binomial cost function was developed to account for over-dispersion. Next, models to predict common alternative splicing types such as exon skipping, alternative 5' splicing and alternative 3' splicing were obtained by composing these building-block models and fitting to data of endogenous splice sites. Our model improves the prediction of SNP effects at 5' splicing site compared to the competitive models on the GEUVADIS dataset (Lappalainen et al. 2013). We expect our model to help for the interpretation of genetic variants associating with common and rare diseases.

## **Graph-based Convolutional Neural Networks for analyzing pathways in cancer**

Hryhorii Chereda

Universitätsmedizin Göttingen, Institut für Medizinische Statistik

[hryhorii.chereda@med.uni-goettingen.de](mailto:hryhorii.chereda@med.uni-goettingen.de)

In recent years deep learning was applied to a wide range of problems in various areas. Such deep learning tools as convolutional neural networks (CNNs) have been shown to work well in natural language processing and computer vision, especially at image classification tasks. Furthermore, CNN's have been applied to bioinformatic challenges like patient stratification tasks. Nowadays, deep learning (including CNNs) is extending to Non-Euclidean domains such as graph-structured data and manifolds. We are planning to map gene-expression data to the vertices of biological pathways and feed this graph-structured data into CNNs in order to classify patients.

The usual CNN architecture consists of three types of layers: convolutional layers, pooling layers, and fully connected layers. The first two layers utilize the structure of the data preparing informative features for the fully connected neural network layers. In our work, we consider three popular, but different approaches developed for application of CNN on graph-structured data. Our research aims to compare these approaches in order to address the question if the use of graph-based CNNs is able to provide valuable classification improvements by utilizing prior pathway knowledge.

## **Learning the Topology of Latent Signaling Networks from High Dimensional Transcriptional Intervention Effects**

Zahra Nasrollah, Achim Tresch, Holger Fröhlich

Institute of Medical Statistics and Computational Biology (IMSB), Faculty of Medicine University of Cologne

znasroll@uni-koeln.de

Data based learning of the topology of molecular networks, e.g. via Dynamic Bayesian Networks (DBNs) has a long tradition in Bioinformatics. The majority of methods take gene expression as a proxy for protein expression in that context, which is principally problematic. Further, most methods rely on observational data, which complicates the aim of causal network reconstruction. Nested Effects Models (NEMs – Markowitz et al., 2005) have been proposed to overcome some of these issues by distinguishing between a latent (i.e. unobservable) signaling network structure and observable transcriptional downstream effects to model targeted interventions of the network.

The goal of this project is to develop a more principled and flexible approach for learning the topology of a dynamical system that is only observable through transcriptional responses to combinatorial perturbations applied to the system. More specifically, we focus on the situation in which the latent dynamical system (i.e. signaling network) can be described as a network of binary state variables with logistic activation functions. We show how candidate networks can be scored efficiently in this case and how topology learning can be performed via Markov Chain Monte Carlo (MCMC).

As a first step, we extensively tested our approach by applying it to several known network motifs (Feed Forward, Feed Backward, Bifan, Diamond, Protein Cascading) over a wide range of possible settings (e.g. different number of observations, time points). Moreover, we evaluated our method with synthetic data generated from ODE systems taken from the literature. As a next step we will evaluate our method with data from the DREAM 8 challenge.

In future work, we also plan to extend our method to incorporate multi-omics data and apply it to patient samples to identify disease related networks.

## Implementing a multilayer framework for pathway data integration, analysis and visualization

Zaynab Hammoud, Frank Kramer

Department of Medical Statistics, University Medical Center Göttingen

zaynab.hammoud@med.uni-goettingen.de

Network theory has been used for many years in the modelling and analysis of complex systems, as epidemiology, biology and biomedicine [1]. As the data evolves and become more heterogeneous and complex, monoplex networks become an oversimplification of the corresponding systems [3]. This imposes a need to go beyond traditional networks into a richer framework capable of hosting objects and relations of different scales [4], called Multilayered Network. These complex networks have contributed in many contexts and fields [1], although they have been rarely exploited in the investigation of biological networks, where their application seems very convenient.[2] In order to fill this gap, we aim to implement a multilayer framework that can be applicable in various domains, especially in the field of pathway modelling. Our idea is to integrate pathways and their related knowledge into a multilayer model, where each layer represents one of their elements. The model offers a feature we call “Selective Inclusion of Knowledge”, as well as a collection of related knowledge into a single graph, like diseases and drugs. The final aim is called personalized medicine, i.e. a medicine focused on the individual and proactive in nature, which would allow a significant improvement in health care by customizing the treatment according to each patient needs.[3] In this poster, we give an overview of the various models of multilayered networks, then we describe the model we are building, and the workflow of implementing it into an R package as well as the future plan.

Keywords: Multilayer Networks, Network Biology, Network Theory, Pathway Modelling

### References

- [1] Kivelä et al., “Multilayer Networks.”. *Journal of Complex Networks* (2014) 2, 203–271
- [2] De Domenico, Porter, and Arenas, “MuxViz: a tool for multilayer analysis and visualization of networks.”. *Journal of Complex Networks* (2015) 3, 159–176
- [3] Boccaletti et al., “The Structure and Dynamics of Multilayer Networks.”. *Physics Reports* 544, 1 (2014) arXiv:1407.0742
- [4] Traxl, Boers, and Kurths, “Deep Graphs - a General Framework to Represent and Analyze Heterogeneous Complex Systems across Scales.”. *Chaos* 26, 065303 (2016) arXiv:1604.00971

## **Conducting gene set enrichment tests in meta-analyses of simulated transcriptomic data and West-Nile virus infected samples**

Kosch R<sup>1</sup>, Becker SC<sup>2</sup>, Jung K<sup>1</sup>

1) Institute for Animal Breeding and Genetics, University of Veterinary Medicine Hannover

2) Institute for Parasitology, University of Veterinary Medicine Hannover

robin.kosch@tiho-hannover.de

### Introduction and aim

Neuroinfections are regularly studied on the level of the transcriptome using either high-throughput microarray or sequencing technologies to compare infected versus uninfected samples. Due to the high-dimensional character of such data, results of individual studies are little robust and have a limited level of evidence. Therefore, meta-analysis of gene expression profiles is an important tool to obtain more stable results. The enrichment of pathways or gene ontology (GO) terms among the differentially expressed genes is helpful to obtain biologically more meaningful results. Our aim is to study analysis pipelines for enrichment analysis in meta-analysis on simulated transcriptomic data and applied biological data from West-Nile virus infected mice.

### Material and methods

We simulated gene expression data in two different ways. First, we build datasets with totally simulated values, drawn from multivariate normal distributions. Second, we utilized a large dataset with real biological outcomes and use its subsets to conduct a meta-analysis. Only gene sets were simulated for the latter dataset. Three analysis pipelines for combining the studies were evaluated: 1) direct merging of the individual data sets (early merging), 2) merging of the individual results of differential expression analysis (intermediate merging), 3) merging of the individual results of enrichment analysis (late merging). In addition to an altering timing of data merging, we compared different testing methods for gene set enrichment analysis.

### Results

By altering the parameters of the constructed studies, we could test the pipelines for various data scenarios. Meta-analysis, comprising several individual studies enhances the accuracy of the early merging, whereas meta-analysis with multiple samples per individual datasets positively affect the late merging. The meta-analysis of WNV-data resulted several enriched genes that were not found in individual studies alone.

### Conclusions

Several Meta-analyses were conducted, giving insight into the accuracy of the applied pipelines.

## **A machine-learning approach towards the prediction of pulmonary hypertension**

Andreas Leha

Universitätsmedizin Göttingen

andreas.leha@med.uni-goettingen.de

The 2015 ESC Guidelines on pulmonary hypertension (PH) suggest to integrate several echocardiographic measurements and signs to classify the probability of PH. However, selection of parameters, cut-offs and in particular the combinatory interpretation of these has not been fully substantiated by evidence. Thus, predicting the need for right heart catheterization (RHC) that should be considered to establish or rule-out a definite diagnosis of PH is not straightforward. Machine-learning offers an unbiased method to derive a classifier selecting the most differentiating variables. Here, we assessed whether machine-learning can expedite PH prediction.

We performed a retrospective cohort study using echocardiographic and demographic data of 90 patients from three European centers with echocardiography and RHC performed within 24 hours. Four machine-learning algorithms (random forest, lasso penalized logistic regression, gradient boosting machines, support vector machines) were evaluated using a 10 times 10-fold cross validation scheme. Their performance in predicting PH based on demographic and echocardiographic variables was assessed by area under the receiver operating curve (AUC) and compared to the best of several examined empirical algorithms.

Based on an invasively determined PAP of  $\geq 25$  mmHg, the cohort contained 68 PH cases (75.6%). The examined machine learning algorithms achieved high prediction accuracies (random forest: AUC 0.87, 95% CI 0.79-0.96) matching those from the best of several examined classical algorithms (Aduen et al.: AUC 0.87, 95% CI 0.78-0.96).

Machine learning, in addition to classical algorithms, can favourably assist in the decision whether to perform invasive PH measurements.

## The Impact of Computational Modeling on Precision Medicine

Jian Li

LMU München

lijian@ibe.med.uni-muenchen.de

The relationship between metabolism and methylation has been considered an important aspect during cancer development and treatment. However, it remains poorly defined how to apply this aspect to improve pre-clinical and clinical treatment outcome. Here, we utilize molecular information to construct a large-scale metabolic model and apply data from the Cancer Cell Line Encyclopedia (CCLE) to perform in silico simulation for investigation of the impact of metabolism on efficiency of chemotherapy. We show that different metabolic components involved in methionine cycle can play a predictive role in chemotherapy treatment. Furthermore, they are potentially also involved in cellular methylation processes. Finally, we show that metabolic pathways for generation of purine and pyrimidine possess high predictive value for chemotherapy treatment. This study demonstrates in a fine manner that computational approaches have strong potential to handle complex issues such as treatment prediction, biomarker identification of systems medicine.

## **Genomewide quantitative analysis of mRNA metabolism with SLAM-seq**

Katharina Moos

Institute for Medical Statistics and Bioinformatics, University of Cologne

katharina.moos@uni-koeln.de

Gene expression regulation is of vital importance for the functionality of the cell. Adaptations of expression are crucial to go through the cell cycle, to carry out specific functions within an organism or to react to environmental changes. Gene expression regulation is achieved by systematic changes in a) mRNA synthesis in the nucleus, b) mRNA transport out of the nucleus into the cytosol, and c) mRNA degradation in the nucleus or d) degradation in the cytosol. Current approaches for the measurement of mRNA metabolism are able to determine synthesis rates as well as cellular degradation rates. However, we are still lacking a method for the genome-wide estimation of nuclear export rates. Here, we want to provide a new biochemical-bioinformatics approach for the measurement of all kinetic parameters of mRNA metabolism. Using an experimental method for metabolic labeling in combination with deep sequencing (SLAM-Seq), we can distinguish transcripts that have been synthesized after the application of the labeling pulse from pre-existing ones. We analyze time series data for nuclear and cytosolic RNA to fit our kinetic model. The most difficult task is a reliable probabilistic modeling of experimental error and bias such as metabolic labeling efficiency. We show in simulations that our model is able to reliably identify its kinetic parameters from realistic count data, and we present first experimental results.

## **Detection of somatic mutations in RNA-Seq data, does it work?**

Alexander WOLFF, Júlia PERERA-BEL and Tim BEIßBARTH

Institute of Medical Statistics, University Medical Center Göttingen

julia.perera@med.uni-goettingen.de

Somatic single nucleotide variants (SNVs) are genomic events with increasing implications in cancer treatment. The gold standard for SNVs detection is whole genome/exome sequencing (WGS/WES) in matched tumor-normal samples, which turns out to be a very costly approach both economically and biologically. On the other hand, RNA sequencing (RNA-Seq) is the most popular technology to study gene expression, novel RNAs, alternative splicing and also, it has the potential for a cost-effective identification of SNVs. SNV calling in RNA-Seq data has been done in matched tumor-normal samples, yet no one has attempted to do so with tumor-only samples. Here we present a method for the identification of SNVs in tumor-only RNA-Seq data putting a special focus on a small panel of clinically relevant SNVs. For evaluation purposes, we analyzed matched tumor-normal WES and tumor-only RNA-Seq data from 14 cancer patients. We compared SNVs detected in RNA-Seq by our method and MuTect2 against the gold standard. We also did a detailed evaluation for a reduced panel of clinically relevant SNVs and reliably identified in RNA-Seq data a subset of mutations (KRAS, NRAS, PIK3CA, BRAF) for which we had pathological annotation. Hence, RNA-Seq rises as a cost-effective and reliable option to detect in parallel gene expression as well as a small panel of clinically relevant SNVs in research.

## TOO MANY ZEROS – A TWO-PART MODEL FOR THE ANALYSIS OF COST DATA IN HEALTH ECONOMY

Iris Reinhard

Biostatistics, Central Institute of Mental Health, Medical Faculty Mannheim / Heidelberg University,  
Germany

iris.reinhard@zi-mannheim.de

In health care research it is common to encounter data characterized by a spike at zero followed by a right-skewed continuous distribution for the positive values. Examples include food consumption in a dietary study, health care utilization and health care expenditures. In the latter case the point mass at zero represents a population of ‘non-users’ who therefore have no costs, while the continuous distribution represents the level of costs for those people who use health services. For statistical analyses in order to understand the influence of therapies, programs, demographic and disease-related variables, alternative approaches are needed to accommodate the discrete and continuous features of the data.

For the identification of possibly influencing factors on semi-continuous right-skewed longitudinal data a two-part model is considered which is based on a two-stage design. The first stage involves modelling the risk for the occurrence of a positive outcome and the second stage models the intensity or the amount of nonzero outcomes. Within that model two sets of covariates / factors can be modeled simultaneously that contribute to separate stages. Data for this analysis come from a controlled prospective intervention study which was investigating the cost-effectiveness of an intervention to reduce compulsory admission into inpatient psychiatric treatment. Analyses for this type of model are performed using a modified SAS-macro of Tooze (2002).

The proposed two-stage model can improve the performance of analyses of semi-continuous health expenditure data. Ignoring the fact that in the intervention study more than 50% of the patients have zero costs may lead to biased results when using the classical linear model approach, even after a possible transformation of the cost data. Tests of prognostic factors in the two parts are leading here to an overlap only in part in the set of predictors. For the second part a gamma regression model could be considered as well and compared via model fit indices.

## **Mathematical modelling of early T helper differentiation and plasticity**

Philippe Robert

Helmholtz-Zentrum Braunschweig

philippe.robert@helmholtz-hzi.de

Quantitative properties of early T helper differentiation: Naive T helper cells integrate a large range of signals into the fate decision of becoming Th1, Th2, Th17 or iTreg (among others), expressing distinct sets of transcription factors and cytokines. We are using mathematical modeling with Ordinary Differential Equations to capture the properties of early T cell effector differentiation in vitro. In order to train the model, the kinetics of expression of master transcription factors and cytokines were monitored under distinct T cell polarizing conditions. In vitro data revealed major latencies in transcription, translation and cytokine secretion, which were important in deciding the timing of cytokine feedbacks and transcription factors expression. Our model successfully reproduced the dynamics of differentiation. These data confirm that 'canonical' in vitro differentiation can be explained by a simple regulatory network. Finally, we show how certain properties of early plasticity can be predicted by the model.

Network-based identification of gene copy number mutations driving oligodendroglioma development

Michael Seifert

IMB TU Dresden

michael.seifert@tu-dresden.de

Oligodendrogliomas represent 4-8% of diagnosed primary human brain tumors. All oligodendrogliomas show a characteristic co-deletion of the chromosomal arms 1p and 19q, but little is known about putative driver genes located in these genomic regions. The occurrence of nearly identical co-deletions in individual oligodendrogliomas does not allow to narrow down the exact location of putative driver genes with standard statistical approaches.

Here, we introduce a novel network-based approach for the identification of putative driver genes in the 1p/19q region. We first learned oligodendroglioma-specific gene regulatory networks [1,2] based on gene expression and gene copy number data of 178 oligodendroglioma patients from TCGA. Next, we used network propagation [1,2] to determine impacts of mutation affected genes (differentially expressed genes within the 1p/19q region) on the expression of known cancer-relevant pathway genes. Comparisons to impacts obtained under random networks revealed 20 putative driver genes that significantly influence the expression of signaling and metabolic pathways. Several of these genes have already been associated with other types of cancer. Moreover, the two top scoring genes, SLC17A7 and ELTD1, have recently been reported to act as tumor suppressor and oncogene in glioblastomas, a closely related tumor type.

We present the first large-scale computational study to pinpoint novel putative driver genes for oligodendrogliomas. Our results indicate that several putative driver genes are located in the 1p/19q region. Generally, our results suggest that our approach is a valuable tool to identify putative tumor drivers in large chromosomal regions affected by DNA copy number mutations.

[1] M. Seifert, B. Friedrich and A. Beyer, Importance of rare gene copy number alterations for personalized tumor characterization and survival analysis, *Genome Biology*, 2016, 17:204.

[2] M. Seifert and A. Beyer, regNet: an R package for network-based propagation of gene expression alterations, *Bioinformatics*, 2018, 34(2), 308-311.

## **Predicting the influence of combination therapies in signaling networks**

Maren Sitte

University Medical Center Göttingen, Institute of Medical Statistics

maren.sitte@med.uni-goettingen.de

The project named “Molecular Mechanisms in Malignant Lymphomas - Demonstrators of Personalized Medicine” compound of research groups of biologists, bioinformaticians and doctors propose to develop prognostic and diagnostic platforms that guide treatment decisions and that support the process of therapeutic target identification in diffuse large B-cell lymphomas (DLBCL). The focus lies on the DLBCL microenvironment as prognostic relevance, which is the foundation of the diagnostic platforms the consortium will establish.

Our aim is to investigate hybrid-models, which will integrate signalling data with existing gene expression data to predict how lymphomas translate signalling stimuli in expression phenotypes. For this approach we analyzed multiplexed proteomics measurements and microarray transcriptomic measurements with Bayesian network approaches as Deterministic Effects Propagation Networks (Bender et al., 2011) and Nested Effects Models (Fröhlich et al., 2008; Markowitz et al., 2005). In a next step we performed a pathway-based data integration and analysis based on prior pathway knowledge using the integration approach which is implemented in R package pwOmics (Wachter A, 2014). The aim is to expand this modelling approach to provide a technique to integrate the direct measurements of phospho-proteins with the indirect measurement of downstream effects.

## Detection of somatic mutations in RNA-Seq data, does it work?

Alexander WOLFF, Júlia PERERA-BEL and Tim BEIßBARTH

Medizinische Statistik

alexander.wolff@med.uni-goettingen.de

Somatic single nucleotide variants (SNVs) are genomic events with increasing implications in cancer treatment. The gold standard for SNVs detection is whole genome/exome sequencing (WGS/WES) in matched tumor-normal samples, which turns out to be a very costly approach both economically and biologically. On the other hand, RNA sequencing (RNA-Seq) is the most popular technology to study gene expression, novel RNAs, alternative splicing and also, it has the potential for a cost-effective identification of SNVs. SNV calling in RNA-Seq data has been done in matched tumor-normal samples, yet no one has attempted to do so with tumor-only samples. Here we present a method for the identification of SNVs in tumor-only RNA-Seq data putting a special focus on a small panel of clinically relevant SNVs. For evaluation purposes, we analyzed matched tumor-normal WES and tumor-only RNA-Seq data from 14 cancer patients. We compared SNVs detected in RNA-Seq by our method and Mutect2 against the gold standard. We also did a detailed evaluation for a reduced panel of clinically relevant SNVs and reliably identified in RNA-Seq data a subset of mutations (KRAS, NRAS, PIK3CA, BRAF) for which we had pathological annotation. Hence, RNA-Seq rises as a cost-effective and reliable option to detect in parallel gene expression as well as a small panel of clinically relevant SNVs in research.

### **Mixed graphical models in metabolomics data analysis optimized with prior knowledge**

Helena U. Zacharias<sup>1</sup>, Karoline Lukaschek<sup>2,3</sup>, Hamimatunnisa Johar<sup>2</sup>, Michael Altenbuchinger<sup>4</sup>, Gabi Kastenmüller<sup>5,6</sup>, KORA study investigators, Karl-Heinz Ladwig<sup>2,7,8</sup>, Fabian J. Theis<sup>1,9</sup>, and Jan Krumsiek<sup>1,5</sup>

<sup>1</sup>Institute of Computational Biology , Helmholtz-Zentrum München, Neuherberg, Germany. <sup>2</sup>Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. <sup>3</sup>Department of Psychosomatic Medicine and Psychotherapy, University of Marburg and Gießen, Germany. <sup>4</sup>Statistical Bioinformatics, Institute of Functional Genomics, University of Regensburg, Regensburg, Germany. <sup>5</sup>German Center for Diabetes Research (DZD) , München-Neuherberg, Germany. <sup>6</sup>Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Neuherberg, Germany. <sup>7</sup>Deutsches Zentrum für Herz-Kreislauf-Forschung (DZHK), Partnersite Munich, Germany. <sup>8</sup>Department of Psychosomatic Medicine and Psychotherapy, Klinikum rechts der Isar, Technische Universität München, Munich, Germany. <sup>9</sup>Department of Mathematics, Technische Universität München, Garching, Germany.

helena.zacharias@helmholtz-muenchen.de

Mixed graphical models (MGMs) are a novel tool for the joint analysis of discrete and continuous variables, extending graphical models to mixed variable types. This analysis method facilitates, e.g., the discovery of important relationships between diseases and the human metabolism, which could lead both to a better understanding of underlying disease mechanisms as well as improved patient care. In the field of metabolomics, a large amount of well researched prior knowledge with regard to specific biochemical pathways is available. Here we demonstrate an easy and time-efficient method for incorporating prior knowledge into the MGM estimation procedure. Our algorithm, which is based on a node-wise LASSO regression, makes use of different weightings of the L1 penalization factors with respect to available prior knowledge. First simulation studies show the superior performance of our MGM-penalization-weighting algorithm with regard to correct edge recovery in comparison to standard MGM estimation methods. Moreover, we demonstrate its value for meaningful data integration with an application to a large-scale serum metabolomics data set, comprising 1,411 participants of the KORA (Cooperative Health Research in the Augsburg Region) follow-up F4 study.

## Linking stem cell function and growth pattern of intestinal organoids

Torsten Thalheim<sup>1</sup>, Marianne Quaas<sup>1,2</sup>, Maria Herberg<sup>1</sup>, Gabriela Aust<sup>2</sup>, Joerg Galle<sup>1</sup>

1 Interdisciplinary Centre for Bioinformatics, Leipzig University, Haertelstr. 16-18, 04107 Leipzig, Germany

2 Department of Surgery, Research Laboratories, Leipzig University, Liebigstr. 19, 04103 Leipzig, Germany

Intestinal stem cells (ISCs) require well-defined signals from their environment in order to carry out their specific functions. Most of these signals are provided by neighboring cells that form a stem cell niche, whose shape and cellular composition self-organize. Major features of this self-organization can be studied in ISC-derived organoid culture. In this system, manipulation of essential pathways of stem cell maintenance and differentiation results in well-described growth phenotypes.

We here provide an individual cell-based model of intestinal organoids that enables a mechanistic explanation of the observed growth phenotypes. In simulation studies of the 3D structure of expanding organoids, we investigate interdependencies between Wnt- and Notch-signaling which control the shape of the stem cell niche and, thus, the growth pattern of the organoids. Similar to *in vitro* experiments, changes of pathway activities alter the cellular composition of the organoids and, thereby, affect their shape.

Exogenous Wnt enforces transitions from branched into a cyst-like growth pattern. Based on our simulation results, we predict that the cyst-like pattern is associated with biomechanical changes of the cells which assign them a growth advantage. As the pattern occurs spontaneously during long term organoid expansion, our results suggest ongoing stem cell adaptation to *in vitro* conditions by stabilizing Wnt-activity. Experimental studies show an incomplete inheritance of the adopted growth pattern indicating an epigenetic origin of the underlying regulation.

Our study exemplifies the potential of individual cell-based modeling in unraveling links between molecular stem cell regulation and 3D growth of tissues.