Workshop on Computational Models in Biology and Medicine 2019

Joint workshop of the GMDS & IBS-DR working groups “Statistical Methods in Bioinformatics” and “Mathematical Models in Medicine and Biology”

March 7th – 8th

Braunschweig Integrated Centre of Systems Biology (BRICS), Germany

Contents

Workshop outline

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

Keynotes

- Ivo Grosse (Halle): "Phylotranscriptomic Hourglass Patterns of Animal and Plant Development and the Emergence of Biodiversity"
- Benjamin Werner (London): "Quantitating somatic evolution in healthy and cancerous human tissues"
- Jan Hasenauer (Bonn): "Mechanistic models of large-scale biochemical reaction networks"

Workshop venue

The workshop will be hosted in the Braunschweig Integrated Centre of Systems Biology (BRICS). Address: Rebenring 56, D-38106 Braunschweig. The lecture hall is easily accessible by public transport. For details, please check in https://www.tu-braunschweig.de/forschung/zentren/brics/kontaktanfahrt.

Organization

The workshop is jointly organized by the GMDS/IBS working groups "Statistical Methods in Bioinformatics" (speakers: Michael Altenbuchinger, University of Regensburg; Klaus Jung, University of Veterinay Medicine Hannover) and "Mathematical Models in Medicine and Biology" (speakers: Markus Scholz, University of Leipzig; Ingmar Glauche, University of Dresden), as well as Michael Meyer-Herrmann and Philippe Robert (Helmholtz Centre for Infection Research) who are also the local organizers.

Contact and local organization

Michael Altenbuchinger, AG Statistical Bioinformatics, Institute of Functional Genomics, University of Regensburg, Am BioPark 9, 93053 Regensburg, Tel. +49 941 943 5095, E-mail: michael.altenbuchinger@klinik.uni-regensburg.de

Philippe Robert, Helmholtz Centre for Infection Research, Rebenring 56, 38106 Braunschweig, Tel.: +49 391 55217, E-mail: philippe.robert@theoretical-biology.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)".

1 Program

Thursday, March 7, 2019

Session 1: Bioinformatics of NGS data (Chair: Klaus Jung)

12:00–12:40 Keynote lecture: Ivo Grosse
Phylotranscriptomic Hourglass Patterns of Animal and Plant Development and the Emergence of Biodiversity

12:40–13:00 Lea Schuh
Simulation and characterization of rare correlated gene expression using stochastic network modeling

13:00–13:20 Martin Treppner
Simulating Single-Cell RNA-Sequencing Data using Negative Binomial Deep Boltzmann Machines

13:20–13:40 Laura Jenniches
Multilevel modeling of HTS data

13:40–14:00 Michael Seifert
Survival differences and associated molecular signatures of DNMT3A-mutant acute myeloid leukemia patients

14:00–15:00 Coffee break & poster session

Session 2: Disease modelling (Chair: Ingmar Glauche)

15:00–15:40 Keynote lecture: Benjamin Werner
Quantitating somatic evolution in healthy and cancerous human tissues

15:40–16:00 Pietro Mascheroni
Cancer cells under pressure: How do glioma cells respond to mechanical forces?

16:00–16:20 Gang Zhao
The temporal pattern of insulin secretion influences liver function: explanation from a mathematical model

16:20–16:40 Michael Huttner
Lyra - containerized microservices for browsing shared biomedical data

16:40–17:00 Coffee break & poster session

Session 3: Disease modelling (Chair: Markus Scholz)

17:00–17:20 Masoud Hoore
Mathematical model of Amyloid beta fibrillization in Alzheimer's disease

17:20–17:40 Maria T. E. Prauße
Verifying hypotheses on pathogenic immune evasion in human whole blood by state-based virtual infection models

17:40–18:00 Shabaz Sultan
Modelling the Immune System's Interaction with the Tumour Microenvironment

18:00–18:20 Ingmar Glauche
Mathematical modeling of therapy response in CML patients reveals the potential for substantial dose reductions in many patients

19:30 Social event at Anders, das Pfannenrestaurant
Address: Am Magnitor 7, Braunschweig, `https://www.das-anders.de/`

Friday, March 8, 2019

Session 3: Machine Learning / Optimization in Computational and Systems Biology (Chair: Michael Altenbuchinger)

8:30–10:00 Educational lecture: Jan Hasenauer *Mechanistic models of large-scale biochemical reaction networks*

10:00–10:30 Coffee break

Session 4: Machine Learning / Optimization in Computational and Systems Biology (Chair: Michael Altenbuchinger)

10:30–10:50 Zahra Nasrollah
Learning the topology of latent signaling networks from high dimensional transcriptional intervention effects

10:50–11:10 Victor Greiff
Mining immune repertoires using machine learning and high-dimensional statistics

11:10–11:30 Darius Schweinoch
Temporal control of the RIG-I-dependent antiviral innate immune response

11:30–11:50 Hryhorii Chereda
Utilizing molecular networks in Convolutional Neural Networks on Graphs to predict the appearance of a metastatic event

11:50–12:10 Alexander Tille
Mathematical model of the factor H mediated self and non-self discrimination by the complement system

12:10–12:30 Ashar Ahmad
Stratifying PD patients by disease progression using advanced machine learning techniques

12:30–12:50: Poster award and workshop closing

2 Keynotes & educational lecture

2.1 Keynote 1: Phylotranscriptomic hourglass patterns of animal and plant development and the emergence of biodiversity

Alexander Gabel[1]*, Hajk-Georg Drost*[1,2]*, Marcel Quint*[3]*, and Ivo Grosse*[1,4]

[1]Institute of Computer Science, Martin Luther University Halle-Wittenberg, Germany
[2]Sainsbury Laboratory Cambridge, University of Cambridge, UK
[3]Institute of Agricultural and Nutritional Sciences, Martin Luther University Halle-Wittenberg, Germany
[4]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany

One surprising observation going back to pioneering works of Karl Ernst von Baer in 1828 and Ernst Haeckel in 1866 is that embryos of different animal species express on average evolutionarily young genes at the beginning of embryogenesis, evolutionarily old genes in mid-embryogenesis, and again evolutionarily young genes at the end of embryogenesis. This phylotranscriptomic hourglass pattern as well as the resulting morphological hourglass pattern that animals of different phyla look morphologically different at the beginning of embryogenesis, morphologically similar in mid-embryogenesis, and again morphologically different at the end of embryogenesis show that trancriptomic and morphologic biodiversity emerges in a nonlinear and even non-monotonic manner.

Focusing our attention on plants, which represent the second major kingdom in the tree of life that evolved embryogenesis, we have found that the phylotranscriptomic hourglass pattern also exists in plant embryogenesis. This observation is surprising as multicellularity and embryogenesis evolved independently in animals and plants and suggests the convergent evolution of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. Moreover, we have found that phylotranscriptomic hourglass patterns also exist in the two main transitions of post-embryonic plant development, germination and floral transition, suggesting the convergent evolution of phylotranscriptomic hourglass patterns in embryonic and post-embryonic plant development.

The origin of these phylotranscriptomic hourglass patterns has remained concealed, but here we find that not only the mean age of expressed genes changes in an hourglass-like manner, but the whole age distribution of expressed genes changes. When studying the entropy of these age distributions as functions of time, we find hourglass patterns that surprisingly are orders of magnitude more significant than the original phylotranscriptomic hourglass patterns of the mean, which might indicate that the phylotranscriptomic hourglass patterns of the entropy are more fundamental than, and possibly even the origin of, the original phylotranscriptomic hourglass patterns of animal and plant development.

2.2 Keynote 2: Quantitating somatic evolution in healthy and cancerous human tissues

Benjamin Werner[1]

[1] The Institute of Cancer Research (London, UK)

A human produces approximately a hundred thousand trillion cells throughout his lifetime, requiring him to copy 10^{25} base pairs of DNA. Also this process is very precise, it is not perfect and we would estimate that humans accumulate a hundred thousand trillion mutations throughout live. It appears to be a miracle how the human body is able to maintain functional integrity. Here we discuss some mechanisms that supress mutation accumulation in healthy human tissues and their failure in cancer. By combining evolutionary theory, mathematical and computational modelling with genomic sequencing, I show how we can measure somatic mutation rates per cell division in healthy and cancerous tissues. The same approach also allows us to measure clonal expansion rates and quantitating the strength of selection in individual human tumours. We find striking differences between individual tumours both in selection strength and clonal expansion rates. Both have implications for the estimated age of tumours and may inform treatment and prevention strategies. I furthermore will present some results of a clinical trail of targeted colon cancer treatment. I will show that a combination of serial cell free tumour DNA sampling in blood and a simple mathematical model allows us to follow resistance evolution within patients in real time. We hope this or similar approaches will support personally adjusted cancer treatment strategies.

2.3 Educational lecture: Mechanistic models of large-scale biochemical reaction networks

Jan Hasenauer[1]

[1] Universität Bonn, Germany

Mechanistic mathematical models are powerful tools in modern life sciences. Similar to experimental techniques, mechanistic models facilitate an assessment of biological processes and hypothesis testing. Furthermore, mechanistic models allow for the integrative assessment of multiple datasets as well as the prediction of latent variables and future experiments. To achieve this, the model structure has to be defined and the unknown model parameters have to be estimated from experimental data. These tasks are already challenging for small-scale biological processes and high-quality data. For large-scale biological processes and realistic data, inference is often still intractable.

In this tutorial, I will provide an overview over state-of-the-art approaches for mechanistic modelling of biochemical reaction network using ordinary differential equations. I will discuss model building, calibration and selection. Amongst other things, I will cover recent developments such as (i) scalable and efficient optimisation methods using adjoint sensitivity analysis and hierarchical formulations, (ii) robust estimation methods using alternative noise models, and (iii) uncertainty analysis and model reduction methods using profiling and sampling.

3 Oral presentations

3.1 Simulation and characterization of rare correlated gene expression using stochastic network modeling

Lea Schuh[1,2,3]*, Eric Sanford*[1]*, Benjamin L. Emert*[1]*, Cesar A. Vargas-Garcia*[4]*, Abhyudai Singh*[4]*, Carsten Marr*[2]*, Arjun Raj*[1]*, Yogesh Goyal*[1]

[1]Department of Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA
[2]Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg 85764, Germany
[3]Department of Mathematics, Technische Universität München, Garching bei München 85748, Germany
[4]Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716, USA

Drug resistance leads to a high relapse rate in melanoma patients. The resistant subpopulation of cells that survives drug treatment has been recently characterized by correlated overexpression of specific genes at the time of drug administration. Accordingly, this rare occurrence of correlated overexpression leads to long-tailed gene expression distributions at the population level where the tails exhibit the drug-surviving cells. Correlated overexpression arises and disappears over time such that cells exhibiting overexpression of a subset of genes are thought to be in a transient-resistant state. However, we lack a mechanistic model for how these cells enter and leave this transient-resistant state. More abstractly, how can we explain the occurrence of rare correlated overexpression and its resulting long-tailed distributions at the population level? What gene regulatory network architectures and kinetic parameters give rise to such a rare behavior? To address these questions, we developed a stochastic model to simulate gene expression dynamics that is able to mimic correlated overexpression. We implemented this model for the subset of 122 weakly-connected, non-isomorphic, symmetric network architectures of varying size and simulated over a million timepoints for 100 different parameter sets using Gillespie's next reaction method. The resulting simulations were classified into showing rare behavior or not. Using this classification, we optimized a decision tree to identify parameters that are important for rare correlated overexpression. We determined two parameters to be most important for producing rare behavior - both of which modulate the strength of gene expression regulation between two genes. We also find that with an increasing connectivity of the gene regulatory networks the system is less able to disintegrate correlated overexpression and hence is less able to produce rare behavior. Together, our approach provides insights into the formation of rare behavior as seen in drug resistance formation in melanoma.

3.2 Simulating Single-Cell RNA-Sequencing Data using Negative Binomial Deep Boltzmann Machines

Martin Treppner[1]

[1] Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Stefan-Meier-Str. 26, D-79104 Freiburg

With the rise of single-cell RNA-sequencing technologies, evaluation of downstream analysis techniques becomes important. Thus, data sets with known underlying structure from realistic single-cell RNA-sequencing simulations are needed for evaluating methods. We developed a corresponding approach based on a generative model, more precisely deep Boltzmann machines (DBMs). For this, we adapted the DBM to estimate the mean parameters of the negative binomial distribution while estimating the dispersion parameter using Fisher scoring. Accordingly, the model allows for the generation of synthetic single-cell RNA-sequencing data.

For illustration, this approach is applied to a subset of the 1.3 million mouse brain cells data set for learning the particularities of the data. Conditional sampling with artificial data is used to investigate the performance of the DBM, i.e. we evaluate whether the DBM can generate plausible gene expression patterns for a group of genes, given the expression level of another gene group. This also has a practical application, as it allows to pose biological questions. For example, we consider a setting in which two genes A and B are exclusively expressed in conjunction. Here, conditional sampling can be used by fixing gene A at a high expression value, expecting that gene B will also be highly expressed in the generated samples. Following this investigation of DBM properties, we apply DBMs for the purpose of generating artificial data based on original mouse brain data.

Generated samples are compared to the original data set by clustering. In doing so, we show that the generated samples resemble the true observations since they exhibit similar patterns.

To conclude, DBMs enable modeling of complex joint gene expression distributions and are thus a valuable method for simulating realistic single-cell RNA-sequencing data.

3.3 Multilevel modeling of HTS data

Laura Jenniches[1]

[1] HIRI, Würzburg, Germany

In recent years, decreasing costs of high-throughput sequencing (HTS) experiments have lead to increasingly complex experimental designs. Standard RNA-seq analysis tools often have difficulties capturing key features of multi-factorial datasets. As a result, these datasets are usually analyzed on a case-by-case basis. Our aim is the development of a flexible toolkit for Bayesian inference and multilevel modeling of HTS data using the Stan platform for high-performance statistical computation. Stan separates model description from implementation which allows for easy distribution and extension of the models. We will briefly present the software tool Stan and outline its advantages in the analysis of multilevel datasets. As a case study, we use a time course RNA stability experiment based on rifampicin treatment and RNA-seq to investigate the targets of a recently discovered global RNA-binding protein, ProQ, in the bacterial pathogen Salmonella Typhimurium. While ProQ's activities remain largely uncharacterized, it is thought to operate in part by protecting sRNA and mRNA from decay. We can quantify this protection activity of ProQ by comparing the decay rates of RNA in the wild-type, a proQ deletion mutant, and a plasmid complementation. As replication is limited, we use hierarchical models to share information across genomic loci. Additionally, the flexibility of Stan allows us to describe non-linear effects with additional parametric dependences and non-parametric models. We will also discuss sampling efficiency, which is greatly influenced by the choice of priors and model implementation. Finally, the results from the statistical modeling process are integrated with CLIP-seq data on ProQ and other RNA-binding proteins (CspCE, CsrA, Hfq) in a network approach to identify pathways regulated by ProQ.

3.4 Survival differences and associated molecular signatures of DNMT3A-mutant acute myeloid leukemia patients

Michael Seifert[1]

[1] IMB TU Dresden, Fetscherstr. 74, D-01307 Dresden

Acute myeloid leukemia (AML) is a very heterogeneous and highly malignant blood cancer. Mutations of the DNA methyltransferase DNMT3A are among the most frequent recurrent genetic lesions in AML. The majority of DNMT3A-mutant AML patients shows fast relapse and poor overall survival, but also patients with long survival or long-term remission have been reported. Underlying molecular signatures and mechanisms that contribute to these survival differences are only poorly understood and have not been studied in detail so far. We applied hierarchical clustering to somatic gene mutation profiles of 51 DNMT3A-mutant patients from the The Cancer Genome Atlas (TCGA) AML cohort revealing two patient subgroups with profound differences in overall survival. We determined associated molecular signatures that distinguish both subgroups at the level of protein-coding genes, miRNAs, and signaling pathways. We further learned gene regulatory networks to predict potential major regulators associated with survival differences between short- and long-lived DNMT3A-mutant AML patients. We found several genes and miRNAs with known roles in AML pathogenesis, but also interesting novel candidates involved in the regulation of hematopoiesis, cell cycle, cell differentiation, and immunity that may contribute to the observed survival differences of both subgroups. Our study contributes to the characterization of molecular alterations associated with survival differences of DNMT3A-mutant patients and could trigger additional studies to develop robust molecular markers for a better stratification of AML patients with DNMT3A mutations.

3.5 Cancer cells under pressure: How do glioma cells respond to mechanical forces?

Pietro Mascheroni[1]

[1] Helmholtz Centre for Infection Research / BRICS, Rebenring 56, 38106 Braunschweig, Germany Germany

The microenvironment of solid tumors is an important player in disease progression. Particularly, the behavior of cancer cells can be modulated by both chemical and mechanical cues. Mechanical forces act at both microscopic and macroscopic levels, triggering gene expression patterns in cells as well as altering tissue perfusion. As a matter of fact, such forces are recognized to affect the response of patients to therapies, even though their exact mechanisms of action remain still elusive. Here we investigate the response of glioma cells to mechanical compression combining mathematical modeling and in vitro experiments. We show that both cell proliferation and migration are significantly affected by an external mechanical pressure, and provide a quantification of these effects in different glioma cell lines.

3.6 The temporal pattern of insulin secretion influences liver function: explanation from a mathematical model

Gang Zhao

Physiological insulin secretion exhibits various temporal patterns, the dysregulation of which is involved in diabetes development. We integrated current knowledge of hepatic insulin signaling into a mathematical model and investigated the impact of the temporal patterns of insulin. Two sets of experimental data allowed the identification of a minimal structure of the model. Although the parameters of the minimal model were not fully identifiable, the model generated robust predictions. Model predictions explained the impact of the temporal patterns and contributed to the understanding of selective hepatic insulin resistance, which is a long-standing paradox in the field.

3.7 Lyra - containerized microservices for browsing shared biomedical data

Michael Huttner[1]

[1] Institute of functional genomics, Regensburg, Germany

Research papers in the biomedical field come with large and complex data sets that are shared with the scientific community as unstructured data files via public data repositories. Examples are sequencing, microarray, and mass spectroscopy data. The papers discuss and visualize only a small part of the data, the part that is in its research focus. For labs with similar but not identical research interests different parts of the data might be important. They can thus download the full data, preprocess it, integrate it with data from other publications and browse those parts that they are most interested in. This requires substantial work as well as programming and analysis expertise that only few biological labs have on board. In contrast, providing access to published data over web browsers makes all data visible, allows for easy interaction with it, and lowers the barrier to working with data from others.

We have developed Lyra, a collection of microservices that allows labs to make their data easily browsable over the web. Currently we provide tools for

(a) insertion of genomic, proteomic, transcriptomic and metabolomic data,

(b) cross linking data from different publications via automatic conversion of over 200 molecular identifier types,

(c) fast data access and search over a JSON API, and

(d) dynamic and interactive visualization in the users web browser.

3.8 Mathematical model of Amyloid beta fibrillization in Alzheimer's disease

Masoud Hoore[1]

[1] SIMM, Rebenring 56, 38106 Braunschweig, Germany

Circadian rhythm and Alzheimer's disease (AD) are known to be linked together. Disturbances in sleep cycle are early symptoms detected in AD patients. Whether the relation between the circadian rhythm of the brain activity and AD is a correlation or causality is unknown. In this study, a mathematical model of Amyloid beta production and fibrillization is derived in order to investigate how the circadian rhythm may play a role in AD. Our in-silico results show that a rise in the brain normal activity is able to initiate AD by triggering Amyloid beta fibrillization. The sensitivity of the system to different parameters is analyzed in order to suggest the best prevention/therapeutic strategies.

3.9 Verifying hypotheses on pathogenic immune evasion in human whole blood by state-based virtual infection models

Maria T. E. Prauße[1,2], *Teresa Lehnert*[1,3], *Sandra Timme*[1], *Kerstin Hünniger*[4,5], *Ines Leonhardt*[3,5], *Oliver Kurzai*[3,4,5] *and Marc Thilo Figge*[1,2,3]

[1] Applied Systems Biology, Leibniz Institute for Natural Product Research and Infection Biology - Hans Knöll Institute, Jena, Germany
[2] Faculty of Biological Sciences, Friedrich Schiller University, Jena, Germany
[3] Center for Sepsis Control and Care (CSCC), Jena University Hospital, Jena, Germany
[4] Fungal Septomics, Leibniz Institute for Natural Product Research and Infection Biology - Hans Knöll Institute (HKI), Jena, Germany
[5] Institute of Hygiene and Microbiology, University of Würzburg, Germany

The immune system protects us constantly against the harm of pathogens which we inhale on a daily basis or which colonize our skin. If the immune response is impaired the risk of pathogens breaching the outer surfaces and reaching the blood stream increases. This oftentimes causes life-threatening systemic infections, which attribute to high morbidity and mortality. Pathogens which are most commonly associated to such severe diseases are the bacterial pathogen Staphylococcus aureus, and the fungal pathogens Candida albicans and Candida glabrata. Together, C. albicans and C. glabrata are responsible for about 70% of all systemic candidiasis cases. As the early clearance of these pathogens from the blood stream is of fundamental importance, the innate immune response in human whole-blood was investigated in a systems biology study combining experimental whole-blood infection assays and state-based modelling (1,2). In this study, we found a population of pathogens that were not cleared by immune cells such as monocytes and polymorphonuclear neutrophils (PMN), but evade the immune response in whole blood. So far, the underlying mechanism causing these immune evasive pathogens could not be identified.

By the means of biomathematical models and the acquired data from whole-blood assays we tested biologically reasonable hypotheses. The established state-based model of whole-blood infection comprises three different immune-evasion mechanisms: The spontaneous mechanism defines the immune-evasion rate with a constant probability (1,2). The second immune-evasion mechanism is dependent on PMN which phagocytose microbes and it presents a time-dependent probability (3). This is biologically justified since PMN secrete proteins upon phagocytosing a microbe for the first time. These proteins may cause unknown effects and could alter the outer surface of the microbe and therefore, may decrease the rate of successful clearance from the blood. The third immune-evasion mechanism entails that a subpopulation of immune-evasive cells already exists prior to the infection. These may arise from the heterogeneity within a microbial population, which developed due to unknown reasons before infecting the host. Furthermore, we test combinations of these immune-evasion mechanisms such as a model combining the pre-existing immune evasion with the spontaneous immune evasion and the PMN-mediated immune evasion, respectively. We fit these five models to the data from whole-blood assays which were conducted with blood samples from healthy donors and were then infected with C. albicans, C. glabrata and S. aureus.

We found that all five immune-evasion models are in agreement with the experimental data from whole-blood infection assays. Nevertheless, we could detect significant differences in the simulation results, like the time course of extracellular killing by antimicrobial peptides and the predicted ratios of alive and killed immune-evasive cells. Our results provide evidence for the existence of pre-existing immune-evasive pathogenic cells. Additionally, we fitted the models to whole-blood infection assays performed with samples from patients who underwent cardiac surgery with extracorporeal circulation which induces a strong inflammatory stimulus. Our results indicate that under the condition of activated immune response, the hypothesis on pre-existing immune-evasive cells alone might not hold true, but a combination of pre-existing evasive cells with the spontaneous or PMN-mediated evasion mechanism is still possible.

We conclude that the hypothesis on pre-existing immune-evasive cells alone is not reasonable when considering all available data sets. Furthermore, we are able to propose new experimental measurements that could provide further insights into the underlying mechanism of pathogenic immune evasion in human blood.

[1] Hünniger and Lehnert et al. (2014) PLoS Comput Biol. [2] Lehnert and Timme et al. (2015) Frontiers in Microbiology. [3] Prauße et al. (2018) Frontiers in Immunology.

3.10 Modelling the Immune System's Interaction with the Tumour Microenvironment

Shabaz Sultan[1]

[1]Department of Tumor Immunology, Radboud Institute for Molecular Life Sciences, Nijmegen, the Netherlands

Antitumoral immune responses are traditionally considered ineffective in cancer patients, but therapies that boost the host's immune system have recently achieved considerable success. There is evidence that tumour infiltration by different types of immune cells is a relevant factor for immunotherapy success. For example, higher infiltration by cytotoxic T cells and lower infiltration by suppressor T cells often correlate with better patient outcome.

We use high resolution spatially resolved image data combined with validated machine learning methodology to identify immune cells and their associated phenotypes within the spatial context of a tumour. We hypothesize that the spatial distribution of different cell types in this context are an emergent property of the underlying dynamic behaviour of immune cells in this environment. We want to investigate if properties that describe this dynamic behaviour are more predictive than cell distribution statistics directly observable from static image data.

To this end we use a custom implementation of the cellular Potts model, which can simulate fine grained cellular behaviour for large, dense cellular environments. We explore the space of dynamic cellular behaviour using this simulation environment and investigate which parameters lead to cell distributions best matching our image derived empirical data.

Our goal is to use this derived dynamic behaviour in a causal model to make predictions on patient prognoses and response to immunotherapy treatment. We have access to image data of the tumour microenvironment of patients treated with different types of immunotherapy and with different levels of response to treatment. We aim to map the relation between dynamic cell behaviour, and the progression of a tumour and efficacy of immunotherapy treatment. This will allow us to better understand why treatment works for certain patients and not others, and make better predictions on patient outcome.

3.11 Mathematical modeling of therapy response in CML patients reveals the potential for substantial dose reductions in many patients

Ingmar Glauche[1]

[1]Technische Universität Dresden, Fetscherstr. 74, D-01307 Dresden, Germany

Continuing tyrosine kinase inhibitor (TKI)-mediated targeting of the BCR-ABL1 oncoprotein is the standard therapy for chronic myeloid leukemia (CML) and allows for a sustained disease control in the majority of patients. While therapy cessation for patients appeared to be a safe option for about half of those patients with optimal response, there has been no systematic assessment of long-term TKI dose de-escalation. We use a mathematical model to analyze and consistently describe biphasic treatment responses from TKI-treated patients from two independent clinical phase III trials. Scale estimates reveal that drug efficiency determines the initial response while the long-term behavior is limited by the rare activation of leukemic stem cells. We use this mathematical framework to investigate the influence of different dosing regimens on the treatment outcome. We provide strong evidence suggesting that TKI dose de-escalation (at least 50%) does not lead to a reduction of long-term treatment efficiency for most patients who have already achieved sustained remission, and that this approach maintains the secondary decline of BCR-ABL1 levels. We demonstrate that continuous BCR-ABL1 monitoring provides patient-specific predictions of an optimal reduced dose that does not decrease the anti-leukemic effect on residual leukemic stem cells. Our results are consistent with the interim results of the DESTINY trial and provide predictions that can be clinically tested. Our results suggest that dose-halving should be considered as a long-term treatment option for CML patients with good response under continuing maintenance therapy with TKIs. We emphasize the clinical potential of this approach to reduce treatment-related side-effects and treatment costs.

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

Zahra Nasrollah[1]*, Achim Tresch, Holger Fröhlich*

[1] Institute of Medical Statistics and Computational Biology, Faculty of Medicine, University of Cologne, Germany

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 – Markowetz 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 reconstruction of several known elementary network motifs over a wide range of scenarios (e.g. different number of observations, length of time series). Moreover, we evaluated our method on synthetic data generated by differential equation systems.

Application of our method to breast cancer proteomics data from the DREAM8 challenge infers most of the known interactions in causal signaling pathways. The inferred network shows considerable overlap with the corresponding protein-protein network reported in STRING.

In another context, we aimed to detect interactions between 20 proteins involved in EGFR signaling which were associated with drug sensitivity in Non Small Cell Lung Cancer. We compare our results against the STRING and HIPPIE database, and against a reconstruction of this network by Paurush et al. using NEMs.

3.13 Mining immune repertoires using machine learning and high-dimensional statistics

Victor Greiff[1]

[1]University of Oslo, Department of Immunology, Nydalen, 0424 OSLO, Norway

The interrogation of immune repertoires is of high relevance for understanding the adaptive immune response in (autoimmune) disease and infection. We and others have recently shown that immune repertoires are much more predictable than previously thought, which is of incredible importance for the precise manipulation and prediction of adaptive immunity. However, as of yet, we lack the computational methods that enable us to understand the construction rules according to which immune repertoires are assembled. Specifically, (i) on the repertoire-level, we lack mathematical methods to measure repertoire similarity across individuals. We will show preliminary results of a network approach that allows quantifying immune repertoire similarity based on the multidimensional combination of immunological features that cover the full dimensionality of the repertoire space. Our networks-based similarity definition explicitly allows many-to-many repertoire comparisons, which implies that we can embed an individual's repertoire within a population's similarity space. (ii) On the sequence level, we lack machine learning methods that enable modeling immune-receptor-antigen binding in the absence of extensive 3D data. Therefore, we trained a collection of machine learning models on large synthetic antibody sequence datasets with varying pattern complexities, which simulate antibody-antigen binding. We tested the models in terms of pattern prediction accuracy and recovery. Taken together, we will show preliminary data that pave the way towards understanding the molecular construction rules of adaptive immunity on the repertoire and single sequence level.

3.14 Temporal control of the RIG-I-dependent antiviral innate immune response

Darius Schweinoch[1], *Jamie Frankish*[2], *Carola Sparn*[2], *Marco Binder*[2], *Lars Kaderali*[1]

[1] Institute for Bioinformatics, University Medicine Greifswald, 17475 Greifswald, Germany
[2] Division Virus-associated carcinogenesis (F170), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

Upon viral infection, antiviral innate immunity pathways induce an antiviral state of host cells to interfere with viral replication and spread. In this system, RIG-I like receptors play a crucial role by sensing viral RNA within the cytoplasm and triggering the production and secretion of interferons (IFNs) and other cytokines. By autocrine signaling, IFN in turn activates the JAK-STAT pathway, leading to massive upregulation of IFN-stimulated genes (ISGs). We developed a model of the RIG-I pathway activation after stimulation with double-stranded RNA and successfully linked it to an existing model of JAK-STAT signaling. Our model describes the activation and deactivation of key pathway components and allows the identification of sensitive steps in the regulation of the antiviral response by directly linking them to the expression of ISGs. In a next step, we are going to apply our model to analyze the dynamic consequences of perturbations of the system imposed by virus-encoded antagonists with known or unknown targets.

3.15 Utilizing molecular networks in Convolutional Neural Networks on Graphs to predict the appearance of a metastatic event

Hryhorii Chereda[1]*, Annalen Bleckmann, Frank Kramer, Andreas Leha, Tim Beißbarth*

[1]University Medical Center Göttingen, Institute of Medical Bioinformatics, 37077 Göttingen, Germany

Gene expression data is commonly available in cancer research and depicts a snapshot of the current status of specific tumor tissues. This high-dimensional data can be analyzed for diagnoses, prognoses, and to suggest treatment plans. Molecular networks (e.g. biological pathways) are represented by graphs detailing interactions between molecules. Gene expression data can be assigned to the vertices of these graphs, and the edges can depict interactions, regulations and signal flow.

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 already shown prominent results in different applications, such as visual object recognition, object detection and speech recognition. Furthermore, CNN's have been applied to bioinformatic challenges such as predicting the effects of mutations in non-coding DNA on gene expression and disease. Recently, CNN's have been extended to include graph-structured data. We are planning to map gene-expression data to the vertices of molecular networks and feed this graph-structured data into CNNs in order to classify patients.

In our work, we used the clinical and gene expression data of breast cancer patients in order to predict the appearance of the metastatic events. The gene expression data were structured by utilizing two molecular networks as prior knowledge. As a first approach, we used the WNT signalling pathway that plays a crucial role in breast cancer progression. This pathway is represented as a directed graph. As the second approach, we incorporated the protein-protein interaction network from HPRD that represents undirected interactions between two proteins. We applied two different CNN methods that are capable to embed undirected and directed connections of graph-structured data.

Our research aims to address the question if the CNNs on graphs are able to provide valuable classification improvements for the prediction of metastatic events utilizing the prior knowledge.

3.16 Mathematical model of the factor H mediated self and non-self discrimination by the complement system

Alexander Tille[1]

[1]Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, Germany

The complement system is part of the innate immune system and plays an important role in host defense against pathogenic infections. By mediating immunological and inflammatory processes it plays a key role in coordinating innate and adaptive immune response. Its main task is the recognition and subsequent opsonization of foreign particles or dysfunctional cells.

In this study, we focus on the alternative pathway of the human complement system. The alternative pathway is activated spontaneously, which leads to a basal level of active complement molecules. Thus a tight regulation mechanism is needed to protect the body's own cells (self-cells) from opsonization. The major regulator of the alternative pathway is the protein factor H. It acts in the fluid phase and can attach to cell surfaces where it controls complement activation effectively. Besides the body's own cells, pathogens like C. albicans also have acquired the ability to bind factor H and thus escape opsonization and cause severe infections.

In order to understand the opsonization process better, we developed a mathematical model of the alternative pathway using ordinary differential equation for surface-bound molecules and partial differential equations for the concentration profiles of fluid phase molecules around a cell. The model focuses on the most important components of the complement cascade: C3b in the fluid phase and on the cell surface as well as inactivated C3b on the cell surface. The other components of the complement system are combined in effective rates that represent the dynamics of the formation of several intermediate products of the cascade. Combining the system into three differential equations allowed us to summarize a large set of unknown parameter into a small set of effective rates, which preserve the dynamics of the complete system.

Using steady state analysis we investigated driving processes of the complement activation and regulation on the cell surface. The model enables a clear distinction between pathogens and self-cells. Based on these results we propose treatment strategies to enhance the opsonization of pathogens while protecting the body's own cells.

3.17 Stratifying PD patients by disease progression using advanced machine learning techniques

Ashar Ahmad[1]

[1]University of Bonn, Endenicher Allee 19C, 53115 Bonn, Germany

Parkinson Disease (PD) is a neurodegenerative disorder known to affect over 10 million people worldwide. Patients suffering from PD exhibit a broad range of signs and symptoms ranging from cognitive decline to impaired motor activities. The progression pattern of the PD patients tends to be of a highly heterogeneous nature. Stratifying PD patients would allow us to identify patient subgroups more likely to respond to clinical trials as well as might help us in understanding the underlying features which lead to faster/ slower progression. This in turn will provide useful background knowledge for guiding future drug discovery for PD.

In the present work we use data from the Parkinson's Progression Markers Initiative (PPMI) which is a popular observational clinical study to comprehensively evaluate PD patients in terms of multi modal data (imaging, biologic sampling and clinical and behavioural assessments). We use this data set to build a statistical machine learning model which can identify distinct patient sub-groups based on their multi-modal longitudinal progression patterns. We employ unsupervised functional data clustering to discover clusters and subsequently make use of these discovered cluster labels to build a classification model based on baseline data, along with genomic and CSF biomarker profiles to predict the future progression pattern. More specifically, ensemble based supervised learning approaches are used with different data integration techniques to build our final classification model. We measure the prediction performance of our classification model on the discovery cohort with a cross-validation scheme.

As a final validation step we plan to use the Parkinson's Disease Biomarkers Program (PDBP) data as external validation data set. The goal here would be to predict and stratify PD patients based on the previously developed statistical model and verify these predictions by looking at the longitudinal progression of the different clinical and behavioural assessments across time.

4 Poster presentations

4.1 Mathematical modeling of Alzheimer's disease

Masoud Hoore[1]

[1] Brics & SIMM HZI, Braunschweig, Germany

Sleepiness during the day and wakefulness at night are the very early symptoms of Alzheimer's Disease (AD). A mathematical model for the Amyloid beta fibrillization in the brain is proposed to explain the relation between neural activity and AD. It is found that the disturbances in the sleep cycle alter the rate of Amyloid beta production and cause their deposition if not treated. It is deduced from the model that the regulatory factors, such as microglia, astrocytes, and Amyloid beta efflux, play important roles in avoiding the disease. Possible treatment strategies can also be suggested from the mathematical model.

4.2 Transcriptional noise in Pseudomonas aeruginosa – From population to single cells

Elisabeth Vatareck[1]

[1] Helmholtz Centre for Infection Research, Inhoffenstr. 7, 38124 Braunschweig, Germany

Transcription and translation are noisy processes. In consequence, even isogenic populations of bacteria exhibit a remarkable variability in gene expression. The resulting phenotypic heterogeneity provides the population with the necessary flexibility to quickly adapt to changing environments.

Pseudomonas aeruginosa is a ubiquitous bacterium that is found in environmental habitats as well as in clinics, being one of the most frequent pathogens for nosocomial infections. Due to its extreme metabolic versatility, the pathogen is able to adapt to various environments, thrive in an animate niche and escape host defense. How phenotypic heterogeneity contributes to its adaptive capacity has so far not been studied yet.

Here, we used a set of transcriptomes from 258 P. aeruginosa strains to identify genes that show high variability between replicated samples. We predicted 292 genes with noisy transcription, among them the whole denitrification pathway and siderophore genes.

The glp locus, responsible for glycerol metabolism, showed the highest transcriptional variability in RNAseq data with a remarkable bimodal distribution. We therefor measured the activity of glp promoters in single cells using flow cytometry. Combining both analyses we could demonstrate that in case of the glp locus transcriptional variability between populations is a result of bimodal transcription within the population. The flexibility of glp activity might enable P. aeruginosa to quickly adapt to the lung environment in which glycerol is a major nutrient.

We demonstrated that transcriptional variability within a population can be predicted from replicated gene expression data. The identified heterogenic traits might play an important role for adaptation and virulence. Further analyses will provide more insights into the mechanisms controlling transcriptional variability in P. aeruginosa and its role for the pathogens capability to thrive in changing environments.

4.3 Evaluation of Computational Pipelines for Meta-analysis of ChIP-seq Data

Ihsan Muchsin[1], *Moritz Kohls, Klaus Jung*

[1] Institute for Animal Breeding and Genetics, University of Veterinary Medicine Hannover, Hannover, Germany

BACKGROUND:
Chromatin ImmunoPrecipitation followed by sequencing (ChIP-seq) is an experimental method to analyse protein interactions with DNA in a high-throughput manner (Park, 2009). Sequencing experiments often involve only a small number of samples. The level of evidence for individual experiment is usually limited and error-prone. Since the data of many experiments and studies are deposited in public repositories such as Gene Expression Omnibus (Barrett et al., 2012) and ArrayExpress (Kolesnikov et al., 2014) the option for meta-analyses to increase the level of scientific evidence exists. Chen et.al have implemented a pipeline for meta-analysis of ChIP-seq data using normalized reads (raw data) merging. Additionally, quantile normalization followed by p-values combination (results merging) has been also studied (Chen, 2015). This study evaluates and refines these existing computational pipelines.

METHOD:
ChIP-seq data are obtained from computer simulation and public repositories, i.e., ENCODE (The ENCODE Project Consortium, 2012) and ArrayExpress. Experimental data from different types of transcription factor, which have distinct binding properties, are selected. In addition, experimental data are manipulated by introducing artificial bindings. After preprocessing, ChIP-seq data lists genomic regions together with a value for the peak intensity reflecting the binding activity of the sample compared to a control. The core part of meta-analysis is to find a common set of binding regions from the different studies. This can be reflected by the union of the regions from all studies or the overlapping regions between the different studies. In addition, a measure of significance (p-value) and other scoring functions are assigned to the set of common peaks. Fisher's method (Fisher, 1925) and Stouffer's Z-score (Stouffer, 1949) method are commonly used for this purpose. In addition, this project also compares raw data merging with results merging.

AIMS:
Various computational pipelines for meta-analysis of ChIP-seq data are compared. The error rates for false positive binding as well as the power for detecting a true binding site are calculated to judge the performance of each pipeline. In addition, for the experimental data, the motif enrichment analysis are performed to discern pipeline that gives the optimal enrichment.

References:

- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., . . . Soboleva, A. (2012, nov). NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Research, 41(D1), D991–D995. doi: 10.1093/nar/gks1193
- Chen, R. (2015). Meta-analysis framework for peak calling by combining multiple ChIP-seq algorithms and gene clustering by combining multiple transcriptomic studies (Doctoral dissertation, University of Pittsburgh).
- Chen, Y., Meyer, C. A., Liu, T., Li, W., Liu, J. S., & Liu, X. S. (2011). MM-ChIP enables integrative analysis of cross-platform and between-laboratory ChIP-chip or ChIP-seq data. Genome biology, 12(2), R11.
- ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414), 57.
- Fisher, R. A. (1992). Statistical methods for research workers. In Breakthroughs in statistics (pp. 66-70). Springer, New York, NY.
- Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y. A., Williams, E., ... & Megy, K. (2014). ArrayExpress update—simplifying data submissions. Nucleic acids research, 43(D1), D1113-D1116.
- Park, P. J. (2009). ChIP–seq: advantages and challenges of a maturing technology. Nature reviews genetics, 10(10), 669.
- Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., & Williams Jr, R. M. (1949). The American soldier: Adjustment during army life.(Studies in social psychology in World War II), Vol. 1.

4.4 A Bayesian Network Approach for Analyzing Epidemiological and Gene Expression Data on Non-Alcoholic Fatty Liver Disease

Ann-Kristin Becker[1]

[1] Institute of Bioinformatics, University Medicine Greifswald, Greifswald, Germany

To get a better understanding of the complex processes involved in non-alcoholic fatty liver disease (NAFLD) and the progression of the disease, a combination of knowledge about conditions inside and outside the liver, which are driving the disease, is necessary. To this end, we combine different epidemiological data sets on NAFLD by an integrative systems biology-based analysis using Bayesian networks (BNs).

Bayesian networks are a type of statistical model that represents a set of variables and their conditional dependencies via compact diagrams. Thus, on the one hand, they provide a consistent mathematical basis for reasoning under uncertainty and, on the other hand, they offer an intuitive and efficient way of representing complex dependency structures. As such, they are suitable for identifying risk factors and distinguishing between direct and indirect influences, and they allow an integrative analysis fitting the multifactorial pathogenesis of NAFLD.

Patient data from SHIP (Study of Health in Pomerania) provide various clinical, socioeconomic and molecular data as well as biomaterials from a relevant number of fatty liver patients. These are supplemented by study data from Gani_Med (Greifswald Approach to Individualized Medicine) as well as tissue samples available in LiSyM (Liver Systems Medicine). In this way, influencing factors and associations inside and outside the liver can be analyzed.

Several adaptations are necessary to efficiently learn Bayesian Network structures from those large epidemiological data sets. First, methods for imputing missing values and appropriate discretization methods are adapted to allow the integration of the very different data types (such as laboratory parameters, pre-existing conditions, socioeconomic parameters, gene expression data) as well as to ensure robust results based on the diverse and incomplete study data.

Moreover, regularization methods for Bayesian networks are developed, which enable the integration of the very high-dimensional molecular data. Apart from this, we include additional biological knowledge from pathway databases, e.g. information about signal transduction pathways, to obtain more robust results for rather low case numbers compared to the number of variables.

4.5 Approaches for generation of artificial sequencing reads for simulation purposes

Moritz Kohls[1]

[1] Institute for Animal Breeding and Genetics, University of veterinary medicine of Hannover, Bünteweg 17p, 30559 Hannover, Germany

BACKGROUND:
Next-generation sequencing (NGS) is regularly used to identify viral sequences in a DNA sample of an infected host (Wang et al., 2013; Hunt et al., 2015; Scheuch et al., 2015). One part of most bioinformatics pipelines is to map sequencing reads or reads assembled to larger contigs to a database of known virus genomes and – if available – to the reference genome of the host. Due to similarities between viruses, mutations, mapping errors or conserved sequences the resulting virus lists can contain false positive and false negative findings. Very few approaches have been implemented to assess the error rates from bioinformatics virus detection pipelines. As one approach, Kruppa et al. have implemented a decoy database including false virus sequences. Additionally, this approach includes false sequences into the fastq-file from the sample. We want to refine the existing decoy strategy to estimate false findings in the bioinformatics virus detection pipeline.

METHODS:
To be able to assess the error rates from bioinformatics virus detection pipelines, we implement a decoy database including false virus sequences to the artificial reference genome and false reads to the fastq-file. In order to create realistic false sequences, main characteristics of the sample sequences should be maintained. For example, the uShuffle algorithm shuffles the sequences while maintaining the kmer-distribution of the complete original sequence. However, the local kmer-distribution depends on the location in the genome which can lead to CpG islands and many other changes in the local kmer-distribution. Hence we create a new algorithm based on inhomogeneous Markov chains which considers this effect and maintains the local kmer-distribution approximately. To build the decoy database, we concatenate the true reference genome with the false reference sequences and map the true and false reads to the concatenated reference genome with Bowtie2.

AIMS:
We count how often a true or false read is mapped to a true or false reference sequence and calculate the false discovery rate as a measure for the quality of the decoy database. We compare different algorithms on different parameters to find out which algorithm in combination with which parameters gives the best result so that we can reduce the risk to detect false discovery viruses in a bioinformatics pipeline.

References:

- Hunt, M., Gall, A., Ong, S. H., Brener, J., Ferns, B., Goulder, P., ... & Otto, T. D. (2015). IVA: accurate de novo assembly of RNA virus genomes. Bioinformatics, 31(14), 2374-2376.
- Kruppa J, Jo WK, van der Vries E, Ludlow M, Osterhaus A, Baumgärtner W and Jung K: Virus detection in high-throughput sequencing data without a reference genome of the host. Submitted manuscript under revision.
- Scheuch, M., Höper, D., & Beer, M. (2015). RIEMS: a software pipeline for sensitive and comprehensive taxonomic classification of reads from metagenomics datasets. BMC bioinformatics, 16(1), 69.
- Wang, Q., Jia, P., & Zhao, Z. (2013). VirusFinder: software for efficient and accurate detection of viruses and their integration sites in host genomes through next generation sequencing data. PloS one, 8(5), e64465.

4.6 Role of bacterial motility during host-pathogen interaction

Matthias Preusse[1], *Sebastian Felgner*[2], *Susanne Häussler*[1,2]

[1] Helmholtz Centre for Infection Research, Molecular Bacteriology, Braunschweig, Germany
[2] Twincore, Molecular Bacteriology, Hannover, Germany

Pseudomonas aeruginosa is an opportunistic pathogen that plays a dominant role as the causative agent of severe acute as well as chronic infections. During an infection, the pathogenic bacteria cause symptoms due to their high rate of reproduction but also due to tissue invasion. This causes an immune response, resulting in common disease associated symptoms. Thus to understand bacterial pathogenicity knowledge on the first host-pathogen contact is crucial. Recent studies demonstrated the importance of flagella and flagellar-driven motility in this process. However, the specific crosstalk between bacteria and host remained elusive.

In this study, we employed a dual-sequencing approach in a macrophage invasion assay. Pseudomonas strains with impaired motility or lack of flagella revealed an altered expression of genes involved in the biosynthesis of LPS, pyoverdine and spermidine. Importantly, these genes were only differentially expressed during contact with the eukaryotic host indicating a significant role during host-pathogen interaction. Furthermore, the altered gene expression increased the immunogenicity (i.e. TNF-α, IL-6) of the flagella variants. Our results demonstrate how single bacterial gene deletions/modifications alter the host-pathogen interaction and host gene expression patterns that affect e.g. phagocytosis via an impaired PI3k/Act activity in macrophages.

In summary, our results highlight the importance of flagella synthesis and motility during host-pathogen interaction and demonstrate the necessity to understand the effect of bacterial genotypes and phenotypes during host contact.

4.7 DMD Method outperforms Sigma Point methods and sampling based methods

Dantong Wang[1]

[1]Helmholtz Zentrum, München, Bahnhof str. 22, 85375, Neufahrn bei Freising, Germany

In order to get comprehensive understanding of cell pathways, mechanistic models are widely used tools. In these models, parameter estimation is a crucial optimization problem, which needs to be solved with the help of experimental data. However, because of the variance from cell to cell, there could be a large difference between parameter values in different cells even with the same cell type. Therefore, mixed effects modeling method is commonly used for the situation, where parameters are not constants, but distributions. These parameters contain two different effects, one is fixed, meaning that this part is the same for all cells; the other one is random, meaning that this part is different from cell to cell and is usually considered as Gaussian distribution.

Considering the mixed effects model mentioned above, method to propagate these distribution-type parameters through a model (mostly OED model) is important. Sampling based methods are most intuitive, but very time consuming. Therefore, sigma point methods are introduced to approximate distribution using fixed number of points, which are selected by fitting several (usually the first two) moments of the distribution, and then propagate them through the model. Even they are much more efficient than the simple sampling based methods, there are still some shortcomings: 1) the number of points used is usually fixed; 2) since they capture only the first two moments in most cases, their accuracy becomes lower when the model is highly nonlinear.

In this paper, we implement a DMD (Dirac Mixture Distribution) method which, as well as in the sigma point methods, approximates Gaussian distribution using several points. But the locations of these points are computed by minimizing the distance between DMD and Gaussian distribution. Hence, 1) the number of points used can be flexible; 2) it captures not only several moments, but the feature of the whole Gaussian distribution. 6 toy models, including one simple ODE model, and JAK /STAT pathway model are implemented to assess the accuracy of this new method. 5 sigma point methods and 2 sampling based methods are used for comparison. Results show that the DMD method is more accurate than the other methods implemented in this paper, especially in models which are highly nonlinear, i.e. ODE models.

4.8 Identification of target genes and signaling networks of Wnt11 in human breast cancer progression

Maren Sitte[1]

[1]University Medical Center Göttingen, Institute of Medical Bioinformatics, Goldschmidtstr. 1, 37077 Göttingen, Germany

The non-canonical Wnt protein Wnt11 is known to be overexpressed in many human cancers including human breast cancer. While Wnt11 overexpression is associated with increased tumor cell migration and invasion, the molecular mechanisms underlying these effects are largely unknown. Therefore, we aimed to identify the cellular signaling network and target genes of Wnt11 that are involved in Wnt11-induced breast cancer invasion.

We first created human MCF-7 breast cancer cells stably overexpressing Wnt11 and characterized the cells by RNA-sequencing. The analysis revealed 42 genes that were significantly differentially expressed in Wnt11-positive cells, many of them involved in cell adhesion. In order to identify signaling molecules involved in Wnt11 signal transduction, we stimulated MCF-7 cells with recombinant Wnt11, performed a reverse phase protein array (RPPA) and set up network models using a bioinformatics approach. To reconstruct the pathways of the measured proteins in the RPPA we applied the method DDEPN and used pwOmics as a integrative approach to generate a network estimated from both data sets. Thereby, we identified a subset of proteins that was phosphorylated after 5-10 minutes of Wnt11 stimulation. To validate the results, we performed Western Blots which confirmed the activation of Akt in Wnt11-stimulated cells. Inhibition of the PI3K-Akt pathway, using specific inhibitors that are also in clinical use, resulted in a reduction of the pro-invasive effect of Wnt11 on breast cancer cell invasion.

Taken together, our analyses shed light on Wnt11-induced signaling pathways involved in breast cancer progression that could be used as targets for therapy.

4.9 Estimating Pharmacokinetic and- dynamic parameters of thrombopoiesis models using clinical data: tradeoff between simplicity and complexity

Yuri Kheifetz[1]

[1]IMISE, Härtel Str. 16-18, 04107 Leipzig

Objectives

An individualized biomathematical model of thrombopoiesis was developed by us to support clinical decision making and optimizing risk management of multicyclic poly-chemotherapy (Kheifetz, Scholz 2019). However, available individual clinical measurements are insufficient for reliable estimates of individual parameters. In order to tackle this problem, we developed and approach to exploit both, direct data of an individual patient and indirect data from other biological and clinical studies.

Methods

Our clinical data consisted of 138 non-Hodgkin's lymphoma patients treated with CHOP-like chemotherapies. We developed a virtual participation methodology in which we simulated each individual parameter set for each patient under respective real clinical condition as well as under conditions of independent population-based experiments measuring 12 averaged biological features. Deviations of simulations from all data points of all types were summed up to individual goal functions to be optimized. We derived relative cytotoxic effects on different precursor cells from available in vitro studies. Relative PD effects of cyclophosphamide, doxorubicin and etoposide were derived from publically available population fitting results of a simple semi-mechanistic model applied to other clinical studies. We derived relative degradation rates of thrombopoietin by megakaryocytes of different ploidies and platelets from data of relative receptor densities. We assumed different parsimonies: Parameters with related biology and low identifiability were set to the same value (for example, chemotherapy PD effects of active stem and pluripotent precursor cells). We compared alternative modeling assumptions using Bayesian information criteria (BIC).

Results & Conclusions

By our complex mechanistic model of thrombopoiesis we integrated existing PK models of cytotoxic drugs. This allows estimation of relevant pharmacometric covariates at an individual patient level.

Taking into account a number of chemotherapy sensitive cells types and drugs, we needed to estimate more than 20 PD different parameters, which connect drug concentrations with cytotoxic effects on different cell types. By using in vitro data, indirect inferences from other studies and parsimony assumptions we managed to reduce this number to only two population and two individual PD parameters. Application of BIC showed the optimality of these decisions.

Our novel methodology of virtual participation of a patient in literature studies enabled reliable fitting of system parameters of numerous precursor cells and ensured consistency of our model with the constantly increasing biological and clinical knowledge about thrombopoiesis. This will greatly improve the chances of model-based clinical decision making in an individual chemotherapy situation.

4.10 Thrombopoietin mediates the maturation of megakaryocytes via the expansion of primitive hematopoietic cells – a model perspective

Andrea Gottschalk[1]

[1] Institute for Medical Informatics and Biometry, TU Dresden

Thrombopoietin (Thpo) is a ligand of the receptor Mpl and the key regulator of megakaryopoiesis and platelet production. Congenital amegakaryocytic thrombocytopenia (CAMT) or myeloproliferative neoplasm (MPN) are human diseases that are caused by mutations resulting in a loss or gain of function of the Thpo-receptor. Although several mathematical models focus on the effects of chemotherapy and irradiation on thrombopoiesis, we aimed to analyze, whether those models are able to describe the dysregulated phenotypes of CAMT or MPN.

We compared an established model by Wentz et al. (Ref) with simpler mathematical models of thrombopoiesis suggesting additional mechanistic regulations regarding the Thpo/Mpl binding or the feedback between different cell stages. By using approximate bayesian computation, we parametrized the different mathematical models based on data from mouse models with various dysregulated phenotypes: (1) Mpl overall deficiency (CAMT), (2) Mpl deficient megakaryocytes and thrombocytes (MPN), (3) Thpo deficiency and (4) Thpo overexpression. We evaluated the quality of the model fit based on the changes of homeostatic levels of Thpo, primitive hematopoietic cells and thrombocytes.

The model comparison showed that the established models of thrombopoiesis cannot account for the clinically relevant phenotypes. Moreover, our analysis suggests two additional feedback mechanisms regulating thrombopoiesis: i) Thpo-binding inhibits the proliferation of early megakaryocytes and ii) primitive hematopoietic cells inhibit the maturation of megakaryocytes.

Our model analysis suggests that additional feedback loops in thrombopoiesis are necessary to explain the compensation mechanism observed in Thpo-receptor related diseases. Further experimental and model studies are required to fully resolve the nature of these regulations.

4.11 Model-driven suggestions regarding Staphylococcus aureus chronic infection are validated experimentally and induce full clearance

Lito A. Papaxenopoulou[1]

[1]BRICS Braunschweig Integrated Centre of Systems Biology, Rebenring 56, 38106, Braunschweig Germany

Staphylococcus aureus is a hazardous bacterium, which is responsible for nosocomial- and community-acquired infections globally. It is notorious for its multidrug resistance, which leads to recurrent or chronic infections, and even life-threatening diseases. In chronic infections, the presence of a population of cells that suppress the function of T cells helps the persistence of the bacterium. These cells are known as Myeloid Derived Suppressor Cells (MDSC) and they consist of heterogeneous groups of immature myeloid cells. In this study, our mathematical model sheds light onto whether the expansion of the MDSC during chronic S. aureus infection takes place in the site of infection or systemically. We conclude that the origin of the proliferation is predominantly systemic, and our conclusion is validated by experimental data. Further analysis of the model suggests perturbation approaches to destabilize such chronic infection equilibria in the system, which could induce clearance. Experiments following up these mathematical predictions were conducted and experimental results confirmed the model-driven suggestions revealing MDSC reduction, recover of T cell function and complete clearance from S. aureus.

4.12 Loss-Function Learning for Digital Tissue Deconvolution

Marian Schön[1]

[1] Institute of Functional Genomics, Statistical Bioinformatics University of Regensburg, Am BioPark 9, 93053 Regensburg, Germany

Gene-expression profiling of bulk tumor tissue facilitates the detection of gene regulation in tumor cells. However, differential gene expression can originate from both tumor cells and the cellular composition of the surrounding tissue. The cellular composition is not accessible in bulk sequencing but can be estimated computationally.

We propose Digital Tissue Deconvolution (DTD) to estimate cellular compositions from bulk sequencing data. Formally, DTD addresses the following 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.

In Görtler et al (2018), we presented a method to adapt the loss function to the deconvolution problem of interest. Here, we introduce the related R package "DTD". It provides all implementations, functions and routines for loss-function learning. Visualization functions are included to assess the quality of a deconvolution model and to gain additional information on the loss-function learning procedure. We present our package in an exemplary analysis, where we estimate immune cell quantities from gene-expression profiles of melanoma specimens. Using loss-function learning we increased the accuracy from a correlation of 29% to 72% between true and estimated cellular proportions.

4.13 A tipping point in tumor-immune dynamics determines patient fate

Jeroen Creemers[1]

[1]Department of Tumor Immunology, Radboudumc, Nijmegen, The Netherlands

Novel treatment modalities such as immunotherapy and targeted therapies are revolutionizing care for patients with metastatic cancers. Despite these advances, treatment effects are very heterogeneous – many treatments do not benefit a majority of the patients. Explanations for these heterogenic effects are sought in mechanisms within the tumor microenvironment. The tumor microenvironment develops from an “arms race” between two complex systems: the developing tumor and our immune system. To investigate the underlying interactions of these complex competing systems, we used in silico modeling based on ordinary differential equations to capture the essential principles and dynamics in anti-tumor immunity with respect to overall survival of cancer patients. The main components of our model are the priming, clonal expansion and migration of T cells from lymph nodes into the tumor microenvironment and the formation of tumor-immune complexes and subsequent killing of developing tumors. These few basic principles enabled realistic simulation of natural courses of disease of cancer patients. While one might expect a gradual influence of our immune system or a malignancy on overall survival of patients (e.g. weak activation of the immune system will result in a small survival benefit, while stronger activation will yield a longer survival benefit), we instead found a ‘tipping point’ or critical state transition within tumor-immune dynamics, meaning either the immune system or the cancer emerges as clear winners form the arms race. We expanded our model with the addition of several treatment modalities, targeting either the tumor-axis (e.g. surgery, radiotherapy, chemotherapy or targeted therapy) or the immune-axis of the system (e.g. checkpoint inhibitors, chimeric antigen receptor T cells or anti-cancer vaccines). Dependent on the duration of the therapy and effect size of the treatment, induction of a temporary or long-term clinical response was possible. In summary, our in silico model predicts the presence of tipping points within anti-tumor immunity. Notably, this prediction is only based on fundamental dynamics shared by most tumors, such that the tipping point should be a robust property of all cancers. Potential implications include biomarker discovery, clinical trial design and ultimately clinical decision making for cancer patients.

4.14 Numerical approach for a reaction-diffusion-chemotaxis model

Camile Fraga Delfino Kunz[1]

[1]Goethe Universität, Ruth Moufang Straße 1, 60438, Frankfurt am Main, Germany

Modelling cells and tissues is an emergent field in biomedical sciences. Some behaviors in biological cells are yet unknown and there is a lot of open questions in this field. Cell migration can occur collectively or individually. On the collective cell migration, cell-cell interactions promote a coordinated behavior and it plays an important role on some biological process, such as immune system, tissue remodeling, wound healing and regeneration, pattern formation, and also in diseases such as cancer.

The group of F. Matthäus has developed expertise in the new area of mathematical biology which combines modeling with image/data analysis, and acts in strong collaboration with experimental partners. Through data from time-lapse microscopy and using particle velocimetry (PIV), it is possible to obtain spatio-temporal velocity distributions, divergence, vorticity, streamlines or pathlines. Based on these data some past works developed agent-based or hybrid mathematical models taking into account mechanical cell-cell interaction, mechanotransduction, chemotaxis, and also interaction with a dynamically changing chemical environment.

In a present project we consider skin patterning in mouse embryos, where cell aggregates form based on a hierarchical process involving a Turing system and cell chemotaxis. Hereby, epidermal cells produce growth factors which interact and form a chemical spot pattern. The epidermal cells react chemotactically to this spot pattern and condense in the areas of the spots. Furthermore, the epidermal cells can move collectively. To better understand this system we study a reaction-diffusion-chemotaxis model. As this model is not chosen for mathematical tractability but with respect to describe the biological process properly, rigorous mathematical analysis is only possible by numerical approaches. We present the numerical approach chosen to analyse the model and some preliminary results.

4.15 An in silico analysis of patient-adaptive Interleukin-2 therapy in autoimmune and inflammatory diseases

Sahamoddin Khailaie[1]

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

Regulatory T cells (Tregs) are suppressor cells that control self-reactive and excessive effector T cells (Tconvs) responses. Breakdown of the balance between Tregs and Tconvs is the hallmark feature of autoimmune and inflammatory diseases. Due to the positive dependency of both populations on Interleukin-2 (IL-2), it is a subtle leverage to restore the healthy immune balance. By employing a mechanistic mathematical model, we studied the IL-2 therapy for stabilizing Treg population and restricting inflammatory Tconv response. We introduced an adaptive control strategy to design the minimal IL-2 dosage.

4.16 Zero-sum regression in action: A prognostic miRNA Signature in DLBCL

Gunther Glehr[1]

[1] Statistical Bioinformatics, University of Regensburg, Am BioPark 9, 93053 Regensburg, Germany

OMICs data sets need preprocessing before analysis. This intrinsically includes the use of reference points like the mean expression of all features, a defined spike-in or the value of housekeeper genes.

Reference points have two major drawbacks: Measuring platforms become incomparable (Altenbuchinger et al. 2017) and noise of the reference point is added to a predictive model (Bates and Tibshirani 2017).

The concept of zero-sum signatures, introduced by Lin et al. 2014 with extensions from Altenbuchinger et al. 2017 and Bates and Tibshirani 2017, enables that a predicted response is free of reference points.

If all possible, unique log-ratios of measurements are used in a LASSO-penalized regression, such signatures directly emerge. However, the feature space expands from p measurements to p choose 2 new features.

Here we use expression levels of 800 micro-RNAs (miRNAs), measured with the NanoString nCounter miRNA system. 228 DLBCL specimens were used to find a predictive signature on all high-count log-ratios for overall and progression free survival. We show that, besides the zero-sum property, the found log-ratio features are predictive but the corresponding single measurements are not.

References

Altenbuchinger, M. et al. (2017). “Reference point insensitive molecular data analysis”. In: Bioinformatics (Oxford, England) 33.2, pp. 219–226. doi: 10.1093/bioinformatics/btw598.

Bates, Stephen and Robert Tibshirani (2017). Log-ratio Lasso: Scalable, Sparse Estimation for Log-ratio Models. url: http://arxiv.org/pdf/1709.01139v1.

Lin, W. et al. (2014). “Variable selection in regression with compositional covariates”. In: Biometrika 101.4, pp. 785–797. issn: 0006-3444. doi: 10.1093/biomet/asu031.

4.17 A multi-source data integration approach reveals novel associations between metabolites and renal outcomes in the German Chronic Kidney Disease study

Helena U. Zacharias[1], Michael Altenbuchinger[2], Stefan Solbrig[3], Andreas Schäfer[3], Mustafa Buyukozkan[1], Ulla T. Schultheiß[4], Fruzsina Kotsis[4], Anna Köttgen[4], Rainer Spang[2], Peter J. Oefner[2], Jan Krumsiek[5], Wolfram Gronwald[2], GCKD study investigators

[1]Institute of Computational Biology, Helmholtz Zentrum München, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany.
[2]Institute of Functional Genomics, University of Regensburg, Am Biopark 9, 93053 Regensburg, Germany.
[3]Department of Physics, University of Regensburg, Universitätsstraße 31, 93053 Regensburg, Germany.
[4]Institute of Genetic Epidemiology, Department of Biometry, Epidemiology, and Medical Bioinformatics, Faculty of Medicine and Medical Center, University of Freiburg, 79106 Freiburg, Germany.
[5]Institute of Computational Biomedicine, Weill Cornell University, New York, NY 10021, USA.

Omics data facilitate the gain of novel insights into the pathophysiology of diseases and, consequently, their diagnosis, treatment, and prevention. To that end, it is necessary to integrate omics data with other data types such as clinical, phenotypic, and demographic parameters of categorical or continuous nature. Here, we exemplify this data integration issue for a study on chronic kidney disease (CKD), where complex clinical and demographic parameters were assessed together with one-dimensional (1D) ^{1}H NMR metabolic fingerprints.

Routine analysis screens for associations of single metabolic features with clinical parameters, which requires confounding variables typically chosen by expert knowledge to be taken into account. This knowledge can be incomplete or unavailable.

We introduce a framework for data integration that intrinsically adjusts for confounding variables. We give its mathematical and algorithmic foundation, provide a state-of-the-art implementation, and give several sanity checks. In particular, we show that the discovered associations remain significant after variable adjustment based on expert knowledge. In contrast, we illustrate that the discovery of associations in routine analysis can be biased by incorrect or incomplete expert knowledge in univariate screening approaches. Finally, we exemplify how our data integration approach reveals important associations between CKD comorbidities and metabolites. Moreover, we evaluate the predictive performance of the estimated models on independent validation data and contrast the results with a naive screening approach.

4.18 Predicting Comorbidities of Epilepsy Patients Using Big Data from Electronic Health Records Combined with Biomedical Knowledge

Thomas Linden[1]

[1]UCB Biosciences GmbH

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 ($\sim$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.