Workshop on Computational Models in Biology and Medicine 2023

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

June 15 – 16, 2023

University of Stuttgart,
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

- Matthias König (Humboldt-University Berlin): "Advancing Liver Function Assessment: Personalized and Stratified Approaches with Standardized Computational Models and Data"

- Jana Wolf (Max-Delbrück-Center for Molecular Medicine (MDC), Berlin): "Modeling signal transduction and gene expression in cancer: from information processing to patient specific-models"

- Lars Kaderali (University of Greifswald): "Mathematical Modeling of Infection and Immune Response"

- Achim Tresch (University of Cologne): "Dimension reduction for spatial transcriptomics data"

Workshop venue

The workshop takes place at the University of Stuttgart Campus Vaihingen, Universitätsstraße 32, 70569 Stuttgart.

Organization

The workshop is jointly organized by the GMDS/IBS working groups "Mathematical Models in Medicine and Biology" (speakers: Nicole Radde, University of Stuttgart; Ingmar Glauche, Technische Universität Dresden) and "Statistical Methods in Bioinformatics" (speakers: Michael Altenbuchinger, University Medical Center Göttingen; Tim, Kacprowski, Technical University of Braunschweig; Markus Wolfien, University Hospital Carl Gustav Carus, Dresden, Germany).

Contact

Prof. Dr. Nicole Radde,
Institute for Stochastics and Applications,
E-mail: nicole.radde@simtech.uni-stuttgart.de
Sebastian Höpfl,
Institute for Stochastics and Applications,
E-mail: sebastian.hoepfl@isa.uni-stuttgart.de

Support

The workshop is funded by the "Deutsche Gesellschaft für Medizinische Informatik,

Biometrie und Epidemiologie (GMDS)", the "Deutsche Region der Internationalen Biometrischen Gesellschaft (IBS-DR)" and the Cluster of Excellence EXC 2075 "Data-Integrated Simulation Science (SimTech)".

1 Program

Thursday, Jun 15, 2023

11:30 Registration opens (with small lunch)

12:45–13:00 Welcome

Session 1: FAIR model and data principles

13:00–13:40 Keynote lecture: Matthias König
Advancing Liver Function Assessment: Personalized and Stratified Approaches with Standardized Computational Models and Data

13:40–13:55 Dagmar Waltemath
Having fun reusing computational biology models? - and why not?

13:55–14:10 Jürgen Pleiss
EnzymeML - Modelling challenges in enzymology and biocatalysis

14:10–14:25 Sebastian Höpfl
Journals in Systems Biology could improve their impact factor by enforcing reproducibility

14:25–15:30 Coffee break & poster session I

Session 2: Dynamical modeling in health and disease

15:30–16:10 Keynote lecture: Jana Wolf
Modeling signal transduction and gene expression in cancer: from information processing to patient specific-models

16:10–16:25 Bachelot Yann
Spatial distancing: Investigation of a defense mechanism for pathogen immune evasion

16:25–16:40 Ulrich Mansmann
Shape-specific characterization of colorectal adenoma growth and transition to cancer with stochastic cell-based models

16:40–17:10 Coffee break

17:10–17:25 Philipp Altrock
Stochastic dynamics of cancer relapse in hematologic malignancies

17:25–17:40 Gavin Fullstone
In silico-guided optimisation of CNS-targeted Therapeutic Antibodies

17:40–17:55 Felix Weidner
Leveraging quantum computing for dynamic analyses of gene regulatory networks

18:00 Dinner (Pizza with vegeterian and vegan options) & poster session II

Friday 16, 2023

8:30–9:00 Brezeln and coffee

Session 3: Statistical analysis and machine learning

9:00–9:40 Keynote lecture: Lars Kaderali
Mathematical Modeling of Infection and Immune Response

9:40–9:55 Hryhorii Chereda
Ensemble-GNN: federated ensemble learning with graph neural networks for disease module discovery and classification

9:55–10:10 Stefan Schrod
Individual treatment effect estimation for survival data

10:10–10:25 Konrad Grützmann
Network-based analysis of heterogeneous patient-matched brain and extracranial melanoma metastases pairs reveals three homogeneous subgroups

10:25–10:40 Elham Shamsara
A COVID-19 informed neural network with adaptive weighting during variants of concern in Germany

10:40–10:55 Poster award and group picture

10:55–11:30 Coffee break

Session 4: Open topics

11:30–12:10 Keynote lecture: Achim Tresch
Dimension reduction for spatial transcriptomics data

12:10–12:25 Ana Stolnicu
Establishing a trustworthy signalling entropy calculation for biological processes analysis

12:25–12:40 Franziska Görtler
ADTD - Adaptive Digital Tissue Deconvolution.

12:40–12:55 Laura Strohmaier
Gene-regulatory networks controlling cell fates downstream of TGFβ-signaling in MCF10A cells

12:55–13:10 Sophia Krix
MultiGML: Multimodal Graph Machine Learning for Prediction of Adverse Drug Events

13:10–13:15 Closing remarks

2 Keynotes

Keynote 1

Advancing Liver Function Assessment: Personalized and Stratified Approaches with Standardized Computational Models and Data

Matthias König

Humboldt-University, Germany

Abstract: Essential prerequisites for the practical application and translation of computational models include: i) reproducibility of results; ii) reusability and extensibility of models; iii) data availability; and iv) strategies for model stratification and individualization. In this study, we present a modeling workflow tailored to these critical aspects, with a focus on liver function tests. Evaluating liver function is a crucial task in hepatology, yet accurately quantifying hepatic function has persisted as a clinical challenge. Dynamic liver function tests offer a promising method for non-invasive in vivo assessment of liver function and metabolic phenotyping. These clinical tests determine liver function through the elimination of a specific test substance, thus revealing information about the liver's metabolic capacity.
We employed whole-body physiologically-based pharmacokinetic (PBPK) models to simulate these tests, which encompass absorption, distribution, metabolism, and elimination processes. PBPK models serve as powerful instruments for investigating drug metabolism and its impact on the human body. In this research, we showcase our efforts in utilizing PBPK models as digital twins for metabolic phenotyping and liver function evaluation [1-6]. To develop and validate our models, we created the first open pharmacokinetics database, PK-DB, containing curated data from over 600 clinical studies. Our models are individualizable and stratifiable, enabling simulation of lifestyle factors and co-administration effects on drug metabolism.
We have applied our models to various clinical inquiries, such as simulating individual outcomes post-hepatectomy using an indocyanine green model and examining the influence of CYP2D6 gene variants through a dextromethorphan model integrated with drug-gene interactions. These models are constructed hierarchically, describing metabolic and other biological processes in organs like the liver and kidneys, connected to whole-body physiology. All models and data are accessible for reuse in a reproducible manner, encoded in the Systems Biology Markup Language (SBML). In this study, we provide an overview of PBPK models and demonstrate how SBML [7-8], COMBINE standards [9], and FAIR principles [10] can facilitate model development, coupling, and reuse.

[1] PK-DB: pharmacokinetics database for individualized and stratified computational modeling. Grzegorzewski J, Brandhorst J, Green K, Eleftheriadou D, Duport Y, Barthorscht F, Köller A, Ke DYJ, De Angelis S, König M. Nucleic Acids Res.

2021 Jan 8;49(D1):D1358-D1364.
doi:10.1093/nar/gkaa990. pmid:33151297
[2] Pharmacokinetics of caffeine: A systematic analysis of reported data for application in metabolic phenotyping and liver function testing. Jan Grzegorzewski, Florian Bartsch, Adrian Köller, and Matthias König. Frontiers in Pharmacology 2022, Vol12.
doi:10.3389/fphar.2021.752826. pmid:35280254
[3] Prediction of survival after hepatectomy using a physiologically based pharmacokinetic model of indocyanine green liver function tests. Adrian Köller, Jan Grzegorzewski, Michael Tautenhahn, Matthias König. Front. Physiol., 22 November 2021.
doi:10.3389/fphys.2021.730418. pmid:34880771
[4] Physiologically based modeling of the effect of physiological and anthropometric variability on indocyanine green based liver function tests. Adrian Köller, Jan Grzegorzewski and Matthias König. Front Physiol. 2021 Nov 22;12:757293.
doi:10.3389/fphys.2021.757293. pmid:34880776
[5] Physiologically based pharmacokinetic (PBPK) modeling of the role of CYP2D6 polymorphism for metabolic phenotyping with dextromethorphan. Grzegorzewski, J., Brandhorst, J., König, M. Front Pharmacol. 2022 Oct 24;13:1029073.
doi:10.3389/fphar.2022.1029073. pmid:36353484
[6] A physiologically based pharmacokinetic model for CYP2E1 phenotyping via chlorzoxazone. J. Küttner, J. Grzegorzewski, HM. Tautenhahn, M. König. bioRxiv 2023.04.12.536571 (preprint).
doi:10.1101/2023.04.12.536571
[7] SBML Level 3: an extensible format for the exchange and reuse of biological models. SM Keating, D Waltemath, M König, F Zhang, A Dräger, C Chaouiya, FT Bergmann, A Finney, CS Gillespie, T Helikar, S Hoops, RS Malik-Sheriff, SL Moodie, II Moraru, CJ Myers, A Naldi, BG Olivier, S Sahle, JC Schaff, LP Smith, MJ Swat, DT, L Watanabe, DJ Wilkinson, ML Blinov, K Begley, JR Faeder, HF Gómez, TM Hamm, Y Inagaki, W Liebermeister, AL Lister, D Lucio, E Mjolsness, CJ Proctor, K Raman, N Rodriguez, CA Shaffer, BE Shapiro, J Stelling, N Swainston, N Tanimura, J Wagner, M Meier-Schellersheim, HM Sauro, B Palsson, H Bolouri, H Kitano, Akira Funahashi, H Hermjakob, JC Doyle M Hucka, and SBML Community members. Mol Syst Biol. 2020;16(8):e9110.
doi:10.15252/msb.20199110
[8] The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 2 Core. M. Hucka, F. Bergmann, C. Chaouiya, A. Dräger, S. Hoops, S. Keating, M. König, N Le Novere, C. Myers, B. Olivier, S. Sahle, J. Schaff, R. Sheriff, L. Smith, D. Waltemath, D. Wilkinson, F. Zhang. J Integr Bioinform. 2019 Jun 20;16(2);.
doi:10.1515/jib-2019-0021.
[9] Specifications of Standards in Systems and Synthetic Biology: Status and Developments in 2022 and the COMBINE meeting 2022. M. König, P. Gleeson, M. Golebiewski, T. Gorochowski, M. Hucka, S. Keating, C. Myers, D. Nickerson, F. Schreiber. J Integr Bioinform. 2023 Mar 29;20(1)..
doi:10.1515/jib-2023-0004. pmid:36989443

[10]Ten Simple Rules for FAIR Sharing of Experimental and Clinical Data with the Modeling Community. Matthias König*, Jan Grzegorzewski, Martin Golebiewski, Henning Hermjakob, Mike Hucka, Brett Olivier, Sarah M. Keating, David Nickerson, Falk Schreiber, Rahuman Sheriff, Dagmar Waltemath. Preprints 2021, 2021080303 (in revision).
doi:10.20944/preprints202108.0303.v2

Keynote 2

Modeling signal transduction and gene expression in cancer: from information processing to patient specific-models

Jana Wolf

Max-Delbrück-Center for Molecular Medicine (MDC), Berlin

Abstract: Signaling pathways are frequently perturbed in cancer. The IKK/ NF-kB and p53 pathways are critical for the cellular response to DNA damage. They regulate the cell fate decision towards apoptosis, survival or senescence and are therefore central in tumor development and response to treatment. We show how the establishment of qualitative and quantitative pathway models based on cell population and single cell data enables an investigation of cellular information processing, the characterization of cellular heterogeneity and the elucidation of possible cross talks between pathways. Moreover, we use clinical data for the development of patient-specific models capturing the different combinations of perturbations per patient as a basis for an identification of optimal target strategies.

Keynote 3

Mathematical Modeling of Infection and Immune Response

Lars Kaderali

University of Greifswald, Germany

Abstract: The past two years of the Covid pandemic have made mathematical modeling of viral infections broadly visible to the general public. While models have been used to predict future infection numbers as much as number of patients in intensive care and on ventilators, there are also less visible models that describe the molecular processes within a host during infection, using a combination of high throughput experimental data, bioinformatics analysis, as well as detailed mathematical modeling using differential equations. These models can impact on drug development and antiviral treatment strategies. The talk will bridge from intracellular models of infection and immune response over models at the tissue and organ level to patient and drug treatment data, and I will also briefly touch upon models used to describe infections at the population level.

Keynote 4

Dimension reduction for spatial transcriptomics data

Achim Tresch

University of Cologne, Germany

Abstract: The availability of techniques for high-resolution spatial transcriptomics measurements has led to a surge of methods for spatial data analysis, with the aim to detect spatial expression patterns and genes that follow these patterns. In this note, we introduce Spatial Components Analysis (SpaCo, available in R), a statistical method that fully exploits the multivariate structure of the data instead of performing gene-wise analyses. The graph Laplacian of the spatial layout of the measurement spots defines a scalar product and a distance metric between expression patterns. Inspired by Principal Components Analysis (PCA) and using this scalar product, SpaCo constructs an ordered basis - the spatial components - of the gene space, which is optimal in minimizing Geary's C, a well-known spatial statistic. We show that SpaCo is superior to PCA when used as primary dimension reduction during data preprocessing while being equally efficient. Regarding scalar multiples as identical, we represent expression profiles of genes as points in projective space endowed with the Fubini-Study metric induced by our scalar product. Spatially distributed genes are identified as those with minimal distance to the projective subspace spanned by the metagene expression profiles of the relevant spatial components.

3 Oral presentations

3.1 Having fun reusing computational biology models? - and why not?

Dagmar Waltemath

University Medicine Greifswald, Medical Informatics Laboratory, Germany

Abstract: Dagmar Waltemath1, Olaf Wolkenhauer2, Christian Winterhalter3, Markus M. Becker4, Matthias König5, David Nickerson6, Irina Balaur7

1Medical Informatics Laboratory, University Medicine Greifswald, Greifswald, Germany,
https://orcid.org/0000-0002-5886-5563
2Systems Biology and Bioinformatics, University of Rostock, Rostock, Germany
3University of Greifswald, University Library, christian.winterhalter@uni-greifswald.de,
https://orcid.org/0000-0001-8618-0337
4Leibniz Institute for Plasma Science and Technology (INP), Greifswald, Germany,
https://orcid.org/0000-0001-9324-3236
5Humboldt University of Berlin, Institut for Biology, Institute for Theoretical Biology, Systems Medicine of Liver, Philippstraße 13, Berlin, Germany,
https://orcid.org/0000-0003-1725-179X
6Auckland Bioengineering Institute, University of Auckland, New Zealand.
https://orcid.org/0000-0003-4667-9779
7Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Luxembourg,
https://orcid.org/0000-0002-3671-895X

Computational modeling and simulation is a standard tool in many disciplines relating to biology and medicine. Virtual experiments simulate the functioning of biomedical processes and systems on various scales and levels of detail. We believe that adherence to FAIR principles supports reproducibility, semantic descriptions of model constituents, and accessibility of all model-related data. The ability to comprehend, and potentially reuse, previous works and resources eases the burden for scientists to develop new models, methods, and tools to make impactful contributions. This desire to reuse has led to a variety of standards, guidelines, and principles to be considered when publishing virtual experiments.

Among them are widely accepted requirements such as those for Good Scientific Practice or the FAIR Guiding Principles for Data Stewardship (Wilkinson et al., https://doi.org/10.1038/sdata.2016.18). The FAIR principles suggest criteria for sharing a virtual experiment with respect to findability, accessibility, interoperability and ultimately reusability of the work. The systems biology community also requests reproducibility of their simulation projects (Niarakis and Waltemath et al., https://doi.org/10.1093/bib/bbac212). For example, the EOSC-funded project

(https://fair-ca-indicators.github.io/) aims to raise awareness for FAIR in the COMBINE community and specifically to promote the assessment of the quality of research outcomes using a coordinated FAIR assessment. We emphasise the benefits of following and applying FAIR principles throughout the model life-cycle, from the model development stage through to after the model is released. The main challenge in this project is achieving a community-level accepted set of FAIR indicators to capture the specifics of the COMBINE resources. For this, we rely on the engagement of community members. We continuously do this community-wide engagement with workshops and training sessions, transparent communication, online dissemination and involvement acknowledgement - all our work components are open and freely-accessible and connected via our project website to major dedicated frameworks such as zenodo for dissemination material, github for code files etc. We also target dissemination of our ongoing work, results, lessons learnt and challenges to the broader community of Computational Biology and Open Science.

Following the domain-specific requirements defined in Minimum Information Guidelines for models (Le Novere et al., https://doi.org/10.1038/nbt1156) and simulations (Waltemath et al., https://doi.org/10.1371/journal.pcbi.1001122) is a first step in the direction of FAIR model sharing. Beyond the minimum requirements, using appropriate standard formats (e.g., SBML, CellML, SBGN, SED-ML, see also König et al., https://doi.org/10.1515/jib-2023-0004), together with semantic information in the form of relevant ontologies (Neal et al., https://doi.org/10.1093/bib/bby087) such as SBO (Courtot et al., https://doi.org/10.1038/msb.2011.77) and software libraries (Waltemath et al., https://doi.org/10.1515/jib-2020-0005) further enhances the FAIRness of a given study.

We will present here a high-level overview of this ecosystem, which is maintained by the COMBINE Coordination Board (https://co.mbine.org/). We will show examples of community efforts for model FAIRification (e.g., https://www.ebi.ac.uk/biomodels/covid-19), and we will briefly discuss why - despite all these community efforts - model reuse is still not as enjoyable and as simple as it should be. In fact, most studies remain non-reproducible (Höpfl et al., https://doi.org/10.1038/s41598-023-29340-2; Tiwari et al., https://doi.org/10.1038/s41598-023-29340-2).

Our hypothesis is that sufficient information, guidance, tool support, and incentives are missing in our community. As one consequence we wish to promote the idea of establishing an open information source for this research domain. This resource will connect disconnected resources, provide metadata about existing standards, models and guidelines and it will support scientists, for example, in discovering appropriate tools and formats, evaluating the reproducibility of their work, measuring the FAIRness of their work over time, and accessing community tools.

IB and DaW received funding from the European Union,Äôs Horizon 2020 research and innovation programme under grant agreement No 101017536. MK was supported by the Federal Ministry of Education and Research (BMBF, Germany) within LiSyM by grant number 031L0054 and ATLAS by grant number 031L0304B and by

the German Research Foundation (DFG) within the Research Unit Program FOR 5151 QuaLiPerF by grant number 436883643 and Priority Programme SPP 2311 Subproject SimLivA by grant number 465194077. DN is supported by National Institutes of Health [grant P41 GM109824] and [grant 3OT3OD025349] and by the New Zealand Ministry of Business, Innovation and Employment: Catalyst Strategic Fund (12 Labours).

3.2 EnzymeML - Modelling challenges in enzymology and biocatalysis

Jürgen Pleiss

University of Stuttgart, Institute of Biochemistry and Technical Biochemistry, Germany

Abstract: Designing complex biocatalytic reaction systems is a highly complex task due to the interdependence of various factors such as the enzymes, reaction conditions, and modeling methods, which affect the choice of a kinetic model and the estimated kinetic parameters. As a result, the reproducibility of enzymatic experiments and the reuse of enzymatic data are challenging. Previously, we have developed the XML-based markup language EnzymeML to enables the storage and exchange of enzymatic data, including reaction conditions, substrate and product time courses, kinetic parameters, and kinetic models. 1 Our approach aims to make enzymatic data FAIR (findable, accessible, interoperable, and reusable). EnzymeML serves as a seamless communication channel between experimental platforms, electronic lab notebooks, tools for modeling enzyme kinetics, publication platforms, and enzymatic reaction databases. 2 EnzymeML is a valuable tool for investigating kinetic models, and model reduction was performed by a general-purpose framework utilizing JAX.

References

1. Range, J. et al. EnzymeML,Äîa data exchange format for biocatalysis and enzymology. FEBS J. 289, 5864,Äì5874 (2022).

2. Lauterbach, S. et al. EnzymeML: seamless data flow and modeling of enzymatic data. Nat. Methods 2023 (2023) doi:10.1038/s41592-022-01763-1.

3.3 Journals in Systems Biology could improve their impact factor by enforcing reproducibility

Sebastian Höpfl

University of Stuttgart, Institute for Stochastics and Applications, Germany

Abstract: Systems Biology is at the forefront of developing and using modeling standards to improve the reproducibility and reuse of models. Still, a recent study by Tiwari et al. showed that only every second model out of a set of 455 models could be directly reproduced (Tiwari et al. 2021). This limited reproducibility shows the need for further efforts or mechanisms that promote reproducibility. Reproducibility of scientific results fosters collaborations and leads ultimately to faster progress in the whole field and increased trust in science. Since these merits are mainly long-term and not immediate for the publishing researcher, it seems crucial that third parties like publishing companies or funding organizations foster reproducibility. For example, the largest German research funding organization (German Research Foundation, DFG) signed 2003 the "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities" and has since been engaged in promoting open source and reusable data and publications. Here, we applied Bayesian Estimation to analyze the citation count of reproducible against non-reproducible models of the periods 1985-2020 and 2013-2020. The Bayesian Estimation Supersedes the t-Test method of J. Kruschke uses the Non-Central T-distribution to handle outliers in the data and estimates the distribution parameters via Markow-Chain-Monte-Carlo sampling (Kruschke 2013). The parameters are mean and sigma parameters for each dataset and a shape parameter that is jointly sampled for both datasets. Posterior predictive distributions for the difference of means and standard deviations and the effect size can be calculated from the posterior distributions of the parameters. From these, decisions are made via the 95% Highest Density Interval (HDI) of the difference of means posterior predictive distribution. Specifically, a decision can be made if the 95% values with the highest credibility (95From 1985 to 2020, there was no credible difference between reproducible and non-reproducible models. However, reproducible models got more citations, starting in 2013, ten years after the introduction of Systems Biology Markup Language (SBML) (Höpfl et al. 2023). There seems to be a tilting point between 2010 and 2013, from a tendency of more citations for reproducible papers to more than 95 % credibility for higher citation rates for reproducible papers. 2013 seems to be a reasonable time when the efforts of the Systems Biology community towards reproducibility pay off, i.e. it needed about ten years from the introduction of the de facto model standards SBML and CellML until the difference in citations became significant. On the one hand, this shows an immediate benefit for the individual researcher. On the other hand, this only pays off over the years and is, for example, for Ph.D. researchers who want to go into industry after their doctorate not relevant. However, journals could benefit directly and in the long run from an increased citation rate. At the same time, stricter reproducibility guidelines of journals would promote reproducibility and could thus strengthen trust in science. Our analysis shows that the

higher citation numbers of reproducible models cannot be explained by the fact that they were published in journals with a higher Journal Impact Factor (JIF). For this purpose, the citation counts were normalized to the respective JIF, and the normalized groups were compared. It was found that reproducibility leads to an increase in the number of citations independent of the respective JIF. The data indicate that, on average, journals in the field of Systems Biology could potentially increase their impact factor by about 30Reproducibility can be substantially improved with little effort. The authors of Tiwari et al. have assembled eight points to improve model reproducibility in a scorecard, whereby reproducibility is already significantly increased if only four of them are fulfilled. It is our appeal that everyone can do something for better scientific work through reproducible work. On the one hand, the individual researchers. On the other hand, the journals and funding organizations. In the end, the joint effort could pay off for the entire scientific community.

References

Höpfl, Sebastian; Pleiss, Jürgen; Radde, Nicole E. (2023): Bayesian estimation reveals that reproducible models in Systems Biology get more citations. In: Sci Rep 13 (1), S. 2695. DOI: 10.1038/s41598-023-29340-2. Kruschke, John K. (2013): Bayesian estimation supersedes the t test. In: Journal of experimental psychology. General 142 (2), S. 573,Äì603. DOI: 10.1037/a0029146. Tiwari, Krishna; Kananathan, Sarubini; Roberts, Matthew G.; Meyer, Johannes P.; Sharif Shohan, Mohammad Umer; Xavier, Ashley et al. (2021): Reproducibility in systems biology modelling. In: Molecular systems biology 17 (2), e9982. DOI: 10.15252/msb.20209982.

Funding

Funded by the Research Unit Programme FOR 5151 QuaLiPerF (Quantifying Liver Perfusion-Function Relationship in Complex Resection - A Systems Medicine Approach) by grant no. 436883643 and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2075 - 390740016. We acknowledge the support by the Stuttgart Center for Simulation Science (SimTech).

3.4 Spatial distancing: Investigation of a defense mechanism for pathogen immune evasion

Bachelot Yann

Leibniz Institute for Natural Product Research and Infection Biology Hans Knöll Institute, Jena, Germany

Abstract: Yann Bachelot(1), Paul Rudolph(1), Sandra Timme(1), Marc Thilo Figge(1,2)

(1)Applied Systems Biology, HKI-Center for Systems Biology of Infection, Leibniz-Institute for Natural Product Research and Infection Biology, Hans-Knöll-Institute (HKI), Jena, Germany (2)Institute of Microbiology, Faculty of Biological Sciences, Friedrich-Schiller-University, Jena, Germany

Understanding the interactions between the immune system and pathogens is crucial to develop improved diagnostics and therapeutic interventions. One of the key functions of the immune system is to efficiently recognize and eliminate foreign agents. This is achieved through a variety of mechanisms, including the labeling of pathogens with opsonins, which facilitate their recognition and uptake by phagocytes, as well as the secretion of molecules by immune cells that can kill pathogens extracellularly. However, some pathogens, such as the yeast Candida albicans, have developed strategies to evade the immune response. In this study, we pursue a modeling approach to investigate and simulate a potential immune evasion mechanism employed by pathogens.

We propose a mechanism, which we refer to as spatial distancing, where we consider a single pathogenic microbe being surrounded by antimicrobial peptides. We hypothesize that, as a means of defense, the pathogen secretes molecules that can bind to these antimicrobial peptides, forming complexes that then diffuse away from the cell. This reduces the concentration of antimicrobial peptides in the vicinity of the microbial pathogen and thereby increasing the chances of its survival. This model suggests that microbial pathogens can evade the immune system based on molecular complex formation and concentration equilibration by complex diffusion.

To investigate this mechanism, we applied two modeling approaches, where the reaction and diffusion of molecules are represented (i) by partial differential equations (PDEs) and (ii) within an agent-based model (ABM). Both modeling approaches simulate a single pathogenic cell in a three-dimensional environment. In the PDE model, molecules are simulated as concentrations that diffuse on a discrete grid according to the concentration gradient. In the ABM, molecules are represented as individual agents in continuous space performing a random walk. The initial state of simulations is characterized by the microbial pathogen being homogeneously surrounded by antimicrobial peptides, which can either be taken up or bound by the molecules secreted by the pathogenic cell. We performed a parameter screening for

both modeling approaches, where we used as a readout the total concentration of antimicrobial peptides that are taken up by the pathogen. In contrast to the PDE model, the ABM provides additional information on the lifetime of the molecular complexes.

The time-dependent spatial distributions of the various molecules revealed that the secretion of molecules by the pathogenic cell induces indeed a reduction in the concentration of antimicrobial peptides in the close vicinity of the microbial cell. This phenomenon was observed across a wide range of parameter values, suggesting that spatial distancing could be a robust and effective immune evasion mechanism pathogens use. The simulation results confirm the idea of spatial distancing as a beneficial defense mechanism for the pathogen and the inhibition of molecules secreted by the pathogen in defense against the antimicrobial peptides could be a possible target for therapeutic interventions.

3.5 Shape-specific characterization of colorectal adenoma growth and transition to cancer with stochastic cell-based models

Ulrich Mansmann

IBE, LMU, Munich, Germany

Abstract: Colorectal adenoma arelesions on the way to colorectal cancer. Screening colonoscopies allow a to remove them directly that reduces rates of cancer incidence and death. Mathematical models of adenoma growth and transition to cancer provide relevant information on how to implement screening strategies. The proposed modelling shows that shape may be of relevance, a factor which is neglected sofar. Stochastic cell-based models are applied to a data set of 197,347 Bavarian outpatients who had colonoscopies from 2006-2009, 50,649 patients were reported with adenoma and 296 patients had cancer. Multi-stage clonal expansion (MSCE) models were fitted. The modelling allws to derive a simple mathematical expression for the hazard ratio of interval cancers which provides a mechanistic understanding of this important quality indicator. The modelling process provides evidence that adenoma shape deserves closer consideration in screening strategies and as risk factor for transition to cancer.

3.6 Stochastic dynamics of cancer relapse in hematologic malignancies

Philipp Altrock

Max Planck Institut for Evolutionary Biology, Ploen, Germany

Abstract: In hematologic cancers, the initial response to therapy can be observed via temporal changes in tumor burden or of its approximate measures during treatment. Once the cancer cell population is small, disease kinetics can be dominated by stochastic processes. Many treatments evoke an initial response and tumor decline, but cancer cell populations that are difficult to detect can persist for extended times. Probabilistic outcomes arise through rare evolutionary events such as mutations that endow fitness advantage, or by neutral evolution and small-population-size fluctuations. In hematologic malignancies, the persisting disease is often called minimal residual disease (MRD). MRD results from an evolutionary process determined by cancer-intrinsic factors such as clonal evolution, and cancer-extrinsic influences such as therapy and cancer-immune system interactions. Through computational and mathematical modeling of nonlinear and stochastic cell population dynamics, we analyze how properties of the dynamics at the beginning of treatment can be used to predict, e.g., timing and nature of relapse. This allows to investigate the influence of stochasticity of MRD dynamics in leukemias and lymphomas to elucidate the laws of disease evolution during complex treatment schedules applied in the clinic.

3.7 In silico-guided optimisation of CNS-targeted Therapeutic Antibodies

Gavin Fullstone

University of Stuttgart

Abstract: Neurodegenerative diseases, such as Alzheimer,Äôs and Parkinson,Äôs disease, affect >7 million people in Europe and cause a profound deterioration in quality of life and mortality. The blood-brain barriers represent a major obstacle in treatment of neurodegenerative diseases, preventing >98% of small drugs and $\sim 100\%$ of macromolecules from reaching the CNS. Receptor-mediated transcytosis (RMT), where the engagement of specific receptors on the apical side of the cerebral endothelium can induce vesicular trafficking across the BBB, offers a potential pathway for successful CNS-delivery. However, it has been demonstrated that receptor-ligand binding properties including affinity and avidity dictate whether the therapeutic is targeted for lysosomal degradation or for basolateral release. To better understand this process and optimise intricate relationships between binding, trafficking and release, we employ data-driven agent-based models of RMT to delineate and optimise CNS-delivery and activity of potential therapeutics. In these models, properties including receptor levels, receptor-ligand affinity, avidity, steric arrangement and transport processes are integrated. These models recapitulate qualities of affinity-dependent trafficking during RMT. We have employed these models to demonstrate how low expression of glioblastoma therapeutic target receptors, TRAIL receptors, at the BBB interferes with the transport of a CNS-targeted TRAIL-receptor agonist. Furthermore, we are now utilising this approach to understand and improve the CNS delivery of a hexavalent TNFR2-agonist for the treatment of neurodegenerative diseases. Hereby, we have integrated our blood brain barrier models into a pharmacokinetic-pharmacodynamic framework to understand therapeutic clearance, delivery and on-site activity. This platform can therefore be utilised in the de novo design and optimisation of CNS-targeted therapeutics.

3.8 Leveraging quantum computing for dynamic analyses of gene regulatory networks

Felix Weidner

Institute of Medical Systems Biology, University of Ulm, Germany

Abstract: Background and Aims:
Boolean networks (BNs) are simple dynamical models for describing gene regulatory networks. Genes are represented as active (1) or inactive (0) at a given time point. The logical operators AND, OR, and NOT describe rules that specify network components' interactions. Stable states of such models are called attractors and correspond to biological phenotypes. The dynamic state space of BNs grows exponentially with the number of components, limiting analyses of larger systems. Motivated by the end of Moore,Äôs law, we investigate the potential offered by alternative hardware, namely quantum computers, for systems biology.
Methods:
The Boolean rules implementing the dynamics of biologically motivated networks are translated into quantum circuits consisting of a set of quantum logic gates.
These circuits can operate on a superposition of all possible states of the network. Thus, a linearly growing number of qubits can capture dynamics in an exponentially growing state space. Additionally, quantum algorithms such as Grover,Äôs algorithm can be used on this hardware to obtain speedups relative to classical approaches. The obtained circuits can be simulated on a local CPU to provide a benchmark and on real quantum processing units of IBM and IonQ.
Results:
We implement Quantum Boolean networks, providing quantum circuits for extracting information about the dynamics of the system.
Among other analyses, this allows for continuous tuning of weights associated with genes,Äô activities to explore specific subspaces of the network,Äôs dynamics while retaining the simplicity of Boolean logic. Furthermore, the outcomes of perturbations simulating biological knockouts and overexpressions can be more quickly screened by perturbing the network with a fixed superposition state.
Quantum search algorithms are adapted to invert the direction of dynamics, allowing for the identification of possible predecessors and initial conditions leading to an attractor, as well as the number of such states. The performance of two real quantum processors based on different hardware paradigms is compared.
Conclusion:
We could demonstrate a proof-of-principle of the use of quantum hardware for systems biology, highlighting the advantages that can be gained by exploiting the principles of superposition and entanglement.

3.9 Ensemble-GNN: federated ensemble learning with graph neural networks for disease module discovery and classification

Hryhorii Chereda

Department of Medical Bioinformatics, University Medical Center Göttingen, Germany

Abstract: Federated learning enables collaboration in medicine by allowing data to be shared across multiple parties without the need to aggregate it in a central cloud. Biological network graphs are a common type of data in the biomedical domain. While machine learning models can be applied to a wide range of data types, graph neural networks (GNNs) are particularly developed to perform different tasks with graphs. For instance, a patient can be represented by a biological network where the nodes contain patient-specific omics features. In this case, GNNs perform graph classification to predict a patient's clinical endpoint. Our Ensemble-GNN approach builds predictive models utilizing PPI networks containing various node features such as gene expression and/or DNA methylation. To do this, Ensemble-GNN derives relevant PPI network communities (subnetworks) and trains an ensemble of GNN models based on the inferred communities. Also, our method allows for the deployment of a federated ensemble of GNNs. Using a large breast cancer dataset of 981 patients and 8469 genes from the Cancer Genome Atlas (TCGA), we show that in a federated setup, models trained on subnetworks locally and shared across multiple parties/clients globally can improve client-specific predictive performance.

3.10 Individual treatment effect estimation for survival data

Stefan Schrod

Universitätsmedizin Göttingen, Germany

Abstract: Outcome prediction is one of the routine applications of precision medicine. However, predicting individual treatment effects for patients remains challenging. In this context, so-called counterfactual analyses were established in recent years. Estimating causal effects from observational data is different from classical machine learning in that we never see the individual treatment effect in our training data; for each patient, we usually see his/her outcome either treated or untreated, but usually not both. In other words, we observe just the outcome after the factual treatment (medication yes/no) and not for the respective counterfactual treatment which was not applied to this patient. Causal reasoning models are used to optimize treatment decisions computationally. However, these models can rarely deal with survival data. We present BITES (Balanced Individual Treatment Effect for Survival data) an approach which combines a potential outcome Deep Neural Network structure with a Cox regression loss function. By simultaneously learning a latent layer data representation, regularized by an Integral Probability Metric, removes bias of imbalanced treatment assignments. We demonstrate that BITES outperforms state-of-the-art methods in both simulation studies and in an application, in which we optimize hormone treatment for breast cancer patients based on six routine parameters.

3.11 Network-based analysis of heterogeneous patient-matched brain and extra-cranial melanoma metastases pairs reveals three homogeneous subgroups

Konrad Grützmann

Institute for Medical Informatics and Biometry, TU Dresden, Germany

Abstract: Melanoma is the most deadly form of skin cancer. The limited treatment success of brain compared to extracranial melanoma metastases is poorly understood. Here, 11 heterogeneous patient-matched pairs of brain and extracranial melanoma metastases were analyzed using melanoma-specific gene regulatory networks learned from public transcriptome and methylome data. This allowed to predict potential impacts of patient-specific driver candidate genes on other genes and pathways, revealing that: (i) pathway impact ratios of the patient-matched metastasis pairs clustered into three robust subgroups, (ii) subgroup-specific downstream pathways were enriched, and (iii) top differentially impacted subgroup-specific genes with known roles in cancer including melanoma (GATA3, BCL11B, FES, LDLR, MST1R, CD247). Patient subgroups and ranking of target gene candidates were confirmed in a validation cohort. Summarizing, network-based impact analyses of heterogeneous metastases pairs predicted individual regulatory differences in melanoma brain metastases, cumulating into three subgroups with specific downstream target genes.

3.12 A COVID-19 informed neural network with adaptive weighting during variants of concern in Germany

Elham Shamsara

Methods in Medical Informatics, Department of Computer Science, University of Tübingen, Germany

Abstract: Throughout history, pandemics such as the swine flu in 2009 and the emergence of COVID-19 in 2019 have led to the development of mathematical models, including compartmental models, to study disease spread. However, accurately estimating parameters during an epidemic can be challenging. Machine learning techniques like neural networks offer alternative solutions to predict and prevent disease spread. Physics-informed neural networks (PINNs) have been developed to incorporate the physics of the problem into NNs. In this regard, disease-informed neural networks (DINNs) were introduced to estimate the parameters of diseases, specifically COVID-19. Although DINNs are a robust procedure for effectively learning the dynamics of disease spread, the estimated parameters in this network are constant. This study extends DINNs to estimate time-varying parameters in COVID-19 using a sliding window approach.
Moreover, a significant challenge in informed neural networks is balancing the gradient flow dynamics of physics loss and forecast errors during training and at different epochs with a constant Œª. To address this issue, a new PINN called adaptive-PINN was developed to adaptively tune Œª at different epochs of training and to balance the gradients of the loss terms. In our study, the mean gradient statistics procedure for loss weighting in adaptive DINNs (ADINNs) is investigated. Our novel sliding window adaptive disease-informed neural network model (SW-ADINN), which is implemented for COVID-19, is based on the last three major variants of concern in Germany (alpha, delta, and omicron). We utilized a Susceptible-Exposed-Infected-Hospitalized-Recovered-Death (SEIHRD) model that considers not only confirmed cases, recovered, and deceased individuals, but also the unreported exposed state and hospitalized patients. PINNs and DINNs can handle missing data if the unreported variable is present in at least one other equation of the model and an initial value for the missing data is given. However, increasing the number of iterations is necessary to obtain reliable results. To avoid this, we first estimated the number of unreported cases (E) by fitting the model to the reported values and numerically solving it by the least squares method. To achieve a better estimation, we split the dataset into sub-datasets and obtained the best-fitted model for each sub-dataset by assessing a set of initial random parameters. The concatenation of the best approximations for E is used as the input for the NN. We set the optimal time window size as 90, which we determined after testing various window sizes and interpolation methods. In a 90-day interval, the neural network is able to learn that the solutions are not strictly increasing or decreasing, which is crucial for modelling the various waves and fluctuations of the virus. Our proposed SW-ADINNs model has a promising capability of estimating the multi-peak behaviour of the training data. To check the potency of the model in a smaller range of multi-peak behaviour, we

examined the model during the Omicron waves with three sub-lineages and several peaks in a shorter time. The fitted outcome of compartments with the root mean square percentage error (RMSPE) of 12.12% in 272 days demonstrated the validity of the model. Additionally, we utilized the SEIHRD model to predict the unobserved compartments using SW-ADINNs. We solved the model with the last seven sets of parameters captured by SW-ADINNs for a period of 18 days ahead. During the forecast period, RMSPE between the mean of the seven solutions and the unobserved values for those 18 days was 7.75% on the last three major variants of concern in Germany and 3.85% on the lineages of Omicron.

Importantly, the forecast period was not included in the model training, indicating the ability of the STW-ADINNs model to generalize to future scenarios. The small RMSPE of between the mean of 7 solutions and the unobserved values of the 18-day forecast period highlights the accuracy of the model and its potential for predicting the spread of COVID-19 in real-world situations. These promising results suggest that our novel approach combining the SEIHRD model with SW-ADINNs may serve as an effective tool for decision-makers to take proactive measures against COVID-19 and other emerging infectious diseases.

3.13 Establishing a trustworthy signalling entropy calculation for biological processes analysis

Ana Stolnicu

University of Ulm, Germany

Abstract: Ana Stolnicu+*, Nensi Ikonomi+*, Johann M. Kraus+, Hans A. Kestler+

Insights into the mechanisms underlying complex biological systems, including cell differentiation, cancer initiation, and development, may be acquired by computing the signalling entropy, given its suitability for capturing variations in multiple processes[1,2]. In the molecular environment, this metric combines data from a protein interaction network (PIN) with information acquired from pattern profiles in order to estimate the biological changes within a system. It is assumed that the available PINs contain a significant amount of inaccuracy, both in terms of false positives and false negatives. In yeast-two-hybrid screens, the false positive and false negative rates have been reported to be as high as 64% and 43%-71%, respectively[3]. Several correction procedures have been proposed to boost the accuracy of a PIN by filtering out irrelevant information, adjusting for experimental bias, and combining data from several sources[4-7]. Aim: While computing the signalling entropy, we aim to simultaneously capture changes in gene expression across different states and minimise the variations in the measure caused by the network's architecture. Our research is focused on identifying potentially reliable interaction networks that may be used to compute the signalling entropy of a biological system by lowering the number of false positive protein interactions. Methods: To accomplish this objective, we begin by investigating how the signalling entropy of a synthetic gene expression dataset is influenced by various percentages of randomly inserted protein-protein interactions in the network. Here, we used a yeast PIN to simulate many networks, each of which features from 10% to 90% of new connections. Next, after integrating distinct PINs with real datasets, we analysed the resulting alterations among the signalling entropies. In this context, we considered the interactions from well-known databases like Pathway Commons (PC)[8], STRING[6,7] and the Biological General Repository for Interaction Datasets (BioGRID)[9], together with the union and the intersection of those, in combination with cell differentiation or cancer datasets. Results: We observed that the entropies decrease as the number of randomly added interactions increase up to a certain point when they begin to rise again, at roughly 50% of newly inserted edges. After adding from 50% to 90% of interactions, the network possibly takes on the characteristics of a random graph, where the nodes all have approximately the same number of connections. When analysing cell differentiation and tumour data, we examined the statistical significance of the entropy differences among the samples. We observed that with BioGRID and PC, the outcomes were better than with STRING and that after applying some filtering, STRING itself yielded superior results. On the other hand, when more interactions were included in the network, for instance, we considered the union of the investigated PINs, the results were greater than after applying a

correction. It's possible that the overall attainability of the connections improves when BioGRID and PC are added to the interactions of STRING, suggesting that the first two databases may be more curated while the latter contains additional predicted interactions, which might be considered false positives. Conclusions: Due to the limited overlap across the databases documenting observed or predicted protein interactions, the generation of a trustworthy interaction network would be essential for computing the signalling entropy. Based on our findings, we hypothesise that the STRING database, more than BioGRID and PC, requires filtering for misleading protein links. Nonetheless, a viable approach would be the unification of the protein interactions from multiple databases which might reduce the rate of false negative interactions while simultaneously reducing the impact of false positive ones on the outcome. Additional research into other protein-interaction databases and different types of perturbations might be required to further analyse the attainability of the protein networks.

References

Banerji, C. R. S. et al. (2013). Cellular network entropy as the energy potential in Waddington's differentiation landscape. Scientific Reports, 3(1):3039.

Cheng, F. Et al. (2016). Investigating cellular network heterogeneity and modularity in cancer: a network entropy and unbalanced motif approach. BMC Systems Biology, 10(S3):65.

Kuchaiev, O. et al. (2009). Geometric De-noising of Protein- Protein Interaction Networks. PLOS Computational Biology, Volume 5(8) :e1000454

Csárdi, G. and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal Complex systems.

Yu, G. et al. (2010). Gosemsim: an r package for measuring semantic similarity among go terms and gene products. Bioinformatics.

Franceschini, A. et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Research, 41(D1):D808,ÄìD815.

Szklarczyk, D. et al. (2015). STRING v10: protein,Äìprotein interaction networks, integrated over the tree of life. Nucleic Acids Research, 43(D1):D447,ÄìD452.

Cerami, E. et al. (2011). Pathway commons, a web resource for biological pathway data. Nucleic Acids Research.

Livstone, M. S. et al. (2011). The biogrid interaction database. Nature Precedings.

+ Institute of Medical Systems Biology, Ulm University, Ulm 89081, Germany * Equal contribution

3.14 ADTD - Adaptive Digital Tissue Deconvolution.

Franziska Görtler

Computational Biology Unit, Department of Informatics, Department of Biological Sciences, University of Bergen, Norway

Abstract: The comprehensive molecular characterization of patients is a key ingredient to facilitate personalized medicine. In cancer research, for instance, there is increasing evidence that cancerous cells shape their environment, making the tumor microenvironment (TME) a complex mixture of diverse cell types, including immune cells. In fact, many modern therapeutic strategies directly involve the immune system; e.g., check-point inhibitors break the communication lines between tumor and immune cells, leading to a re-activation of the immune system. Consequently, analyses of cancer as well as a multitude of other diseases have to account for the cellular composition.

Modern experimental techniques such as single-cell RNA sequencing (scRNA-seq) or Fluorescence-activated Cell Sorting (FACS) facilitate direct measurements of cellular compositions. However, frequently these are not readily available due to high costs or experimental efforts. Data from large prospective trials or from large public repositories such as The Cancer Genome Atlas Program (TCGA) predominantly provide high-dimensional molecular patient data which were generated from bulk tissue. Bulk data, in contrast to scRNA-seq, are measuring a complex mixture of diverse molecular contributions of multiple individual cells. Thus, analyses of these data are confounded by the underlying cellular contributions, limiting our ability to resolve the causal reason of an observed phenotype. In addition, bulk profiles likely contain signals of cell types not originally considered for model development; for instance, hidden contributions from cancer cells likely do not share any similarity with the considered reference profiles.

Although many different approaches were meanwhile suggested for cell-type deconvolution, they rarely account for hidden cell type contributions and cellular environmental effects on the reference profiles. We systematically address the issues of hidden contributions in mixtures and adaption of immune cells to the tumor microenvironment within a single approach named Adaptive Digital Tissue Deconvolution (ADTD). First, ADTD builds on our previous work on Digital Tissue Deconvolution (DTD) and uses gene weights for optimized deconvolution. Second, it augments the deconvolution by estimates of a background profile together with corresponding cellular proportions across a set of investigated bulks. Third, ADTD adapts its reference profiles to the investigated bulks. Intriguingly, the latter also resolves cell-type specific regulation from bulk transcriptomics data. ADTD was verified in simulations based on single-cell RNA sequencing data of healthy and breast cancer tissue, and further, we demonstrate its application to resolve cell-type specific molecular differences between breast cancer subtypes.

With our (A)DTD algorithm we are getting informations concerning:
1. Cell type distribution for the cell types of interest in the mixtures.
2. Information concerning the cell type(s) not covered by the reference profiles quantitatively and qualitatively (amount and profile of the uncovered part).
3. Information on the adaptations to the tumor microenvironment for the regarded cell types. This information can lead to further understanding the interplay between cancer and immune cell types as well as leading new biomarkers for treatment options.

3.15 Gene-regulatory networks controlling cell fates downstream of TGFβ-signaling in MCF10A cells

Laura Strohmaier

Institute of Biomedical Genetics, University of Stuttgart, Germany

Abstract: The transforming growth factor-β (TGFβ) signaling pathway regulates different cellular outcomes and is involved in various diseases including autoimmune diseases, fibrosis, and cancer. Its dual role in tumorigenesis ranges from tumor suppressor at early stages to tumor promoter, inducing tumor progression at late stages. Upon TGFβ ligand-receptor interaction, SMAD transcription factors translocate to the nucleus and induce transcription of hundreds of target genes, thereby controlling cell cycle progression and cell fate decisions such as epithelial-mesenchymal-transition (EMT). As SMAD proteins show low DNA binding affinity, co-activators and co-repressors interact with the SMAD complex to direct it to its target gene promotor and control its activity. Depending on the SMAD signaling dynamics, TGFβ treated cells show different phenotypic outcomes (Strasen et.al, 2018). However, the translation of SMAD dynamics into different gene expression programs and the contribution of distinct co-activators and co-repressors resulting in different cell fate decisions are unknown. Here we use a genome-wide ordinary-differential-equation (ODE) model to describe the gene expression kinetics of target genes. A simple ODE model only considering time-resolved microscopy data of the nuclear/cytoplasmic SMAD2 ratio described $\sim 60\%$ of differentially expressed target genes. Genes sharing delayed or biphasic expression kinetics, suggesting the involvement of complex regulatory networks, were unexplainable by the simple ODE model. To describe the expression kinetics of these candidates, it is likely required to incorporate complex controls through co-activators and co-repressors into the model. To this end, we identified nine co-factors of SMADs whose transcription is induced in response to TGFβ treatment. Knockdown experiments of these co-factors followed by bulk RNAseq analysis detected SNAI1, SNAI2, SKIL, SKI, RUNX1 and JUNB as key regulators, influencing target gene expression at different time points post-stimulation. These candidates will be included in the ODE model and analyzed by scRNAseq besides phenotypic assays such as EMT characterization. Taken together, we have used a genome-wide systems biology approach to determine key regulatory candidates involved in SMAD signal decoding and cell fate decision.

3.16 MultiGML: Multimodal Graph Machine Learning for Prediction of Adverse Drug Events

Sophia Krix

Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Sankt Augustin, Germany

Abstract: Adverse drug events constitute a major challenge for the success of clinical trials. Several computational strategies have been suggested to estimate the risk of adverse drug events in preclinical drug development. While these approaches have demonstrated high utility in practice, they are at the same time limited to specific information sources and thus neglect a wealth of information that is uncovered by fusion of different data sources, including biological protein function, gene expression, chemical compound structure, cell-based imaging, etc. In this work we propose an integrative and explainable Graph Machine Learning approach (MultiGML), which fuses knowledge graphs with multiple further data modalities to predict drug related adverse events. MultiGML demonstrates excellent prediction performance compared to alternative algorithms, including various knowledge graph embedding techniques. MultiGML distinguishes itself from alternative techniques by providing in-depth explanations of model predictions, which point towards biological mechanisms associated with predictions of an adverse drug event.

4 Poster presentations: session I

4.1 Eulerian Parameter Inference: A new paradigm in probabilistic inference based on a random variable transformation

Vincent Wagner

University of Stuttgart, Germany

Abstract: At its core, nearly every modeling task requires the estimation of model parameters from experimental data. Probabilistic inference, in contrast to point estimation, not only returns one value per parameter, but a whole distribution, thereby inherently estimating the identifiability and uncertainty associated with the parameters.

We here present Eulerian Parameter Inference (EPI), a probabilistic inference method based on the concept of random variable transformation. Formally, the input of EPI is a deterministic simulation model that can be evaluated for individual parameters and a data distribution that is assumed to be generated by an underlying parameter distribution. EPI estimates this parameter distribution from an observed output distribution. In practice, we often deal with data in form of individual samples, from which we have to estimate a distribution by using established density estimation approaches. EPI transforms the estimated data distribution into a parameter distribution that is consistent with the observed data. This can be done by only using point-wise evaluations of the simulation model and approximations of its derivatives with respect to the parameters, which directly returns a density value in the parameter space. In particular, we do not require an explicit formulation of the inverse mapping from the output to the parameters. If grid-based evaluations are not feasible for higher dimensions, Markov-Chain Monte-Carlo sampling techniques can be used to create a sample from the parameter density.

Except for density estimation techniques and Markov-Chain Monte-Carlo sampling, EPI is parameter-free and provably correct if the parameter inference problem is well-posed.

Besides academic examples, we apply EPI to a diverse set of models ranging from algebraic equations over chaotic maps to ordinary differential equation systems, thereby proving its practical applicability.

4.2 CRESCENT: Stratification of Chronic Pain Patients using Machine Learning

Sophie Thiesbrummel

Bielefeld University, Germany

Abstract: Medical secondary care units often struggle with long waiting times for patient appointments. In rheumatology, a waiting time of several months until a patient is first seen can lead to impairments in well-being and health: A delayed initiation of an effective anti-inflammatory disease modifying treatment may have serious health consequences, e.g. irreparable joint damage in patients with rheumatoid arthritis. In addition, patients suffer from persistent pain while not receiving adequate treatment.
Muskuloskeletal pain may originate broadly from three prevalent conditions: an inflammatory rheumatic disease, osteoarthritis and chronic pain syndromes. Long waiting times occur when primary care units experience difficulties in attributing pain to a specific disease and thus refer a large number of patients to rheumatology, many of them actually suffering from chronic pain conditions such as fibromyalgia.
The aim of this study is to improve the stratification of inflammatory rheumatic diseases, chronic pain disorders and osteoarthritis, being well aware of the fact that these conditions are not mutually exclusive, but may occur simultaneously. To address this research question, we applied to all patients presenting to rheumatology a simple questionnaire targeting patients with fibromyalgia and measured C-reactive protein, a laboratory blood marker for inflammation. Based on these two easily measurable variables we attempt to improve the disease stratification. Furthermore, we use machine learning such as binary or multi-label classification and Bayesian approaches and analyse patient data under special consideration of vital signs, laboratory values and questionnaire information. Improving the stratification of pain patients can help to correctly identify patients in need of anti-inflammatory therapy, which will increase the patients,Äô life quality.

4.3 preCICE: A sustainable and user-friendly coupling ecosystem for partitioned simulations

Ishaan Desai

Institute for Parallel and Distributed Systems, University of Stuttgart, Germany

Abstract: What if we could build complex and efficient multi-physics simulations by easily plugging together the tools we already have at hand? Prototyping such complex and efficient simulations has been possible since longer with the free/open-source coupling library preCICE, which provides sophisticated numerical coupling methods and scalability on ten thousands of compute cores [Bungartz et al., preCICE - A fully parallel library for multi-physics surface coupling, CompFluids, 2016, https://doi.org/10.1016/j.compfluid.2016.04.003].
Today, it is significantly easier to design partitioned simulations by selecting from a list of ready-to-use integrations with widely-used simulation codes, following a unified and actively maintained online documentation, and connecting with an expanding community of users and contributors, counting more than 100 research groups worldwide. This growing ecosystem of subprojects creates challenges in structuring and automating the development, documentation, testing, and continuous integration from unit to system level. This poster will present the challenges and lessons learned in growing preCICE from an as-is coupling library to a sustainable, batteries-included ecosystem [Chourdakis et al., preCICE v2: A sustainable and user-friendly coupling library [version 1; peer review: 2 approved] Open Res Europe 2022, 2:51, https://doi.org/10.12688/openreseurope.14445.1].

4.4 Common Data Models for Genetic Rare Diseases as a Basis for Personalized Medicine

Najia Ahmadi

Institute for Medical Informatics and Biometry (IMB), TU Dresden, Germany

Abstract: Rare Diseases (RD) are numerous, geographically scattered, and mostly hereditary (1-6). In the United States, a disease affects less than 200,000 people and in Europe 5 per 10,000 people is considered rare (2). Around 7,000 RDs are known to affect 3.5,Äì5.9% of the population (2,3). It is challenging to recognize genetic disorders solely based on clinical features and an accurate diagnosis of such a genetic RD is only possible via discerning the precise molecular cause, i.e., a genotype that may explain the clinical phenotype (5). However, because of the genetic and phenotype variability associated with RDs, the lack of in-depth knowledge about every potential gene variation, and the individual patient journey, the diagnosis of these diseases require a more individualized approach. Here, a Common Data Model (CDM) that can facilitate comprehensive documentation and integration of both genotype and phenotypic information of patients is vital to derive new patient-centric insights. One prominent model is the Observational Medical Outcomes Partnership (OMOP) CDM, which is well-suited for Harmonizing and analyzing longitudinal data and serves as a basis for predictive models. Here, we demonstrate in our works i) how a CDM needs to be developed and applied in the clinical setting and ii) to what extent a CDM can be the basis of Artificial intelligence (AI) based predictions.

i) In a recent scoping review (Sc-R), we identified the state-of-the-art methods to develop CDMs in the health domain (7,8). The articles included showing the diversity of methods used to develop a CDM in different healthcare domains. We highlight the need for more specialized CDM development methods for medical contexts and propose a development process that will ease the conception of new CDMs. In the next step as part of the SATURN project (9), a generalizable OMOP-based RD-CDM will be applied for three use cases Endocrinology, Gastroenterology, and Pneumology (two university clinics, Dresden and Frankfurt).

ii) We also investigated the suitability of OMOP CDM for genomic data sharing and analysis. We found out that when Variant Call Format (VCF)-based data is mapped to FHIR, it can be transferred to OMOP with 100% success (10). Importantly, OHDSI offers a platform called ATLAS (11) and other tools for cohort definition and analysis using AI methods. According to our own works and in line with current literature, OMOP CDMs lead to standardized data-driven studies for multiple clinical sites and enable a solid basis for AI-based predictions (e.g., Patient level Prediction [PLP]) (12,13).

In summary, CDMs for genetic RDs may play an essential role in facilitating early prediction, diagnosis, and improvement of personalized cancer care, as well as biomarker

discovery.

References:

1. Zoch M, Sedlmayr B, Knapp A, Bathelt F, Helfer S, Schmitt J, et al. Interdisziplinärer Versorgungspfad und mögliche IT-Unterstützung für Menschen mit Seltenen Erkrankungen in Deutschland. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen. 2021;165:68-76.
2. Wakap S, Lambert D, Olry A, Rodwell C, Gueydan C, Valérie L, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. European Journal of Human Genetics. 2019 Sep 16;28.
3. Aymé S, Schmidtke J. Networking for rare diseases: a necessity for Europe. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2007 Dec;50(12):1477-83.
4. Bick D, Jones M, Taylor SL, Taft RJ, Belmont J. Case for genome sequencing in infants and children with rare, undiagnosed or genetic diseases. J Med Genet. 2019 Dec;56(12):783-91.
5. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. Nat Rev Genet. 2018 May;19(5):253,Äì68.
6. Chu SY, Weng CY. [Introduction to Genetic/Rare Disease and the Application of Genetic Counseling]. Hu Li Za Zhi. 2017 Oct;64(5):11-7.
7. Ahmadi N, Zoch M, Kelbert P, Noll R, Schaaf J, Wolfien M, et al. Methods used in the development of Common Data Models for health data ,Äì A Scoping Review Protocol. 2022 Apr 27 [cited 2022 May 11]; Available from: https://osf.io/rza84/
8. Methods used in the development of Common Data Models for health data ,Äì a Scoping Review [Internet]. JMIR Preprints. [cited 2023 Jan 11]. Available from: https://preprints.jmir.org/preprint/45116
9. SATURN Projekt [Internet]. [cited 2023 Mar 10]. Available from: https://www.saturn-projekt.de/ 10. Peng Y, Nassirian A, Ahmadi N, Sedlmayr M, Bathelt F. Towards the Representation of Genomic Data in HL7 FHIR and OMOP CDM. Stud Health Technol Inform. 2021 Sep 21;283:86-94.
11. ATLAS: Home [Internet]. [cited 2023 Jan 2]. Available from: https://atlas-demo.ohdsi.org//home
12. A comparative patient-level prediction study in OMOP CDM demonstrates advantages for native and web-based implementations [Internet]. 2023 [cited 2023 Mar 9]. Available from: https://www.researchsquare.com
13. Ahmadi N, Peng Y, Wolfien M, Zoch M, Sedlmayr M. OMOP CDM Can Facilitate Data-Driven Studies for Cancer Prediction: A Systematic Review. International Journal of Molecular Sciences. 2022 Jan;23(19):11834.

4.5 On the role of stem cell-niche interactions in AML patients: insights from different mathematical modeling approaches.

Ingmar Glauche

IMB / TU Dresden, Germany

Abstract: Background: Acute myeloid leukemia (AML) is a severe disorder of the blood forming system which is commonly treated with cytotoxic drugs (e.g. Cytarabine and Anthracyclines). The continuous monitoring of measurable residual disease (MRD) levels during and after therapy has developed into an essential tool to control treatment success. AML progression and treatment outcome do not only depend on the underlying mutations, but also on the cell of origin, the local microenvironment, metabolic conditions or the individual immune response. Especially the roles of the local microenvironment and the resulting stem cell-niche interactions have been discussed in recent years and it is yet unclear how these are linked to disease dynamics and treatment response. While mathematical models can add to a quantitative understanding of disease development and therapy response, differences in their underlying assumptions can lead to diverging interpretations about potential targets within the stem cell-niche interaction.

Aim: We aim to compare two different mathematical models of AML with respect to their ability to describe treatment response data and analyze how different assumptions about stem cell-niche interactions affect strategies for therapy optimization.

Methods: We compare two currently published models of AML progression and treatment, one developed by our group (Hoffmann et al., J. R. Soc. Interface 2020), one presented by Pedersen et al. (Stem Cells, 2023). We assess their suitability to describe disease dynamics and treatment outcomes by fitting each model to 275 clinical MRD time courses and respective therapy cycles. Furthermore, we perform a systematic screening with univariate alterations for relevant parameters to evaluate their influence regarding velocity and depth of remission during induction therapy as well as the length of remission thereafter.

Results: We fit both models to optimally describe the clinical data set. Our results show that both models are capable of describing significant reductions of disease levels (up to MRD negativity) during therapy, although they adhere to the assumption that the cytotoxic kill acts equally on proliferating healthy and leukemic cells. Our analysis of the remission behavior after induction therapy suggests that a moderate activation of both healthy and leukemic quiescent cells during cytotoxic treatment can potentially improve the clinical outcome as more of the leukemic cells are actively targeted. Interestingly, we observed that the two models produce conflicting results when it comes to altering the potential of the leukemic cells to revert from an active proliferating into a more quiescent state. While in the model

of Hofmann et al. inhibition of the reattachment has a minor negative effect on the duration of remission, there is an positive and more pronounced effect predicted by the model of Peterson et al. This feature is due to the assumption that subsequent stem cell divisions can only happen if the cells have intermediately reentered a quiescent state. It needs to be validated in an experimental or in a clinical setting whether this assumption can be uphold, especially for the rather unregulated proliferation of leukemic cells. Such empirical results are essential to support or dismiss certain model assumptions and the resulting quantitative predictions.

Conclusion: Our approach demonstrates that systematic model analysis can be used to infer mechanisms of stem cell-niche interactions and to identify potential interventions in order to optimize treatment strategies. Our results indicate that such interventions may target the stem HSC niche interaction during treatment or after the induction phase.

4.6 Disease Map Integration into a Systems Medicine Workflow - A Use Case

Julia Scheel

University of Rostock, Germany

Abstract: Background: Disease Maps are a community-driven Systems Medicine approach to represent and model disease mechanisms. Disease maps serve both as a knowledge base and analytical tools for advanced Omics data integration and interpretation, as well as hypothesis generation. Although this approach offers a multitude of options, most Disease Maps are not fully leveraged by their target research community, yet.
Objective: This work highlights the application of the placenta specific Disease Map NaviCenta [1] to enrich proteomics analysis in the case of placental dysfunction and presents a versatile computational workflow. From a medical perspective, the placenta remains the key to maternal diseases, such as Pre-eclampsia (PE) and Intrauterine growth restriction (IUGR), however the pathophysiology of these disorders remains elusive and a NaviCenta enriched workflow might contribute towards the overall understanding.
Methods: Term placentas were collected from nulliparous preeclampsia and intrauterine growth restriction pregnancies, along with age and body mass index matched healthy controls. Quadrupole-orbitrap mass spectrometer was used in data dependent analysis mode to identify differentially expressed proteins between conditions. In brief, the finally developed workflow contains Perseus [2] , ClusterCompare [3], NaviCenta Disease Map [1] , and the TriplexRNA database [4]. A core regulatory network based on log2 fold change and rewiring scores, including regulating miRNAs and transcription factors, was constructed using Cytoscape DyNet.
Results and Conclusion: Our results show how Disease Maps can be utilized to identify new potential regulatory proteins in PE and IUGR pathology. The identified miRNA triplexes further open a novel avenue of investigation for placenta research. We further demonstrate that embedding the use of NaviCenta into a larger computational workflow supports the generation of more meaningful results and will contribute towards placenta research.
References:
[1] https://www.sbi.uni-rostock.de/minerva/index.xhtml?id=NaviCenta
[2] https://maxquant.net/perseus/
[3] G Yu, LG Wang, Y Han, QY He. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS: A Journal of Integrative Biology. 2012, 16(5), 284-287
[4] Schmitz et al. NAR https://triplexrna.org/

4.7 Splicing decisions in a three-exon minigene: Insights from mathematical modeling

Panajot Kristofori

Institute of Biomedical Genetics, University of Stuttagrt, Germany

Abstract: Pre-mRNA splicing is a complex process that generates RNA isoforms with different exonic compositions. Dysregulated splicing can lead to the retention of introns and impaired translation, but the factors that regulate splicing fate remain poorly understood. To shed light on this process, we employed a systematic mutagenesis assay on a three-exon minigene (RON) and analyzed genome-wide RNA-seq data from the ENCODE project.

Our results revealed a non-monotonic, U-shaped relationship between splicing efficiency and the inclusion level of the alternative exon, with the lowest splicing efficiency observed at intermediate inclusion levels. We developed a kinetic post-transcriptional model to test hypotheses using extensive mutagenesis data of all RNA isoforms of RON minigene, which suggested that partial recognition by the initial spliceosome machinery underlies the non-monotonic relationship between splicing efficiency(Intron removal) and inclusion of the middle exon. This dependency was also accurately predicted by the model in different cell lines and for various RNA binding protein knockdowns.

Furthermore, our findings indicate that U1 binding in the exon is enough to prevent skipping of the alternative exon, contributing to the non-monotonic behavior observed in splicing efficiency. Our study provides insights into splicing regulation and highlights the importance of quantifying all splicing outcomes, including unproductive ones, to gain a comprehensive understanding of splicing regulation. By considering both productive and unproductive splicing outcomes, our work suggests potential targets for therapeutic intervention and contributes to the growing understanding of the complex molecular mechanisms underlying pre-mRNA splicing and the role of alternative splicing in gene expression regulation.

4.8 Identification of genome-wide expression differences between patient-matched intra- and extracranial melanoma metastases pairs using Hidden Markov Models

Theresa Kraft

IMB TU Dresden, Germany

Abstract: Overcoming the therapy resistance of melanoma brain metastases (intracranial metastases) is a crucial step to improve therapy response of affected patients. Therefore, key molecular mechanisms that distinguish treatment-resistant intracranial from treatable extracranial metastases need to be determined. Melanoma metastases from the same patient are more similar to each other than to metastases in the same tissue from other patients. This strengthens the need for a personalized analysis. Thus, we compared RNA-sequencing data of 16 intracranial metastases with 21 patient-matched extracranial metastases in a personalized way using a three-state Hidden Markov Model (HMM) with state-specific Gaussian emission densities to identify altered genes for each individual metastases pair. An in-depth analysis of the predicted gene expression alterations across all patients led to three major findings: (i) especially cytokine signaling, calcium signaling and ECM-receptor interaction were most frequently altered, (ii) immune-relevant genes showed most frequently decreased expression in intra- compared to patient-matched extracranial metastases, and (iii) intracranial metastases were associated with a brain-like phenotype expression program. These general findings are in good accordance with other studies comparing intra- and extracranial melanoma metastases. Moreover, a candidate gene set was identified that includes 103 genes that were differentially expressed in the same manner in 69% (11 of 16) of all patients. This gene set contains known immune-relevant genes (e.g. CCL19, CLEC10A, CD8B, CD79A) and potential cancer therapy targets (e.g. CLEC10A, CXCL11, GPR68). Overall, our study contributes to a better characterization of key pathways and genes that could play a role in therapy resistance of melanoma brain metastases.

4.9 Developing a data-integrated simulation framework to predict responsiveness to targeted cancer therapeutics

Fabian Klötzer

Institute of Cell Biology and Immunology, University of Stuttgart, Germany

Abstract: The wide variety of cancer types represents a big challenge for cancer therapy. Case-to-case heterogeneity demands individualized treatments that can vary in choice, dosing and combination of anti-cancer agents. Modeling the sensitivity of tumor cells to anti-cancer therapeutics can provide a cost effective in silico estimation of treatment efficacy. Therefore, we are developing a data-integrated, mechanistic ODE model to predict mitochondrial outer membrane permeabilization (MOMP) within cancer cells. MOMP is a crucial step in executing apoptotic cell death, resulting from complex interactions of the druggable BCL-2 protein family. Simulating BCL-2 protein interactions allows for introduction of mechanistic domain knowledge into the model. This improves the model specificity, allows to investigate regulatory mechanisms that have not been fully understood yet, and can test for the sensitivity to anticancer agent combinations (BH3 mimetics). As a first step, we determined treatment recommendations for 325 different cancer cell lines using a prototype model implementation. Input data were averaged protein amounts from cell populations, thus so far neglecting the heterogeneity between individual cells of the same population. To incorporate heterogeneity in the model, we will use in house experimental single cell data to re-parameterize the model to simulate cell populations at single cell resolution. This interplay between data-based modeling and model-based experimental design is hoped to increase the explanatory power of the model and the accuracy of optimal treatment predictions.

4.10 Unravelling of tumor microbiome and its implication for oncogenic signalling

Linh Dang

Department of Medical Bioinformatics, University of Medical Center Göttingen, Germany

Abstract: Malignant tumors have been found to harbor bacteria and fungi, which can impact tumor proliferation, progression, and response to therapy. To better understand this phenomenon, we plan to comprehensively characterize the tissue and microbiome of young colorectal cancer (CRC) patients, identify oncogenic microbiome signatures, and investigate their influence on oncogenic signaling pathways that contribute to tumor development and progression. Our approach will involve several steps. First, we will perform 16S rRNA sequencing of tumor and adjacent normal tissues from patient samples. Second, we will conduct genome-wide DNA methylation and host RNA sequencing analysis of the tumor tissues to identify microbiota-sensitive epigenetic and transcriptomic signatures, and we will associate bacterial and fungal clusters with inflammatory processes and immune responses. Finally, we will integrate microbiome and multi-omics datasets to uncover a small set of candidate genes involved in inflammatory and immune responses, as well as key pathways.

4.11 Long-read genome sequencing and RNA sequencing resolve an intronic LINE-1 insertion in the APC gene in a so far unsolved adenomatous polyposis family

Alexandra Anke Baumann

Institute for Clinical Genetics, University Hospital Carl Gustav Carus at the Technische Universität Dresden, Germany

Abstract: Introduction: Familial adenomatous polyposis is caused by pathogenic variants in the tumor suppressor gene Adenomatous Polyposis Coli (APC). The genetic cause for a family with familial adenomatous polyposis across 4 generations could not be identified by cancer gene panel sequencing (94 genes), array-CGH, exome sequencing, and investigation of specific intronic variants. Methods: Long-read genome sequencing (PacBio), short-read genome sequencing, short-read RNA-sequencing, and further validations were performed in different tissues of multiple family members. Results: Long-read genome sequencing (PacBio) of one family member resolved a 6 kb insertion of a LINE-1 element in intron 8 of the APC gene that could be detected but not deciphered with short-read genome sequencing (Illumina). The structural variant was validated on DNA level in several family members and investigation of RNA revealed aberrant splicing.
Conclusion: This case study supports the utility of long-read DNA sequencing techniques and complementary RNA approaches to tackle unsolved cases.

4.12 Predicting the risk of cytokine release syndrome induced by cancer immunotherapies using risk factors learnt from COVID-19.

Philippe Robert

University of Basel

Abstract: Background: Cancer immunotherapies such as CAR-T-cells or bispecific antibodies can induce over-activation of the immune system, leading to Cytokine Release Syndrome (CRS). The risk for CRS can be dose-limiting in the application of cancer immunotherapies. Similarly, infections, such as Covid-19, can cause CRS by uncontrolled immunopathology.
Hypothesis: CRS immunopathogenesis is multifactorial and difficult to define. High IL-6 serum levels poorly predict the CRS severity. Quantitative integration of patient-specific factors to predict the individualized risk of developing CRS allows adapting the dosage of cancer-immunotherapeutics while minimizing adverse effects and, therefore, can improve the individual benefit-risk ratio and health outcomes for patients.
Methods: We mine large longitudinal health records of patients with CRS across multiple conditions to delineate risk factors of CRS. We integrate individualized CRS risk factors into a mathematical mechanistic model for CRS onset.
Results: Using a cohort of 39 000 Covid-induced CRS, we isolated CRS risk factors that align with risk factors of CRS during cancer immunotherapies. We developed a mechanistic model for cytokine dynamics and CRS onset during SARS-CoV2 infection or cancer immunotherapy. We run /in silico/ trials on virtual patients and predict /in silico/ the impact of individualized treatment on reducing adverse effects.
Conclusions: We show a proof of concept that risk factors predictive of CRS can be identified on large-scale patient datasets to predict individualized drug effects.

4.13 A hybrid population model enables the study of single-cell heterogeneity in an epigenetic memory system.

Viviane Klingel

University of Stuttgart, Institute for Stochastics and Applications, Germany

Abstract: Single-cells measurement techniques like flow cytometry allow heterogeneity in biological systems to be quantified efficiently. However, many modelling approaches currently cannot capture this behavior, as often only an average cell is covered in commonly used ODE models. Single-cell data can of course be reduced to its mean to be usable with these models, but this leads to a loss of information in the data and, more importantly, is only an accurate description if the data is close to a normal distribution. Especially in the case of bimodal distributions, which can for example occur in bistable systems, averages are poor descriptions of the data and lack the ability to reproduce important features of the system. The synthetic epigenetic memory system from Graf et al. (2022) is such a particular system. It is characterized by the ability to switch from an OFF- to an ON-state through a transient metabolic trigger. This ON-state is sustained via positive feedback based on DNA methylation. A large part of the cells can remember this state for many days, but eventually, more and more cells switch back to the OFF-state. In the experimental data, this is observable as two subpopulations, ON- and OFF-cells. We aim to capture this bimodality by a tailored model which describes heterogeneous single-cell trajectories. While many stochastic modeling approaches exist for similar problems, we choose a hybrid model here. This model combines the simulation speed of differential equations with a stochastic process describing cell division, as well as distributed parameters. The simulated population is then compared to the data using Gaussian mixture models. We show that the model is able to reproduce experimental single-cell data and, in addition, gives insights into mechanisms for the ON to OFF switch in individual cells.

4.14 Coupling of SBML Micromodels to PDE Macromodels for liver simulation

Steffen Gerhäusser

Institute of Structural Mechanics and Dynamics in Aerospace Engineering, Stuttgart, Germany

Abstract: Steffen Gerhäusser1, Lena Lambers1,2, Luis Mandl1, Matthias König3 and Tim Ricken1
1Institute of Structural Mechanics and Dynamics in Aerospace Engineering, Faculty for Aerospace Engineering and Geodesy, University of Stuttgart, Stuttgart
2Experimental Transplantation Surgery, Department of General, Visceral and Vascular Surgery, Jena University Hospital, Jena, Germany
3Systems Medicine of the Liver Lab, Institute for Theoretical Biology, Humboldt-University Berlin, Berlin, Germany

The liver provides essential metabolic functions for the human body and is therefore important for overall health. Nowadays, the modern western lifestyle affects the liver in a negative way, resulting in an increased number of liver tumors and non-alcoholic fatty liver diseases (NAFLD). In this context, computer-based simulations can function as digital twins of the patient to estimate the effect of drug treatment or surgical intervention. Using the human liver as an example, we present the coupling of micromodels for cell metabolism and function and spatially resolved macromodels for tissue deformation and perfusion. The metabolic processes at the cellular scale can be modeled as a biochemical reaction network resulting in a system of ordinary differential equations (ODEs). Such a system can be represented with the System Biology Markup Language (SBML) format and solved with the simulation engine libRoadRunner[1]. The calculated reaction rates of the SBML microsimulation are then communicated to the macroscale and included in the balance equations of the macromodel. As a representative example for such a microsimulation we developed the SPT-model, which describes the conversion of a substrate (S) to a product (P) while forming a toxic byproduct (T). The macromodel is set up of a system of coupled partial differential equations (PDEs), which describe the function-perfusion interaction between the tissue and the blood flow in a homogenized approach. This approach is based on the theory of porous media (TPM)[2] and is discretized with the finite element method (FEM). For the simulation of the coupled function-perfusion relationship on the macroscale, we are using the FEM framework FEBio[3]. Since two different frameworks are used for the different scales, software specific aspects of the coupling like performance and software sustainability are also addressed and alternatives to the conventional coupling are offered. Alternatively, the coupling can be performed with the coupling library preCICE[4], leaving the domain specific solvers in their designated programming language.

[1] Welsh, C. [,Ä¶], König, M., et al. Bioinformatics(39)1 2023
[2] Ehlers,W. Foundations of multiphasic and porous materials(2002)
[3] Maas SA, et al. Journal of Biomechanical Engineering, 134(1) 2012

[4] Gerasimos C., et al., Open Research Europe, 2022

4.15 Transcriptional regulator identification using prize-collecting Steiner trees

Gihanna Gaye Galindez

Division Data Science in Biomedicine, Technische Universität Braunschweig, Germany

Abstract: Transcription factors play important roles in maintaining normal biological function, and their dysregulation can lead to the development of diseases. Identifying candidate transcription factors involved in disease pathogenesis is thus an important task for deriving mechanistic insights from gene expression data. We developed Transcriptional Regulator Identification using Prize-collecting Steiner trees (TRIPS), a workflow for identifying candidate transcriptional regulators from case-control expression data for a disease or phenotype of interest. In the first step, TRIPS combines the results of differential expression analysis with a gene module identification step to retrieve perturbed modules comprising an expanded gene list. TRIPS then solves a prize-collecting Steiner tree problem on a gene regulatory network, thereby identifying candidate transcription factors that control the gene modules retrieved in the previous step. Our results indicate that the gene module identification step improves the identification of transcription factors putatively driving disease development. We compare TRIPS to relevant methods for inferring key regulator genes using publicly available disease datasets and show that the proposed workflow can recover known disease-associated transcription factors with high precision. Furthermore, we perform perturbation analysis to demonstrate the robustness of TRIPS.

4.16 From Modeling Synthetic Morphogenetic Systems to Design

Amatus Beyer

Institute for Stochastics and Applications (ISA), Germany

Abstract: Morphogenesis is the process by which cells form a defined shape and is central to the development of organized multicellular systems. The mechanism behind this process is not yet fully understood. Using synthetic biological building blocks and mathematical modeling, we aim to design structures capable of showing predefined patterns when seeded with cells. We start with a 2D structure based on a synthetic Notch morphogenetic system proposed by Toda et al., (2020) and aim to extend it to an output resembling a triangle. We plan to investigate this using 3D time-lapse microscopy and extend the framework to more complex 3D structures. This bottom-up engineering approach, called synthetic development, creates new opportunities to better understand morphogenesis and design complex tissues with desired functions not only for fundamental research but also for regenerative medicine.

5 Poster presentations: session II

5.1 Integrating ML within mechanistic models in Clinical Pharmacology: Necessity and Benefits of adding structure to Scientific Machine Learning Models

Diego Felipe Valderrama Nino

Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin, Germany; Bonn-Aachen International Center for Information Technology (B-IT), University of Bonn, Bonn, Germany

Abstract: INTRODUCTION
Over the recent years, machine learning applications (ML) in clinical pharmacology have grown substantially. The development of several ML techniques has been driven by the increasing availability of high dimensional data (e.g. biomarkers) and the potential for analyzing multimodal data. However, most of the current approaches use ML techniques as black boxes, and therefore there are only a few that have proposed interpretable architectures which integrate mechanistic knowledge.

Nowadays, there are approaches called Scientific Machine Learning (SciML) which integrate mechanistic frameworks and machine learning architectures to benefit from the advantages of both approaches. While an ML architecture helps to manage multimodal (and potentially high dimensional) data, unknown mechanisms, and missing values, the mechanistic framework enables the integration of the physiological constraints and domain knowledge in a set of equations.

In this work we focused on predictive models for pharmacokinetics (PK) and introduced a SciML approach that consists of a compartmental PK model coupled with a neural network term to describe an unknown absorption mechanism, while simultaneously estimating other parameters of drug distribution and elimination.

METHODS
We consider the problem of fitting pharmacokinetic data of a specific drug that exhibits a Weibull distribution and create insilco patient data following an NLME approach with both additive and proportional error. The data consisted of 800 patients that were allocated into eight dose groups ensuring the same number of subjects in each dose group and was simulated for 70 days. Two different PK sampling scenarios were considered:

Sampling 1: Concentrations available every 12 hours.
Sampling 2: Only seven available measurements per cycle in three cycles (only after the 1st, 3th, and 5th) at 0, 12, 24, 48, 72, 96, 144 hours after the dose was administered

For testing, two general schemes were considered: 1) prediction of a new patients

set and 2) prediction on the same patients set. For each set, the scenarios were defined as follows:

Extrapolation: Dosing schedule is the same as in the training data
No dose after: We stopped dosing after day 42
Dose missing: We assumed no dosed on the 49th day
New dosage amount: We change the dosage for all patients to 150mg and 230 mg.
To verify the impact of sample size, we trained all models with a different number of patients. We use MAPE to compare against state-of-the-art methods i.e Lu et. al. and Bräm et. al., and an adapted version of Bräm et. al. model which does not need to be initialized and which we call Bräm w/o 1st PKC.

MODEL
We used a standard ODE system to describe the drug dynamics in a depot and a central compartment. This scheme enabled us to capture the known physiology such as apparent clearance (CL) and apparent volume (V) in the architecture itself. In detail, a neural network was trained to predict the unknown absorption rate based on the data of the Depot compartment and the difference in time since the last dose was given to the patient. While CL and V were learned as parameters in each backpropagation step.

RESULTS
Compared models correctly predicted the mean concentration for each of the eight different dose groups throughout training and extrapolation when employing an "easy" scenario, i.e. a dataset created with an additive error for 680 patients and sampling 1. Only Lu et al. reveals a discrepancy in the last cycles, indicating that this model will perform worse for a lengthy extrapolation task.
The results for a more realistic scenario i.e., a dataset generated with a proportional error for 48 patients and sampling 2, reveal that the compared models cannot fit the training set with a small sample size, generating under fitted values or predicting similar concentrations for all doses close to the average of the first cycle concentration.
In contrast to such tendencies, our suggested PK-SciML model produces reliable predictions even when only a limited number of patients are used, and it extrapolates well to other sampling strategies. These results are consistent with additional experiments that were conducted using intermediate patient populations and sampling densities.
When testing on a set of new patients, where only our model and the Bräm w/o 1st PKC approach can be used, both models perform good when testing in an "easy" scenario, while only our approach is still accurate for the realistic scenario.
For the no dose scenario, Lu et. al. predictions follow the expected behavior while the Bräm et. al. approach predicts a steady state. In the dose missing scenario, both architectures under predict values at the time the dose is removed. In both scenarios for the Bräm w/o 1st PKC model we observe a lack of fit to the training data and a high prediction error. In contrast, our model accurately predicts the concentration in all the scenarios that had been tested. This is further demonstrated by the fact

that our model's estimates of CL and V are rather close to the data's original values.

DISCUSSION

This paper demonstrates that machine learning models perform better when pharmacokinetic modelling principles are incorporated into them. Even with less intensively sampled datasets, our PK-SciML model captures an unknown absorption process when other pure ML-based models fail. Moreover, because the apparent volume and apparent clearance are concurrently learned with the absorption rate, this architecture creates new opportunities for the creation of hybrid models that are more straightforward to comprehend than typical black-box machine learning models.

Another main remark is that the proposed architecture deals with relevant and realistic PK modeling scenarios, such as predictions for new situations and new patients where no data is available and therefore models that need to be initialized like Lu et al and Bräm et. al. cannot be used.

One significant limitation in our model is that it only predicts population trajectories, disregarding the individual effects of individuals on drug concentration. This might be resolved by learning random effects that enable the model to generate patient-specific concentrations and encoding some patient-specific data into the ODE system. We would leave this concept, nevertheless, for future attempts to improve.

5.2 Membrane confinement and curvature-induced stabilization of PIP3 lipid dynamics in three-dimensional Dictyostelium cells.

Marcel Hörning

University of Stuttgart, Germany

Abstract: The plasma membrane of cells is a dynamic structure that is responsible for maintaining the integrity of the cell while facilitating communication with the environment. The membrane is composed of lipids and proteins that are organized in a fluid mosaic structure. One of the key lipids that plays a critical role in cell signaling is phosphatidylinositol 4,5-bisphosphate (PIP2), which is converted to phosphatidylinositol 3,4,5-trisphosphate (PIP3) by the enzyme phosphoinositide 3-kinase (PI3K).

PIP3 is a key regulator of many cellular processes, including cell proliferation, migration, and survival. The dynamic regulation of PIP3 levels in the plasma membrane is therefore critical for proper cellular function. This regulation is achieved through a complex interplay between PI3K activity, PIP3 degradation, and PIP3 diffusion.

The diffusion of PIP3 in the plasma membrane is influenced by several factors, including the lipid composition of the membrane, the presence of proteins, and the curvature of the membrane. The diffusion of lipids in the plasma membrane has been shown to be highly heterogeneous, with the diffusion rate varying depending on the local environment. This heterogeneity is due in part to the presence of membrane proteins and lipid domains that can act as diffusion barriers or facilitators.

In addition to these factors, the curvature of the plasma membrane has also been shown to affect PIP3 diffusion. The curvature of the membrane can reduce the lateral diffusion of lipids by increasing the frequency of lipid-protein interactions and reducing the available diffusion space. This curvature-induced reduction in diffusion has been demonstrated experimentally using giant plasma membrane vesicles (GPMVs) with controlled curvature [1].

In single Dictyostelium cells, we discovered that waves of PIP3 are self-regulated through mechanical deformation of the plasma membrane, even in the absence of an actin network. This regulation occurs due to the underlying excitability and shape of the membrane, resulting in horizontal, vertical, and transient spot dynamics of PIP3 domains [2]. Additionally, we observed that the diffusion of these domains is strongly influenced by highly curved membranes, such as those found at the leading edge of protrusions and macropinocytic cups [3].

To gain a deeper understanding of these experimental observations, we conducted in-silico simulations of the mechanical effects of PIP3 lipid signaling using realistic, three-dimensional plasma membranes as observed in vitro [2], along with a validated

reaction-diffusion scheme [3, 4] and the stochastic tau-leap method [5]. Our simulations revealed that the PIP3 lipid signaling is regulated by variations in membrane deformation, excitability, and curvature-dependent diffusibility. We identified that the membrane confinement and curvature plays a critical role for the stabilisation of the PIP3 lipid dynamics.

[1] Yuan, F. Alimohamadi, H., Bakka, B., and Stachowiak, J.C. (2021). Membranfô bending by protein phase separation. PNAS 118, e2017435118
[2] Hörning, M., and Shibata, T. (2019). Three-dimensional cell geometry controls excitable membrane signaling in dictyostelium cells. Biophys. J. 116, 372,Äì382.
[3] Hörning, M., Bullmann, T. and Shibata, T.. (2021) Frontiers in Cell and Developmental Biology, 9, 2467
[4] Arai Y, et al. (2010) Self-organization of the phosphatidylinositol lipids signaling system for random cell migration. Proceedings of the National Academy of Sciences 107(27):12399,Äì12404.
[5] Cao Y, Gillespie DT, Petzold LR (2006) Efficient step size selection for the tau-leaping simulation method. The Journal of Chemical Physics 124(4):044109.

5.3 A Dynamic modeling approach predicting the heterogenous phenotypes of pancreatic neuroendocrine tumors

Nensi Ikonomi

Institute of Medical Systems Biology, Ulm University, Germany

Abstract: Background and Aims:
Pancreatic neuroendocrine tumors (PanNETs) are a rare tumor entity characterized by a largely unpredictable progression and increasing incidence in developed countries. The molecular pathways involved in the development of PanNETs are still poorly understood, and specific biomarkers are lacking. In addition, the heterogeneity of PanNETs makes their treatment challenging, and most approved targeted therapeutic options for PanNETs lack objective responses. A limited number of tumor drivers have been identified at the molecular level. Of these, menin-1 (MEN-1) is one of the most frequently reported first-hit mutations in patient cohorts, and its impact on PanNET development has also been studied in mouse models. We applied a systems biology approach integrating a dynamic modeling strategy, classification approaches, and patient expression profiles to predict and capture the heterogeneity of PanNETs. We also studied tumor progression and resistance mechanisms to clinically approved treatments such as the mammalian target of rapamycin complex 1 (mTORC1) inhibitors.

Methods:
Boolean network models are simple dynamical models where nodes (genes or proteins) can either be active (1) or inactive (0) at a given time point. Regulatory dependencies are described and integrated via the use of logical operators AND, OR, and NOT. Stable states of such models are called attractors and correspond to biological phenotypes. Since this modeling approach does not require the integration of frequently unavailable kinetic parameters, it allows for the construction and dynamic investigation of rather large molecular crosstalks. On these grounds, we built a Boolean network model based on an expert-based modeling strategy, integrating experimental evidence from a variety of cellular, in vivo, and human models and data. Our model integrates the main crosstalks involved in PanNETs, resulting in a final model of almost 60 nodes and 200 interactions. The resulting model was applied to simulate commonly reported PanNETs drivers in patient cohorts, such as MEN1, Death domain associated protein (DAXX), Tuberous Sclerosis (TSC), as well as wild-type tumors. These results were further integrated and validated through expression data analysis and a foreign body classification approach based on PanNETs expression data with the corresponding mutational landscape. In particular, the expression data were binarised and compared with the activities of the long-term behavior of the model. The foreign classification approach, instead, aimed to assess the degree of association of different PanNET mutations and compare these with the general phenotypes predicted by the model. We also investigated the possibility of predicting new tumor drivers, both as first and second hits after MEN1 loss. This has been done by investigating the 'basin of attraction'

of cancer-related long-term behaviors predicted by the model. Here, we focus on the biased activities of simulation starting states to predict potential tumor drivers.

Results:
Our approach was able to predict PanNET-related phenotypes under a variety of described PanNET-related mutations. Further integration of these results with those from expression data binarization and classification approaches helped capture this tumor type's heterogeneity. Furthermore, we were able to predict potential new tumor drivers and their impact on disease severity. Finally, we were able to evaluate the effect of mTORC1 inhibitors on PanNETs harboring different mutations, approaching a personalized medicine scenario. In this context, we were also able to link mTORC1 inhibition to the emergence of resistance phenotypes, for which we were able to predict a potential molecular mechanism.

Conclusion:
We have established a systems biology-based strategy to model a rare and heterogeneous tumor entity such as PanNETs. Here, our strategy not only predicts tumor phenotypes but also supports the hypothesis of tumour drivers, response to treatment and emergence of resistant phenotypes.

5.4 An amortized approach to non-linear mixed-effects modeling based on neural posterior estimation

Jonas Arruda

LIMES - University Bonn, Germany

Abstract: Non-linear mixed-effects models are a powerful tool for analyzing heterogeneous populations in biology, medicine, engineering, and related fields. However, fitting these models can be challenging, particularly when the description for individuals is complex and the space of latent variables is high-dimensional. We introduce a new parameterization approach for non-linear mixed-effects models that salvages the power of machine learning.
Our approach employs amortizing inference, based on invertible neural networks, to sample from the entity-specific posterior distribution, which is then used to infer the population-level parameters. We apply this approach to problems from cell biology and pharmacology and demonstrate that it is scalable and more efficient than established approaches, in particular when it comes to uncertainty quantification. Furthermore, the proposed approach provides results for complex, e.g. stochastic, models for which established methods fail. Our new approach represents a promising development in the field of non-linear mixed-effects modeling and has the potential to significantly enhance the analysis of heterogeneous populations.

5.5 Stochastic Modeling of Multivariant Infectious Disease

Vincent Wieland

LIMES - University Bonn, Germany

Abstract: Especially in the recent years understanding infectious disease dynamics to predict the evolution of an epidemic became of mayor interest for researchers. A commonly used approach are compartmental models based on ordinary differential eqations. However, this does not capture the intrinsic stochasticity of the dynamics. This work presents a formulation of compartmental models using stochastic differential equations instead. It extends this to a model for a viral disease with different variants of the virus. Additionally, it shows one way to assess the parameters, e.g. infection rates, of these models in a Bayesian inference framework. Namely, the method of particle filters is presented and applied to stochastic differential equations. As a real-world showcase we develop a stochastic differential equation model for the spread of COVID-19 in Ethiopia and apply the introduced methodology to obtain posterior distributions for the infection and recovery rates.
This work shows that stochastic differential equations can be used for modeling and inference for infectious disease in order to provide more realistic models, accounting for the intrinsic stochasticity especially in small populations. With the widely applicable particle filter at hand and rising computational capacities, we consider this approach as promising and applicable to many kinds of stochastic models in the field of modelling disease progression in the future.

5.6 Differential Entropy as an indicator of the development of pluripotent murine ICM cell populations.

Francisco Prista von Bonhorst

Université Libre de Bruxelles, Belgium

Abstract: During cell differentiation, a common population of pluripotent cells progressively acquires distinct cellular fates. The fate of a cell in this population is determined by the expression of specific transcription factors (TFs). This process is complex and involves a network of gene regulation and intra-cellular signaling. In addition to these elements, noise coming from different sources (transcriptional, cell-cell variability, etc. . .) is often required for cells to acquire distinct fates [1, 2].

In our work, we focus on the early mammalian embryogenesis. The first cell fate decision gives rise to two distinct populations of cells: the Inner Cell Mass (ICM) and the Trophectoderm (TE). The second cell fate decision arises when ICM cells give rise to Epiblast (Epi) and Primitive Endoderm (PrE) cells. While PrE cells will give rise to extra-embryonic tissues, Epi cells, are at the origin of all the cell types that will contribute to the future organism. At a genetic level, Epi cells are characterized by a high level of Nanog and low level of Gata6 transcripts, while PrE cells are characterized by a low level of Nanog and a high level of Gata6 transcripts [3, 4]. Other components also play a crucial role in this decision, such as the FGF/ERK signaling pathway, in particular Fgf4, that acts as an intercellular signal [5].

During differentiation cells start from a pluripotent state, develop to a progenitor state and then acquire the final differentiated state. Because this evolution is associated with the passage of the cells through the ‚Äôhesitant‚Äô progenitor state, it is expected that entropy measurements may capture the changes in uncertainty. Extensive research on intRA-cellular entropy of cell populations during development has been carried out, while literature is scarcer for intER-cellular entropy. Theoretically, intRA-cellular entropy should decreasealong these differentiation states and intER-cellular entropy should first increase between the pluripotent and progenitor state, to then decrease between the progenitor and differentiated state [6].

Our purpose is to calculate and study the temporal evolution of the entropy of the expression of various genes related to the differentiation process described above and, compare it with theoretical expectations. In addition, we aim at identifying the genes whose stochastic variations initiate Epi/PrE specification. Indeed, a group of still poorly identified pluripotency genes, coordinated by Nanog, initiate the evolution towards the Epi identity [4]. To this end, we analysed various datasets of single cell gene expression profiles (RTqPCR and scRNA-seq) during this cell fate decision, from various sources of mice embryos [3, 4, 7]. We then developed a methodology to calculate entropy from these various datasets, by converting data from these experiments into mRNA counts, or a quantity that is linearly proportional to it. This enabled us to fit the distributions of the mRNA counts to gamma

distributions, theoretically obtained considering the bursting property of transcription [8]. We then calculated the differential entropy of the fitted distributions and plotted their temporal evolution.

Our methodology leads to results that are in agreement with literature, with an increase, and then decrease in intER-cellular differential entropy levels for certain TFs that are related to the ICM to Epi/PrE transition cell fate, such as Sox2. IntRA-cellular differential entropy also shows a decreasing tendency along the differentiation process. We will include human datasets to compare the results between the two different species.

References:
[1] B. D. MacArthur and I. R. Lemischka, ‚ÄúStatistical Mechanics of Pluripotency,,Äù Cell, vol. 154, pp. 484,Äì 489, Aug. 2013. Publisher: Elsevier.
[2] C. Robert, F. P. v. Bonhorst, Y. D. Decker, G. Dupont, and D. Gonze, "Initial source of heterogeneity in a model for cell fate decision in the early mammalian embryo," Interface Focus, June 2022. Publisher: The Royal Society.
[3] G. Guo, M. Huss, G. Q. Tong, C. Wang, L. Li Sun, N. D. Clarke, and P. Robson, "Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst," Developmental Cell, vol. 18, pp. 675-685, Apr. 2010.
[4] N. Allègre, S. Chauveau, C. Dennis, Y. Renaud, D. Meistermann, L. V. Estrella, P. Pouchin, M. Cohen-Tannoudji, L. David, and C. Chazaud, "NANOG initiates epiblast fate through the coordination of pluripotency genes expression," Nature Communications, vol. 13, p. 3550, June 2022. Number: 1 Publisher: Nature Publishing Group.
[5] M. J. Pokrass, K. A. Ryan, T. Xin, B. Pielstick, W. Timp, V. Greco, and S. Regot, "Cell-Cycle-Dependent ERK Signaling Dynamics Direct Fate Specification in the Mammalian Preimplantation Embryo," Developmental Cell, vol. 55, pp. 328-340.e5, Nov. 2020.
[6] O. Gandrillon, M. Gaillard, T. Espinasse, N. B. Garnier, C. Dussiau, O. Kosmider, and P. Sujobert, "Entropy as a measure of variability and stemness in single-cell transcriptomics," Current Opinion in Systems Biology, vol. 27, p. 100348, Sept. 2021.
[7] E. Posfai, S. Petropoulos, F. R. O. de Barros, J. P. Schell, I. Jurisica, R. Sandberg, F. Lanner, and J. Rossant, "Position- and Hippo signaling-dependent plasticity during lineage segregation in the early mouse embryo," eLife, vol. 6, p. e22906, Feb. 2017. Publisher: eLife Sciences Publications, Ltd.
[8] C. Albayrak, C. A. Jordi, C. Zechner, J. Lin, C. A. Bichsel, M. Khammash, and S. Tay, "Digital Quantification of Proteins and mRNA in Single Mammalian Cells," Molecular Cell, vol. 61, pp. 914-924, Mar. 2016.

5.7 Personalized prediction of mortality risks in chronic kidney disease patients

Bence Oláh

Peter L. Reichertz Institut für Medizinische Informatik (PLRI), Hannover, Germany

Abstract: The identification of chronic kidney disease (CKD) patients at an increased risk of death is crucial to improve clinical decision making and risk mitigation strategies.
We have developed new mortality risk prediction models for CKD patients based on routine patient parameters. For the model development, we have analyzed the German Chronic Kidney Disease (GCKD) study, a prospective, observational, national cohort study of CKD patients, mostly CKD stage G3. We have investigated 4,081 GCKD study participants with a mean follow-up time of 5.86 ± 1.48 years. In total, 515 death events (13 %) occurred within this observation period.
We trained a least absolute shrinkage and selection operator (LASSO) Cox proportional hazards (PH) algorithm with hyperparameter tuning within a 10-fold internal cross-validation. We assessed two different LASSO Cox PH models corresponding to two different hyperparameter settings according to internal partial likelihood deviance optimization. The more regularized ‚Äúlambda-1sd‚Äù LASSO Cox PH model included 22 predictors.
We assessed its performance in comparison to published state-of-the-art mortality risk prediction models developed for CKD and hemodialysis patients, i.e., the Floege- [1] and Bansal- [2] equations, within a resampling approach. Our new models have shown better performance in terms of concordance index and net reclassification improvement, and better calibration and ROC curve characteristics on the GCKD test data sets.
In conclusion, the two new mortality risk prediction models based on easily accessible routine patient parameters have the potential to improve CKD patient care. Their performance will be further assessed in external validation cohorts and they will be made available as an easy-to-use online service for fast implementation in routine patient care.
[1] Jürgen Floege et al. Development and validation of a predictive mortality risk score from a European hemodialysis cohort. DOI: 10.1038/ki.2014.419
[2] Nisha Bansal et al. Development and Validation of a Model to Predict 5-Year Risk of Death without ESRD among Older Adults with CKD.
DOI: 10.2215/CJN.04650514

5.8 Applying a GAN-based classifier to improve transcriptome-based prognostication in breast cancer

Cristiano Guttà

Insistute of Cell Biology and Immunology, University of Stuttgart, Germany

Abstract: Established prognostic tests based on limited numbers of transcripts can identify high-risk breast cancer patients, yet are approved only for individuals presenting with specific clinical features or disease characteristics. Deep learning algorithms could hold potential for stratifying patient cohorts based on full transcriptome data, yet the development of robust classifiers is hampered by the number of variables in omics datasets typically far exceeding the number of patients. To overcome this hurdle, we propose a classifier based on a data augmentation pipeline consisting of a Wasserstein generative adversarial network (GAN) with gradient penalty and an embedded auxiliary classifier to obtain a trained GAN discriminator (T-GAN-D). Applied to 1244 patients of the METABRIC breast cancer cohort, this classifier outperformed established breast cancer biomarkers in separating low- from high-risk patients (disease specific death, progression or relapse within 10 years from initial diagnosis). Importantly, the T-GAN-D also performed across independent, merged transcriptome datasets (METABRIC and TCGA-BRCA cohorts), and merging data improved overall patient stratification. In conclusion, the reiterative GAN-based training process allowed generating a robust classifier capable of stratifying low- vs high-risk patients based on full transcriptome data and across independent and heterogeneous breast cancer cohorts.

5.9 Differential network analysis in cancer research: application of jointly estimated Gaussian Graphical Models (GGM) for comparison of healthy pancreas and PDAC tumor tissue networks

Karly Conrads

University Medical Center Göttingen, Department of Medical Bioinformatics, Germany

Abstract: Differential gene expression (DEG) analysis is commonly used to compare groups in biomedical research, such as healthy and disease states. The resulting gene list ranges in the 10s-1000s and necessitates further down-stream analyses for context (like Go-term enrichment and GSEA). Not only is this a time-consuming and potentially biased process, but there is also evidence that the outcome of such DEG analyses is more the result of changes induced by the disease, as opposed to changes causative for the disease. The causative differences which underlie the healthy and disease state are often the result of subtle, differential up-stream interactions. Differential analysis of jointly estimated networks, in addition to DEG analysis, allows for the integration of significant downstream effects (DEGs) with their more subtle upstream causes (differential network connections) which might be missed with DEG analysis alone. To this end joint-graphical lasso (JGL) is used to estimate tissue-specific, transcription-based Gaussian graphical models (GGM) from nearly 500 healthy pancreas and PDAC tumor tissues samples. The output networks can then be compared for shared and differential gene interactions, in combination with differential gene expression, which adds context and allows for the analysis of key disease specific mechanism to identify potential druggable targets.

5.10 Challenges in Machine-Learning based detection of Measurable Residual Disease in AML patients

Friedemann Uschner

Institute for Medical Informatics and Biometry (IMB), Faculty of Medicine, TU Dresden, Germany

Abstract: The detection of Measurable Residual Disease (MRD) in patients with Acute Myeloid Leukemia (AML) during and after the treatment is an important diagnostic step to prospectively identify leukemia recurrence and adjust patient therapy. Complementing a range of molecular techniques, Multiparametric Flow Cytometry (MFC) has emerged as an inexpensive and fast method to monitor MRD. However, the diagnostic analysis of this data is usually done manually via a sequential, 2-dimensional gating procedure that is time-consuming, rater-dependent and requires a high level of expertise.

To enable a broader use of this MFC-based approach for MRD monitoring and to speed-up clinical treatment decision, we designed a Machine Learning (ML) pipeline to automate the quantification of residual disease levels. We investigate how different ML algorithms can be applied to this aim and how the preparation of training data influences the output qualitatively. The general approach uses the raw multi-dimensional MFC data of 449 AML patients to identify features of 32 pre-defined aberrant cell populations and subsequent patient classification and, thus, compliments the manual gating procedure. Comparing our results with this reference via patient survival we show the general feasibility of ML-based approaches as a tool to support faster detection of recurring malignancies. However, while this approach enables classification on patient-level of up to 80% accuracy, we show that the highly imbalanced nature of the data sets still limits this classification and poses a severe problem for the reliable quantification of malignant cell populations. We highlight the challenges and pitfalls that have to be solved in order to determine levels of MRD on this single-cell level and improve upon patient level classification as well.

5.11 Accelerating Biological Network Reconstruction with Machine Learning and Natural Language Processing

Marietta Hamberger

Institute of Medical Systems Biology, University Ulm, Germany

Abstract: Reconstructing biological networks from scientific literature is imperative for identifying potential therapeutic targets and comprehending complex biological systems. In particular, Boolean network models offer a solution for building regulatory circuits without the often-unavailable kinetic parameters [1]. However, manual network reconstruction is a labor-intensive and time-consuming process, often subject to bias and a lack of rigor.
Recent advances in natural language processing (NLP) and machine learning (ML) offer the possibility of automating this task.
This research examines the feasibility of establishing Boolean models through a hybrid approach that incorporates both rule-based and recent ML algorithms.
Methods:
We developed a first take on a hybrid model converting gene regulatory interactions from natural language statements into Boolean equations.
To achieve this goal, we fine-tuned the BioBERT language model for named entity recognition (NER) and integrated a hybrid rule-based approach for relation extraction (RE) [2]. We trained and evaluated our models on additional biomedical corpora and compared their performance against state-of-the-art models in the field.
Results:
Our study successfully replicated the results reported by the BioBERT authors on the NER task of four entity types (species, chemicals, diseases, genes). Additionally, preliminary results showed promising performance on further biomedical entity types and RE tasks. These findings indicate the potential to improve the accuracy and efficiency of network reconstruction using our approach, which could considerably expedite the workflow and offer a more comprehensive and efficient strategy for network reconstruction.
Conclusion:
By combining rule-based and ML algorithms, our approach presents a promising solution to the manual construction of biological networks from scientific literature. Our study emphasizes the potential of biomedical NLP techniques to pave the way for a more efficient and reproducible approach to network modeling.
In subsequent research, we plan to investigate the reliability of our approach for larger-scale and diverse corpora, explore the use of other ML techniques, and incorporate expert-in-the-loop feedback for additional quality control.
References
[1] Kauffman, Stuart. "Homeostasis and differentiation in random genetic control networks." Nature 224.5215 (1969): 177-178.
[2] Lee, Jinhyuk, et al. "BioBERT: a pre-trained biomedical language representation model for biomedical text mining." Bioinformatics 36.4 (2020): 1234-1240.

5.12 Time-series of transcriptome data reveal non-coding RNAs that can be incorporated into modelling approaches

Daria Meyer

University of Jena, Germany

Abstract: Non-coding RNAs (ncRNAs) are a special class of RNA that are transcribed from the DNA, but not translated into protein. Beside about 4,000 ncRNA classes in Rfam, an RNA families database (Griffiths-Jones et al.), the number of long ncRNAs is estimated to be more than 50,000 (Iyer et al.). Often ncRNAs have regulatory functions; but for most ncRNAs no function is yet known. Here, we propose a method to determine:

(i) differentially expressed ncRNAs suggesting a potential function in a given experimental setup

(ii) possible targets in the transcriptional 'behaviour' as the ncRNAs themself

(iii) a proof-of-principle for integrating regulation via ncRNAs into pathway-based modelling approaches

We show a clear functional relationship of ncRNAs using portal vein ligation (PVL), a surgical procedure that reduces perfusion in part of the liver. Specifically, we compare sham surgery to PVL, at 2 and 5 days, in both ligated and non-ligated liver lobes. Furthermore, we predict targets for selected ncRNAs by correlating transcriptome patterns with ncRNA expression. Finally, we propose to integrate ncRNAs into existing modelling approaches, such as pathway models of liver regeneration after partial hepatectomy (Köller et al.) or gene regulatory networks (Gebert et al.).

References:

Gebert, Jutta, Nicole Radde, and G-W. Weber. "Modeling gene regulatory networks with piecewise linear differential equations." European Journal of Operational Research 181.3 (2007): 1148-1165.

Griffiths-Jones, Sam, et al. "Rfam: an RNA family database." Nucleic acids research 31.1 (2003): 439-441.

Iyer, Matthew K., et al. "The landscape of long noncoding RNAs in the human transcriptome." Nature genetics 47.3 (2015): 199-208.

Köller, Adrian, et al. "Prediction of survival after partial hepatectomy using a physiologically based pharmacokinetic model of indocyanine green liver function tests." Frontiers in physiology (2021): 1975.

5.13 EnzymeML-based workflow for FAIR enzyme kinetics

Max Häußler

IBTB - University of Stuttgart, Germany

Abstract: Data management according to FAIR data principles is an emerging best practice in science.[1] In biocatalysis, the EnzymeML format implements these principles, enabling storing and working with biocatalytical (meta)data.[2] Furthermore, modeling results can be stored together with data on which the results are based on. Together with PyEnzyme, a Python interface for EnzymeML, powerful analysis workflows can be implemented, which exceed the capabilities of analysis in spreadsheet applications. Thus, the EnzymeML infrastructure supports the field by providing structured documentation to experimental data. Hence, enabling the implementation of FAIR analysis pipelines.
Here we present an EnzymeML-based workflow for kinetic parameter estimation, enabling a continuous data flow from raw data to kinetic parameters. Thereby, analytical raw data together with calibration data is used for accurate concentration calculation, enabling description of linear and non-linear relationships between the analytical signal and the analyte concentration. After calibration, the respective information is directly applied to the EnzymeML data model containing analytical raw data and yielding an EnzymeML data model with concentration data. Thereafter, concentration data from the EnzymeML data model is mapped to EnzymePynetics, a tool for kinetic parameter estimation based on time-course measurement data. After kinetic model selection, the modeling results are again written back to the EnzymeML data model and serialized to the standardized EnzymeML format.

[1] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9.

[2] Lauterbach, S., Dienhart, H., Range, J., Malzacher, S., Spöring, J. D., Rother, D., ... Pleiss, J. (2023). EnzymeML: seamless data flow and modeling of enzymatic data. Nature Methods, 1-3.

5.14 Continuous reference intervals determined with the Shiny application AdRI

Sandra Klawitter

Ostfalia Hochschule für angewandte Wissenschaften ‚Äì Hochschule Braunschweig/Wolfenbüttel, Germany

Abstract: Sandra Klawitter1,2, Georg Hoffmann2, Tim Kacprowski3,4, Frank Klawonn1,5

1 Ostfalia University of Applied Sciences, Department of Computer Science, Salzdahlumer Str. 46/48, 38302 Wolfenbüttel, Germany
2 Trillium GmbH Medizinischer Fachverlag, Jesenwanger Str. 42b, 82284 Grafrath, Germany
3 Peter L. Reichertz Institute for Medical Informatics of Technische Universität Braunschweig and Hanover Medical School, Division Data Science in Biomedicine, Rebenring 56, 38106 Braunschweig, Germany
4 Braunschweig Integrated Centre of Systems Biology, Technische Universität Braunschweig, Rebenring 56, 38106 Braunschweig, Germany
5 Helmholtz Centre for Infection Research, Biostatistics, Inhoffenstr. 7, 38124 Braunschweig, Germany

Background: Reference intervals are an important part of the interpretation of medical laboratory results. Especially in children and adolescents, their limits can change very rapidly with age [1]. We propose continuous methods to better represent the age-dependent progression of reference intervals. A user-friendly Shiny application called AdRI (Age-dependent Reference Intervals), available at https://github.com/SandraKla/AdRI, has been developed for this purpose.
Methods: Generalized additive models for position, scale, and shape parameters (GAMLSS) were developed by Rigby and Stasinopoulos in 2005 and implemented in the gamlss R-package, which provides a variety of features and capabilities for univariate statistical regression modelling and statistical learning [2]. The purpose of the Shiny application AdRI is illustrated using 40 biochemical markers from the Canadian Laboratory Initiative on Pediatric Reference Intervals (CALIPER) [3].
Results: Depending on the additive term used, we obtain different smoothed percentile curves of laboratory values. For ALP, the GAMLSS with P-splines for females and Decision Trees for males has the lowest Generalized Akaike Information Criterion (GAIC). All models recognize that the range for ALP of the models is wider after birth and then narrows to follow a relatively constant course over a long period of childhood. With the onset of puberty, the range widens again for both sexes and decreases sharply towards adulthood. Conclusion: We demonstrate the superiority of continuously modeled reference intervals and provide the Shiny application AdRI to make the technique easily accessible to clinicians and other experts. The influence of pathological values, hyperparameters and the distribution of data over age should be the subject of further investigation, given that the database of laboratory

values in newborns and children is generally small. However, the expertise of a laboratory physician is still required to select the correct GAMLSS for the age-dependent course of laboratory values.

References

[1] Wilson SM, Bohn MK, Madsen A, Hundhausen T, Adeli K. LMS-based continuous reference percentiles for 14 laboratory parameters in the CALIPER cohort of healthy children and adolescents. Clin Chem Lab Med. 2023;10.1515/cclm-2022-1077.
doi:10.1515/cclm-2022-1077
[2] Rigby R.A. and Stasinopoulos, M.D. Generalized additive models for location, scale and shape. Appl. Statist. 2005;54, part 3, pp 507-554.
https://doi.org/10.1111/j.1467-9876.2005.00510.x
[3] Colantonio DA, Kyriakopoulou L, Chan MK, et al. Closing the gaps in pediatric laboratory reference intervals: a CALIPER database of 40 biochemical markers in a healthy and multiethnic population of children. Clin Chem. 2012;58(5):854-868.
doi:10.1373/clinchem.2011.177741

5.15 A Python package for (Adaptive) Digital Tissue Deconvolution

Malte Mensching-Buhr

UMG Goettingen, Medizinische Bioinformatik, Germany

Abstract: Spatial transcriptomics (ST) provides spatially distributed RNA profiles of tissues specimens. Individual spots of an ST slide can contain signals from multiple cells. Cell-type deconvolution algorithms allow to computationally reverse engineer the underlying spatial distribution of different cell types, providing insights into the complex interplay of immune cells and the signals they exchange. The performance of such cell-type deconvolution algorithms can be substantially improved via weighted gene spaces. Here, we present a python package which learns such gene spaces based on a machine learning approach called Digital Tissue Deconvolution (DTD). Furthermore, we present an implementation of Adaptive Digital Tissue Deconvolution (ADTD), which also deals with suboptimal reference profiles and potentially unknown background contributions.