Workshop on Computational Models in Biology and Medicine 2017

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

March 2nd-3rd, 2017

University of Veterinary Medicine Hannover, Germany

Workshop outline

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

Keynote speakers

- Vanessa Didelez, Leibniz Institute for Prevention Research and Epidemiology, Bremen
- Korbinian Strimmer, Imperial College, London
- Arne Traulsen, MPI for Evolutionary Biology, Plön

Workshop venue

The workshop will be hosted in the lecture hall "Bayerhörsaal" at the clinical center "Klinikum am Bünteweg" of the University of Veterinary Medicine Hannover, Bünteweg 9, 30559 Hannover. The lecture hall is easily accessible by public transport. For details, please check the University's web page.

Organization

The workshop is jointly organized by the GMDS/IBS working groups "Statistical Methods in Bioinformatics" (Klaus Jung, University of Veterinay Medicine Hannover; Holger Fröhlich, University of Bonn) and "Mathematical Models in Medicine and Biology" (Markus Scholz, University of Leipzig; Ingmar Glauche, University of Dresden).

Contact and local organization

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

Dr. Jochen Kruppa
Institute for Animal Breeding and Genetics
University of Veterinary Medicine Hannover
Bünteweg 17p, 30559 Hannover, Germany
Tel.: +49 511 953 8869
E-mail: jochen.kruppa@tiho-hannover.de

Support

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

Program

Thursday, March 2, 2017

12:45-13:00 Workshop Opening

Session 1 (Chair: Klaus Jung)
13:00-13:25 Julia Perera-Bel
How to report somatic variants in molecular tumor boards
13:25-13:50 Jochen Kruppa
The kmerPyramid as visualization tool for k-mer distributions in viruses and bacteria
13:50-14:15 Benjamin Engelhardt
A Bayesian approach for estimating hidden variables in ODE based models in systems biology
14:15:14:40 Florian Stuhler
Computational modeling reveals mechanisms of action of Bcl2-inhibitors in the development of novel anti-cancer drugs

14:40-16:00 Coffee break & poster session

Session 2 (Chair: Markus Scholz)
16:00-16:45 Keynote 1 Vanessa Didelez
Challenges for Mendelian randomisation analyses
16:45-17:10 Marco Grzegorczyk
Bayesian inference of semi-mechanistic network models
17:10-17:35 Manuel Nietert
Track descriptors - automated characterization of cell tracks
17:35-18:00 Marvin Böttcher
Modeling chronic myeloid leukemica emergence and treatment

19:30 Social event at the restaurant Meiers Lebenslust
(U6 from Bünteweg/TiHo direction Nordhafen and leave at Aegidientorplatz)

Friday, March 3, 2017

Session 3 (Chair: Holger Fröhlich)

08:30-09:15 Keynote 2 Korbinian Strimmer
An entropy approach for integrative genomics and network modeling
09:15-09:40 Ashar Ahmad
Towards Clinically More Relevant Dissection of Patient Heterogeneity via Survival based Bayesian Clustering
09:40-10:05 Simon Klau
priorityLASSO: a hierarchical method for patient outcome prediction based on multi-omics data taking practitioners' preferences into account
10:05-10:30 Michael Seifert
Importance of rare gene copy number alterations for personalized tumor characterization and survival analysis

10:30-11:00 Coffee break

Session 4 (Chair: Ingmar Glauche)

11:00-11:45 Keynote 3 Arne Traulsen
Telomere distributions as a window into stem cell dynamics
11:45-12:10 Matthias Horn
A single-cell based model explains patterns of clonal evolution in primary and relapsed follicular lymphoma
12:10-12:35 Yuri Kheifez
Modeling individual time courses of thrombopoiesis during multi-cyclic chemotherapy
12:35-13:00 Michael Altenbuchinger
Reference point insensitive molecular data analysis

13:00-13:10 Workshop closing

Abstracts of Talks

Towards Clinically More Relevant Dissection of Patient Heterogeneity via Survival based Bayesian Clustering

Ashar Ahmad

University of Bonn, Institute for Computer Science

ashar@bit.uni-bonn.de

Disease sub-type identification has traditionally been explored in an unsupervised machine learning paradigm which involves clustering of patients based on available -omics data, such as gene expression. A follow-up analysis involves determining the clinical relevance of the molecular sub-types such as that reflected by comparing their disease progressions. The above methodology, however, fails to guarantee the separability of the sub-types based on their subtype-specific survival curves.

We propose a new algorithm, Survival based Bayesian Clustering (SBC) which simultaneously clusters heterogeneous -omics and clinical end point data (survival or progression free survival) in order to discover clinically relevant disease subtypes. For this purpose we formulate a novel Hierarchical Bayesian Graphical Model which combines a Dirichlet Process Gaussian Mixture Model (DPMM) with an Accelerated Failure Time (AFT) model. In this way we make sure that patients are grouped in the same cluster only when they show similar characteristics with respect to molecular features across data types (e.g. gene expression, mi-RNA) as well as survival times.

We extensively test our model in simulation studies and apply it to cancer patient data from the Breast Cancer data set and the TCGA repository. Notably, our method is not only able to find clinically relevant sub-groups, but is also able to predict cluster membership and survival on test data in a better way than other competing methods.

Reference point insensitive molecular data analysis

Michael Altenbuchinger

Statistical Bioinformatics, Institute of Functional Genomics, University of Regensburg

michael.altenbuchinger@ukr.de

Abstract: In biomedicine, every molecular measurement is relative to a reference point, like a fixed aliquot of RNA extracted from a tissue, a defined number of blood cells, or a defined volume of bio-fluid. Reference points are often chosen for practical reasons. For example, we might want to assess the metabolome of a diseased organ but can only measure metabolites in blood or urine. In this case the observable data only in
directly reflects the disease state. The statistical implications of these discrepancies in reference points have not yet been discussed.

Reference points are closely linked to data normalization and preprocessing. If we normalize data to a common mean, we generate a data internal reference point: each measurement of a specific feature is expressed relative to the mean of all features. For instance, gene expression profiling measures abundances of mRNA transcripts and normalization to a common mean gives abundances relative to the mean expression level of all genes. This prescription is not unique, e.g. profiles can be normalized to a constant value of one or several housekeeping features. In the latter case we choose another data internal reference point: the mean of the housekeeping features.

Here we show that reference point discrepancies compromise the performance of regression models like the LASSO. As an alternative, we suggest zero-sum regression for a reference point insensitive analysis.

As the LASSO, zero-sum regression performs feature selection by penalizing the l1 norm. However, zero-sum regression additionally enforces the sum over the regression weights to equal zero. This constraint has important practical implications. First, we are unbiased concerning the “best” reference point (or “best” normalization). Second, sparse zero-sum models are “truly sparse”. We illustrate these two aspects by contrasting zero-sum regression with the standard LASSO in a simulation study and in an application that integrates intestinal microbiome analysis with metabolomics. Furthermore, we illustrate that zero-sum regression yields universal models that apply beyond specific platforms. This is demonstrated for the classification of diffuse large B-cell lymphomas into their cell-of-origin subtypes.

Modeling chronic myeloid leukemica emergence and treatment

Marvin Böttcher

MPI for Evolutionary Biology

boettcher@evolbio.mpg.de

Chronic myeloid leukemia (CML) is a cancer of the hematopoetic (blood) system. A mutation in hematopoetic stem cells (HSC) leads to abnormal cell proliferation behavior which ultimately leads to the diagnosis. However, this mutation may not have a direct effect on the hematopoetic stem cells, but instead only the differentiated cells have changed differentiation and growth characteristics. Treatment with Imatinib or similar drugs target exactly this mutated cell behavior and essentially reverses the effect. This treatment is very effective both in terms of long-term survival and also much lower side effects than traditional cancer treatments. Since the stem cells are not targeted by therapy, lifelong treatment with Imatinib is necessary. However, in several clinical studies it has been shown that stopping treatment not always leads to CML relapse, but many patients stay below the cancer detection level for many years after stopping the treatment.

We use a hierarchical compartment based model (Werner et al. 2011, Plos Comp. Biol.) to capture this disparity in outcomes and to learn more about the involved dynamics and risk factors. Thereby we simulate thousands of virtual patients starting with a single mutated hematopoetic stem cell and follow each of them through disease progression, diagnosis, treatment and eventually relapse over time. Since the number of active hematopoetic stem cells and the first progenitor cells is very small, stochastic effects play a major role, eventually resulting in the different outcomes for the patients. However, the output of fully differentiated blood cells is very large which makes a full stochastic simulation unfeasible. Therefore we use a multiscale approach with stochastic simulations in the smaller compartments coupled to deterministic equations in the large compartments where the relative stochastic fluctuations are small (Lenaerts et al. 2010 Haematologica).

Interestingly, not all patients develop CML even when starting from them same initial conditions of a single leukemic stem cell. Additionally, the time of diagnosis varies greatly from patient to patient. The same is also true for the effects of treatment: the reduction in tumor burden under treatment is highly variable as well as the probability for relapse. Eventually our stochastic simulations might help to identify factors that predict the course of disease and if or when it is possible to safely stop the treatment.

Challenges for Mendelian randomisation analyses

Vanessa Didelez

Leibniz Institute for Prevention Research and Epidemiology, Bremen

didelez@leibniz-bips.de

Mendelian randomisation (MR) refers to situations where a genetic predisposition can be exploited as an instrumental variable (IV) to estimate the causal effect of a modifiable risk factor or exposure on an outcome of interest. For example, the ALDH2 gene is associated with alcohol consumption, and has therefore successfully been used as an IV to estimate the causal effect of alcohol on outcomes related to coronary heart disease. MR analyses have become very popular especially recently with the increased availability of GWAS data. This gives rise to the following challenges:
It is common that several SNPs are found to be associated with an exposure of interest, i.e. there are potentially numerous IVs; if these are all valid IVs, methods for multiple instruments are called for.
It is also common that many of these numerous potential IVs are only weakly associated with the exposure of interest; the phenomenon of weak IV bias is well-known for the simple case, and of course it also affects the multiple IV case; hence methods for multiple weak IVs are needed; it has been proposed to combine SNPs into an allele score, i.e. a single hopefully stronger IV, but this can lead to bias if done in a data-driven manner.

Further it is unlikely that all such SNPs are actually valid instruments for the causal effect of interest; they could for instance have pleiotropic effects or violated the IV conditions in other ways. Some first proposals to deal with such violations of assumptions suggest methods that do not require knowledge of which IVs are valid and which aren't, e.g. similar to Egger regression in meta analyses. Data is often only available from different sources, one with instrument-exposure data, and a different source with instrument-outcome data. This means inference has to be based on two bivariate samples (possibly with additional covariates), instead of a joint sample. Two-stage-least-squares (TSLS) can be adapted to this case as `two-sample TSLS', but more robust methods would be desirable.

Even less information is available if an MR analysis has to make do with summary data (often when based on case-control studies, but also otherwise). This means that from a number of primary analyses we only have measures of the instrument-exposure and (possibly from different studies) measures of the instrument-outcome associations.

Moreover, typical data available for MR analyses often comes from case-control studies, i.e. we have a binary outcome and sampling is conditional on case or control status; linear models are not appropriate in this case and if the retrospective nature of the sampling is ignored this can induce selection bias; even if data was sampled prospectively, selection bias can occur if e.g. volunteering is related to exposure/outcome status. In this context it is particularly important to ensure that the chosen meth-

ods for analysis have the null-preservation property, i.e. are consistent under the null-hypothesis of no causal effect.

In this presentation, I will give an overview over the above challenges as well as existing approaches to tackle them, their strengths and limitations.

A Bayesian Approach for estimating hidden variables in ODE based models in systems biology

Benjamin Engelhardt

University of Bonn, Institute for Computer Science, Algorithmic Bioinformatics

engelhar@bit.uni-bonn.de

Models of biological systems become more and more complex. These complex models are often formulated as ordinary differential equation (ODE) systems and allow for obtaining insights into biological processes. This is typically done on the basis of a deep understanding of the underlying biochemical reaction system. However, a major difficulty is that biological systems are naturally open and specifying their boundaries is a highly non-trivial task. Most researchers in systems biology are thus faced with the still unsolved issue to find a compromise between model complexity and the limited amount of knowledge, data and time. However, hidden variables as well as erroneous system variable interactions could lead to large deviations from the fitted model to measured data.

So far most researchers have addressed this issue in a trial and error like manner. We present a systematic, fully algorithmic approach, which detects hidden variables as well as missed interactions between these variables in ODE based mechanistic models. For this purpose we embed the ODE system into a probabilistic graphical modeling framework.
Within this framework we then propose a Markov Chain Monte Carlo based algorithm to infer the state of hidden variables.

We successfully applied our approach to several simulated networks, essential network motifs as well as literature known complex systems with given data. Our algorithm was able to identify correctly a missing reaction in the EPO receptor pathway and several others as the photomorphogenic UV-B signaling in plants and the heterotrimeric G-Protein cycle in yeast. Altogether we report progress to revise ODE based models of molecular reaction systems based on measured data.

Bayesian Inference of Semi-Mechanistic Networkmodels

Marco Grzegorczyk

Johann Bernoulli Institute (JBI), Groningen University

m.a.grzegorczyk@rug.nl

A topical and challenging problem for statistics is to infer the structure of complex systems of interacting units.In many scientific disciplines such systems are represented by interaction networks described by systems of differential equations.
My presentation is about a novel semi-mechanistic Bayesian modelling approach for infering the structures and parameters of these interaction networks from data. The inference approach is based on gradient matching and a non-linear Bayesian regression model. My real-world applications stem from the topical field of computational systems biology, where researchers aim to reconstruct the structure of biopathways or regulatory networks from postgenomic data.
My focus is on investigating to which extent certain factors influence the network reconstruction accuracy. To this end, I compare not only (i) different methods for model selection, including various Bayesian information criteria and marginal likelihood approximation methods, but also (ii) different ways to approximate the gradients of the observed time series. Finally, I cross-compare the performance of the new method with a set of state-of-the art network reconstruction networks, such as Bayesian networks. Within the comparative evaluation studies I employ ANOVA schemes to disambiguate to which extents confounding factors impact on the network reconstruction accuracies.

A single-cell based model explains patterns of clonal evolution in primary and relapsed follicular lymphoma

Matthias Horn, Markus Kreuz, and Markus Loeffler

Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany

matthias.horn@imise.uni-leipzig.de

Introduction: Germinal centers (GCs) provide signals for B cells to become high affinity antigen-detecting cells. In this process, somatic hypermutation (SHM) and clonal selection play a pivotal role, resulting in the accumulation of mutations in immunoglobulin (Ig) genes. Follicular lymphoma (FL) represents the prototype of GC B cell-derived lymphomas. The disease can be controlled but not cured, with a median survival of more than 18 years after combined immunochemotherapy. In the majority of patients relapses are observed, in some cases additionally associated with transformation to more aggressive diffuse large B-cell lymphoma. Objective: In order to gain insight into processes underlying clonal evolution and lymphoma relapse, we applied a mathematical modeling approach. We particularly sought to conclusively explain qualitatively different patterns of FL evolution, such as divergent, sequential and no evolution, and analyze whether the observed inter-patient heterogeneity is of relevance for clinical decision-making. Materials and Methods: We developed a single-cell based mathematical model of physiological GC reaction to study the dynamics of GC expansion and B cell affinity maturation. Furthermore, we applied our model to the situation of FL emergence and relapse. We compared our modeling results to phylogenetic trees reconstructed from clinical measurements (sequences of Ig heavy chain variable gene rearrangements) of primary and relapse tumor in FL patients [Loeffler et al., Leukemia 29(2):456-63, 2015]. Results: Based on parameter changes in a single cell, representing malignant transformation, the model is capable of reproducing typical features of lymphoma emergence and relapse. Specifically, the different patterns of evolution observed in FL patients can be fully explained, based on variation of the timepoints of interfollicular cell migration and the absolute number of migrating FL cells. Evolution is stochastic regarding timepoints and number of cells migrating and thus cannot be predicted. As a consequence, different patterns of evolution within the same patient are predicted to be observed. Importantly, our model predicts complete cessation of SHM after clonal dominance of FL cells within a GC. Conclusion: We found a comprehensive mathematical model of physiological GC reaction, dynamics of FL emergence and heterogeneity of clonal evolution. Due to the stochasticity of evolution we caution to draw clinically relevant conclusions from evolutionary profiles of tumor cells (e.g., with respect to driver mutations). Suppression of interfollicular cell migration through lymphatic vessels already in early disease stages might be a therapeutic goal.

Modeling individual time courses of thrombopoiesis during multi-cyclic chemotherapy

Yuri Kheifetz

IMISE (The Institute for Medical Informatics, Statistics and Epidemiology)

yuri.kheifetz@imise.uni-leipzig.de

Question

Decreased platelet counts, called thrombocytopenia, is a major dose-limiting side effect of dose-intense cancer chemotherapies. However, standard coursers of many chemotherapies result in considerable variability in drug induced platelets dynamics. A major challenge of individualized medicine is to take all relevant factors into account for optimal risk management using individualized modeling. Actually there is a gap between simple phenomenological predictive population models and complex biomathematical modeling of averaged data. Another problem is an inter-occasional variability of platelets response to multi-cyclic chemotherapy. Current statistical model cannot identify patients in the beginning of the treatment, which display sever thrombocytopenia only at the late cycles (Ziepert et al. 2008). In order to tackle these problems, we revised a biomathematical model of average human thrombopoiesis under chemotherapy (Scholz et al. 2010) towards modelling individual time courses taking into account accumulating long-range effects of chemotherapy on a bone-marrow.

Methods

We fitted model parameters hierarchically. 25 population and 10 individual parameters were estimated using dense time series of three patients treated with BEACOPP chemotherapy (Engel et al. 1999). We used these estimators as population parameters as well as prior values in order to fit individually 12 parameters for sparser data of selected 135 patients from the German non-Hodgkin‘s lymphoma trial group applying CHOP-like chemotherapies (Pfreundschuh et al. 2004a; Pfreundschuh et al. 2004b). Individual deviations from treatment protocol were also considered. The individual parameter estimations used simultaneously information from other studies published in literature by incorporating it into the respective weighted likelihood functions, assuming a virtual participation of the patients in the corresponding experiments. This additional information included osteoblasts count change during multi-cyclic chemotherapy (Li et al. 2015), average dynamics of TPO, platelets and megakaryocytes of recombinant-TPO treated healthy patients (Harker et al. 2000) as well as a platelets dynamic after labeled platelets transfusions to patients with different degrees of thrombocytopenia (Hanson and Slichter 1985).

Results

Several new biological insights were discovered and modeled. We have hypothesized a bi-phasic stimulation of thrombopoiesis and we described detailed megakaryocytes populations of ploidies 2-128 as well as proplatelets. We proposed complex TPO-mediated regulation of megakaryocytes’ commitment to either endomitosis or to proplatelets formation, or to the hypothesized dormant

state. The cumulative decrease in average platelets level during multi-cyclic chemotherapy was attributed to interactions between quiescent and active stem cells compartments as well as to the accumulating injures of bone-supporting osteoblasts. We found an optimal tradeoff between goodness of fit and overfitting for most of the patients.

Conclusions

We established a model of individual thrombopoiesis response to chemotherapy. Heterogeneity between patients can be traced back to heterogeneity of a few model parameters. These long-term systems changes during multi-cyclic poly-chemotherapies explain increasing thrombocytopenia severity in the later treatment cycles comparing to that of the early ones. This model is being used currently for a posteriori next-cycle dosage adjustment and we are studying its predictive potential.

References

Engel, C.; Loeffler, M.; Franke, H.; Schmitz, S. (1999): Endogenous thrombopoietin serum levels during multicycle chemotherapy. In: British journal of haematology 105 (3), S. 832–838.

Hanson, S. R.; Slichter, S. J. (1985): Platelet kinetics in patients with bone marrow hypoplasia: evidence for a fixed platelet requirement. In: Blood 66 (5), S. 1105–1109.

Harker, L. A.; Roskos, L. K.; Marzec, U. M.; Carter, R. A.; Cherry, J. K.; Sundell, B. et al. (2000): Effects of megakaryocyte growth and development factor on platelet production, platelet life span, and platelet function in healthy human volunteers. In: Blood 95 (8), S. 2514–2522.

Li, Sidan; Zou, Dehui; Li, Changhong; Meng, Hengxing; Sui, Weiwei; Feng, Sizhou et al. (2015): Targeting stem cell niche can protect hematopoietic stem cells from chemotherapy and G-CSF treatment. In: Stem cell research & therapy 6, S. 175. DOI: 10.1186/s13287-015-0164-4.

Pfreundschuh, Michael; Trumper, Lorenz; Kloess, Marita; Schmits, Rudolf; Feller, Alfred C.; Rube, Christian et al. (2004a): Two-weekly or 3-weekly CHOP chemotherapy with or without etoposide for the treatment of elderly patients with aggressive lymphomas: results of the NHL-B2 trial of the DSHNHL. In: Blood 104 (3), S. 634–641. DOI: 10.1182/blood-2003-06-2095.

Pfreundschuh, Michael; Trumper, Lorenz; Kloess, Marita; Schmits, Rudolf; Feller, Alfred C.; Rudolph, Christian et al. (2004b): Two-weekly or 3-weekly CHOP chemotherapy with or without etoposide for the treatment of young patients with good-prognosis (normal LDH) aggressive lymphomas: results of the NHL-B1 trial of the DSHNHL. In: Blood 104 (3), S. 626–633. DOI: 10.1182/blood-2003-06-2094.

Scholz, Markus; Gross, Arnd; Loeffler, Markus (2010): A biomathematical model of human thrombopoiesis under chemotherapy. In: Journal of theoretical biology 264 (2), S. 287–300. DOI: 10.1016/j.jtbi.2009.12.032.

Ziepert, M.; Schmits, R.; Trumper, L.; Pfreundschuh, M.; Loeffler, M. (2008): Prognostic factors for hematotoxicity of chemotherapy in aggressive non-Hodgkin's lymphoma. In Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 19 (4), pp. 752–762. DOI: 10.1093/annonc/mdm54.

priorityLASSO: a hierarchical method for patient outcome prediction based on multi-omics data taking practitioners' preferences into account

Simon Klau, Vindi Jurinovic, Roman Hornung, Tobias Herold, Anne-Laure Boulesteix

IBE München

simonklau@ibe.med.uni-muenchen.de

In the last few years, bioassay technology improvement and cost reduction have made collecting several types of high-dimensional "omics" data (denoted as "modalities" from now on) in the same study for the same patients feasible. For example, methylation data, copy-number data and transcriptomic data may be available for the same patient cohort. Other examples include microRNA expression, proteomic data, metabolomic data, and single nucleotide polymorphisms (SNPs). Although using omics markers from a single modality (e.g., transcriptomic data) for the prediction of patients' clinical outcome (such as survival time or response to therapy) in the context of personalized medicine has been a well-studied topic, it is not clear how different modalities should be handled in this context.

In practice, it is often possible to make a guess on the utility of the different modalities for prediction based on previous literature or medical scientists' priori knowledge. Furthermore, medical scientists often have "preferences" for one or several modality/ies in the sense that they would prefer a prediction rule which uses (primarily) variables from this/these modality/ies, for example for financial or technical reasons – because the different data modalities are typically not equally cheap and easy to collect in the lab. In this context, we present and illustrate a new simple and pragmatic method based on Lasso regression called priorityLASSO, which incorporates the "priority sequence" defined by the practitioners of the form, say, "clinical modality > gene expression modality > methylation modality > …".

The method consists of a sequence of M Lasso regression steps, where M is the total number of considered modalities. The first step consists in fitting a (generalized) regression model via Lasso to the most prioritized modality, in the second step Lasso is applied to the second modality while considering the linear predictor fitted in the first step as an offset, and so on until the Mth modality. Our method is very fast even for large numbers of modalities, easily interpretable and transportable since it is based on Lasso regression. It is applicable to all types of outcome variables and predictor variables handled by Lasso. It does not necessarily yield optimal models in terms of prediction accuracy, but poses a convenient solution, because it represents a good compromise between accuracy on the one hand and practical issues and common requests from practitioners on the other hand. Note that the principle of priorityLASSO can be simply adapted to variants of Lasso regression (such as, e.g., elastic nets).

We perform an extensive simulation study which shows that, especially in the case of strong correlation between omics modalities, priorityLASSO tends to give more weight to highly prioritized modalities than methods ignoring priorities. Finally, the use of the method in practice is illustrated through an application of the prediction of the resistance of acute myeloid leukemia patients to induction

treatment. This analysis is based on large scale gene expression, extensive genetic testing and clinically implemented variables of 666 AML patients treated on two consecutive clinical trials. A user-friendly implementation of priorityLASSO method is publicly available as an R package on CRAN.

The kmerPyramid as visualization tool for k-mer distributions in viruses and bacteria

Jochen Kruppa, Klaus Jung

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

jochen.kruppa@tiho-hannover.de

Bioinformatics methods often incorporate the frequency distribution of nulecobases or k-mers in DNA or RNA sequences, for example as part of metagenomic or phylogenetic analysis. Because the frequency matrix with sequences in the rows and nucleobases in the columns is multi-dimensional it is hard to visualize. We present the R-package 'kmerPyramid' that allows to display each sequence, based on it's nucleobase or k-mer distribution projected to the space of principal components, as a point within a 3-dimensional, interactive pyramid. Using the computer mouse, the user can turn the pyramid's axes, zoom in and out and identify individual points. Additionally, the package provides the related frequency distribution matrices of about 2.000 bacteria and 5.000 viruses, respectively, calculated from NCBI genbank. The 'kmerPyramid' can particularly be used for intra- and inter species comparisons. The kmerPyramid is based on principal component analysis (PCA) that is used to project the multi-dimensional matrix of nucleobase and k-mer frequencies in the 3-dimensional space. PCA, as a method for dimension reduction, has already been demonstrated to preserve relevant information when exploring these frequencies (Dodsworth et al., 2013; Podar et al., 2013; Imelfort et al., 2014).

References:

[1] Dodsworth, J. A. et al. (2013). Single-cell and metagenomic analyses indicate a fermentative and saccharolytic lifestyle for members of the OP9 lineage. Nat Commun, 4, 1854.
[2] Imelfort, M. et al. (2014). GroopM: an automated tool for the recovery of population genomes from related metagenomes. PeerJ, 2.
[3] Podar, M. et al. (2013). Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. Biology Direct, 8, 9.

Track descriptors - automated characterization of cell tracks

Manuel M. Nietert, Michael Haberl, Francesca Oboardi, Alexander Flügel, Franziska Linke, Dieter Kube, Julia Gross, M. Shariful Islam, Tim Beissbarth, Tim Friede

University Medical Center Göttingen - UMG, Germany

manuel.nietert@med.uni-goettingen.de

Cell migration is a fundamental process in cell biology and is tightly linked to many important hysiological and pathological events such as the immune response, wound healing, tissue differentiation, metastasis, embryo genesis, inflammation and tumor invasion. Experimental advances provide techniques to observe migrating cells both in vitro and vivo. These imaging techniques provide two-dimensional or even three-dimensional movie data (image stacks) to be analyzed. These image stacks are the primary data source used to identify certain features (shapes within the image) as cells. By tracking the displacement of the so defined cells over the time axis, which defines their relative movement from time frame to time frame, the movement path of the cell is derived (a trajectory). These coordinate vector lists describing the track positions can be mathematically described / translated as feature vectors. Using the feature vector representations we demonstrate the feasibility to derive population classifications for analyzing cell track data and the concurrent related statistical analysis to characterize differential conditions: e.g. chemotaxis experiments. Additionally we will present means to extend the analysis to incorporate the local context of the tracks.

References:
Leukemia. 2016 Aug 18. doi: 10.1038/leu.2016.232
Oncogene. 2016 Jun 6. doi: 10.1038/onc.2016.183

How to report somatic variants in molecular tumor boards

Julia Perera-Bel

University Medical Center Göttingen

julia.perera@med.uni-goettingen.de

The understanding of complex diseases, such as cancer, is becoming more comprehensive with the improvements of high-throughput technologies e.g. next-generation sequencing. However, advances in technology platforms and bioinformatic tools contrast with the scarce implementation of cancer genomics in clinical practice. One reason for this situation is that pathologists and oncologists have to face thousands of genomic alterations and unravel their clinical relevance. Accordingly, the scientific community has claimed the need of a comprehensive knowledge database [1] as well as decision support platforms [2, 3] for the interpretation and reporting of genomic findings in clinical practice e.g. in molecular tumor boards.

Towards this end, we have developed a framework for reporting genomic data relying entirely on public knowledge. The report is tailored towards genomic alterations with predictive evidence on drug response. In particular, suggested treatment options are classified according the stage of development of the drug (approved, clinical trials or pre-clinical studies) and the cancer type for which the predictive association exists.

To test the approach we applied it to a TCGA dataset comprising 3184 samples. The results showed a substantial increase on the number of patients with treatment options when drugs in clinical trials were included, compared to only approved drugs. Also, we assessed the clinical utility of the reports on cancer patients whose treatment was based on their genomic profiles. The reports included the treatment decided by the expert panel.

We present a method to report treatment options based on the genomic profile of the patient. It is designed as a supporting tool for all clinicians, biologists and bioinformaticians working with genomic characterization of patients in clinical routine and face complex decisions regarding treatment options.

Importance of rare gene copy number alterations for personalized tumor characterization and survival analysis

Michael Seifert

Medical Faculty Carl Gustav Carus, Institute for Medical Informatics and Biometry (IMB), TU Dresden

michael.seifert@tu-dresden.de

Copy number alterations (CNAs) of large genomic regions are frequent in many tumor types, but only few of them are assumed to be relevant for the cancerous phenotype. It has proven exceedingly difficult to ascertain rare mutations that might have strong effects in individual patients. Here, we show that a genome-wide transcriptional regulatory network inferred from gene expression and gene copy number data of 768 human cancer cell lines can be used to quantify the impact of individual patient-specific gene CNAs on cancer-specific survival signatures [1]. The model was highly predictive for gene expression in 4,548 clinical samples originating from 13 different tissues. Focused analysis of tumors from six tissues revealed that in an individual patient a combination of up to 100 gene CNAs directly or indirectly affect the expression of clinically relevant survival signature genes. Importantly, rare patient-specific gene CNAs (less than 1% in a given cohort) often have stronger effects on signature genes than frequent gene CNAs. Subsequent integration with genomic data suggests that frequency variation among high-impact genes is mainly driven by gene location rather than gene function. Survival analyses on independent tumor cohorts revealed tumor-type specific trends indicating that rare gene CNAs can be as important as frequent gene CNAs for the prediction of patient survival. Our framework contributes to the individualized quantification of cancer risk, along with determining individual key risk factors and their downstream targets.

Keywords: cancer genomics, network biology, network inference, network propagation, gene copy number mutations

[1] Seifert M, Friedrich B, Beyer A, Importance of rare gene copy number alterations for personalized tumor characterization and survival analysis, Genome Biology, 2016, 17:204, https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1058-1.

An entropy approach for integrative genomics and network modeling

Korbinian Strimmer, Takoua Jendoubi

Imperial College, London

k.strimmer@imperial.ac.uk

Multivariate regression approaches such as Seemingly Unrelated Regression (SUR) or Partial Least Squares (PLS) are commonly used in vertical data integration to jointly analyse different types of omics data measured on the same samples, such as SNP and gene expression data (eQTL) or proteomic and transcriptomic data. However, these approaches may be difficult to apply and to interpret for computational and conceptual reasons.

Here we present a simple alternative approach to integrative genomics based on using relative entropy to characterise the overall association between two (or more) sets of omic data, and to infer the underlying corresponding association network among the individual covariates. This approach is computationally inexpensive and can be applied to large-dimensional data sets. Moreover, it may also be viewed as a special form of a latent-variable multivariate regression model.

We illustrate this approach by analysing publicly available metabolomic and transcriptomic data from the DILGOM study.

Computational modeling reveals mechanisms of action of Bcl2-inhibitors in the development of novel anti-cancer drugs

Florian Stuhler, Yasmin Hashambhoy-Ramsay, Nicole Radde

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

florianstuhler@hotmail.com

B-cell lymphoma 2 (Bcl2) is a highly-discussed target for novel anti-cancer drugs. Based on an extensive dataset that contains amounts of proteins of the apoptotic pathway in TRAIL and Bcl2 inhibitor ABT-263 mono-and combination therapy studies comprising more than 40 cell lines, we have developed applied two complementary mathematical approaches in order to predict responsiveness to Bcl2 inhibition and drug efficacy. First, a Gaussian process regression model was trained with protein levels from different cell lines that have been treated with TRAIL and Bcl2 inhibitors in different doses. A leave-one-out cross validation showed that the accuracy of our trained model is not satisfying. Furthermore, the model lacks predictive power, which might be due to the fact that the number of proteins that were included in the analysis (eight in total) could not account for the overall variability in the response across cell lines. In a second approach, we adapted an existing apoptosis model for HCT116 cell lines for our specific purposes. Using this model, we were able to qualitatively confirm published hypotheses about links between caspase-8 activity, the effect of Bcl2 inhibition and cell death .

Abstracts of Posters

Bringing Pathway Knowledge to Systems Medicine Approaches

Florian Auer, Tim Beißbarth and Frank Kramer

Department of Medical Statistics, University Medical Center Göttingen

florian.auer@med.uni-goettingen.de

In modern Systems Medicine approaches the aim is to look at increasingly complex interactions of complete signaling pathways in order to get a more holistic view for individualized treatment decisions. Individualized treatment decisions and newly developed specialized drugs warrant the need to broaden the focus in individualized medicine from singular biomarkers to pathways.
On the other hand pathway databases offer vast amounts of knowledge on biological networks, freely available and encoded in semi-structured formats [BCS06, SAK+09]. The efficient re-use of pathway knowledge and its integration into bioinformatic analyses enables new insights for researchers in systems medicine.
However, the vast amount of published data on molecular interactions makes it increasingly challenging for life science researchers to find and extract the most relevant information. Currently, the tools to use this information and integrate it in a clinical context are still lacking.
Our idea is to compose an analysis pipeline in order to enable patient-specific systems medicine analyses in a university hospital setting. Our poster will present a workflow for visualizing pathway information and integrating omics data within an interactive online application, utilizing state of the art technology[FLH+16, R C14, KBK+ 13, FBBL15] and well-established standard data models[DCP+ 10, HFS+ 03, PCW+ 15].

References

[BCS06] Gary D. Bader, Michael P. Cary, and Chris Sander. Pathguide: a pathway resource list. Nucleic Acids Research, 34(Database issue):D504–506, January 2006.

[DCP+10] Emek Demir, Michael P Cary, Suzanne Paley, et al. The BioPAX community standard for pathway data sharing. Nature Biotechnology, 28(9):935–942, September 2010.

[FBBL15] Silvia Frias, Kenneth Bryan, Fiona S. L. Brinkman, and David J. Lynn. CerebralWeb: a Cytoscape.js plug-in to visualize networks stratified by subcellular localization. Database, 2015:bav041, January 2015.

[FLH+ 16] Max Franz, Christian T. Lopes, Gerardo Huck, et al. Cytoscape.js: a graph theory library for visualisation and analysis. Bioinformatics, 32(2):309–311, 2016.

[HFS+ 03] M. Hucka, A. Finney, H. M. Sauro, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics, 19(4):524–531, March 2003.

[KBK+ 13] F. Kramer, M. Bayerlova, F. Klemm, A. Bleckmann, and T. Beissbarth. rBiopaxParser–an R package to parse, modify and visualize BioPAX data. Bioinformatics, 29(4):520–522, February 2013.

[PCW+ 15] Dexter Pratt, Jing Chen, David Welker, et al. NDEx, the Network Data Exchange. Cell Systems, 1(4):302–305, October 2015.

[R C14] R Core Team. R: A Language and Environment for Statistical Computing. 2014.

[SAK+ 09] Carl F. Schaefer, Kira Anthony, Shiva Krupa, et al. PID: the Pathway Interaction Database. Nucleic Acids Research, 37(Database issue):D674–D679, January 2009.

Reconstruction of hierarchical differentiation processes based on the temporal clonal readout in distinct hematopoietic lineages

Christoph Baldow

Institute for Medical Informatics and Biometry (IMB), Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden
christoph.baldow@tu-dresden.de

The differentiation process of hematopoietic stem cells towards mature blood cells is generally depicted as a hierarchical decision process. Although the principle structure of such a branching tree is widely accepted, increasing evidence about complementary and alternative differentiation pathways question the concept of a fixed hierarchy. Clonal tracing studies allow quantifying to which extend and in which lineages marked hematopoietic stem cells contribute to the production of mature blood cells. However, the reconstruction of a hierarchical decision tree based on the temporal clonal readout in distinct blood lineages is challenging.

We use a computational approach to investigate how and under which conditions the original branching tree can be faithfully reconstructed. To this end we use an agent based model to generate prototypic differentiation processes through a hierarchical decision tree and produce clonal readouts in different lineages over time that closely resample the data available from corresponding experimental and clinical studies. Applying a numerical optimisation procedure, we estimate the extend of necessary data (e.g. with respect to number of marked clones and timing of subsequent measures) that is needed for the reconstruction process in order to obtain identifiable results. We complement the model-based investigations with available data from clonal tracing studies and test our suggested approaches for reconstruction of hierarchical decision trees in real-world settings.

Our analysis is a necessary and complementary prerequisite to support experimental approaches for the reconstruction of hierarchical, hematopoietic decision trees. In particular, the modeling approach allows, based on the available data, to estimate whether certain decision processes can be uniquely reconstructed.

Using Big Data to Predict Risks of Future Comorbidities of Epilepsy Patients

Thomas Gerlach, Holger Fröhlich

UCB Pharma GmbH

t.gerlach1@gmx.de

Epilepsy is a complex brain disorder characterized by repetitive seizure events. Epilepsy patients often suffer from various and severe physical and psychological co-morbidities. While general comorbidity prevalences and incidences can be estimated from epidemiological data, such an approach does not take into account that actual patient specific risks can depend on various individual factors, including medication. This motivates to develop a machine learning approach for predicting risks of future comorbidities of epilepsy patients.

In this work we used Big Data from electronic health care records (> 4 Million observations), which provide a time resolved view on an individual's disease and medication history. We enriched these data with information from several databases (DisGeNet, TTD, KEGG, DrugBank, SIDER, …) to capture putative biological effects of observed diseases and applied medications and extracted >5,000 features from >40,000 epilepsy patients. Having performed a first descriptive analysis of our data we plan as a next step to compare different machine learning approaches, such as Random Survival Forests and deep learning techniques for predicting future comorbidity occurrence after first epilepsy diagnosis. Given the observed differences in comorbidity incidences for different anti-epileptic drugs (AEDs), the approach may be used to optimize treatment w.r.t. comorbidity risk profiles of individual patients in the future.

Altogether we see this project as a step towards a better personalized treatment of epilepsy patients.

Mathematical modelling of ageing-related changes in the polarity of haematopoietic stem cell divisions

Ingmar Glauche

ingmar.glauche@tu-dresden.de

Hematopoietic stem cells (HSCs) are able to maintain their own population while at the same time they are able to generate specialized progeny. Conceptually, asymmetric divisions allow one daughter cell to become differentiated while the other retains stem cell potential in contrast to symmetric divisions promoting daughter cells to adopt equivalent stem cell fates. One intensely debated paradigm holds that asymmetric distribution of cellular components to the daughter cells determines their fate, although experimental evidence in the hematopoietic system is rare. Recent findings show that quiescent not-cycling young HSCs present with a clear polar distribution of Cdc42, tubulin and other polarity proteins in the cytosol. Moreover, it was demonstrated that the polar distribution of Cdc42 depend on Cdc42 activity levels, which are progressively lost in aged HSCs. We hypothesize that the targeted alteration of Cdc42 activity level, which translates into polarity changes, represents a regulatory principle to influence the balance of asymmetric/symmetric HSC divisions upon ageing and rejuvenation.

We designed a mechanistic mathematical model to investigate a potentially causative link between cell polarity, mode of division and resulting fates of the daughter cells. We suggest an ODE-based model, in which a transcriptional auto-regulative feedback for total Cdc42 concentration establishes the bistability required for the differentiation switch. Under the assumption that Cdc42 activity is directly coupled to protein polarity, we are able to demonstrate that the symmetry of division and cell fate is determined by the Cdc42 concentration right after cell division. If this is substantially altered between daughter cells due to an asymmetric mode of division, distinct states are achievable for the cells. The model predicts that (I) asymmetric outcomes should be more likely for young HSCs, while (II) a higher incidence of symmetry divisions in aged HSCs preferentially promotes self-renewal and leads to a progressive expansion of this cell type.

Together with the experimental data our model suggests that Cdc42 has multiple features: while on one hand it appears as a direct regulator of cell polarity, on the other hand it also qualifies as a potential candidate driver to manifest the division-induced decision between HSC self-renewal and differentiation. This work provides the first evidence that disruption of the ability of undergoing asymmetric division is linked to stem cell ageing and rejuvenation and to the establishment of cell polarity before mitosis. This connection reinforces the possibility that the molecules controlling or modulating cell polarity serve as an important class of regulators to determine stem cell regenerative potential.

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

Zahra Hajseyed Nasrollah, Achim Tresch, Holger Fröhlich

Cologne University

znasroll@uni-koeln.de

Data based learning of the topology of molecular networks, e.g. via Dynamic Bayesian Networks (DBNs) has a long tradition in Bioinformatics (e.g. Pe'er et al., 2001). The vast majority of methods take gene expression as a proxy for protein expression in that context, which is principally problematic. Furthermore, only a fraction of authors focus on causal network reconstruction using interventional data.

Nested Effects Models (NEMs – Markowetz et al., 2005) have in the past been proposed to overcome some of these issues by distinguishing between a latent (i.e. unobservable) signaling network structure and observable transcriptional downstream effects of targeted interventions. NEMs have been developed further by a number of authors (e.g. Fröhlich et al., 2011), but are still rather limited models.

The goal of this project is to come up with a more principled and flexible approach for learning the topology of a dynamical system that is only observable through transcriptional responses of possibly combinatorial and arbitrary complex perturbation experiments. More specifically, we here focus on the situation that the latent dynamical system (i.e. signaling network) can be described as a network of binary state variables with logistic activation functions. We show, how candidate networks can be scored efficiently in this case and how topology learning can be achieved via adaptive Markov Chain Monte Carlo (MCMC). In the future we plan to extend our method to incorporate multi-omics data and apply it to TCGA patient samples to identify disease related networks.

Modeling individual time courses of thrombopoiesis during multi-cyclic chemotherapy

Holger Kirsten

Universitaet Leipzig, IMISE

holger.kirsten@imise.uni-leipzig.de

Question

Decreased platelet counts, called thrombocytopenia, is a major dose-limiting side effect of dose-intense cancer chemotherapies. However, standard coursers of many chemotherapies result in considerable variability in drug induced platelets dynamics. A major challenge of individualized medicine is to take all relevant factors into account for optimal risk management using individualized modeling. Actually there is a gap between simple phenomenological predictive population models and complex biomathematical modeling of averaged data. Another problem is an inter-occasional variability of platelets response to multi-cyclic chemotherapy. Current statistical model cannot identify patients in the beginning of the treatment, which display sever thrombocytopenia only at the late cycles (Ziepert et al. 2008). In order to tackle these problems, we revised a biomathematical model of average human thrombopoiesis under chemotherapy (Scholz et al. 2010) towards modelling individual time courses taking into account accumulating long-range effects of chemotherapy on a bone-marrow.

Methods

We fitted model parameters hierarchically. 25 population and 10 individual parameters were estimated using dense time series of three patients treated with BEACOPP chemotherapy (Engel et al. 1999). We used these estimators as population parameters as well as prior values in order to fit individually 12 parameters for sparser data of selected 135 patients from the German non-Hodgkin‘s lymphoma trial group applying CHOP-like chemotherapies (Pfreundschuh et al. 2004a; Pfreundschuh et al. 2004b). Individual deviations from treatment protocol were also considered. The individual parameter estimations used simultaneously information from other studies published in literature by incorporating it into the respective weighted likelihood functions, assuming a virtual participation of the patients in the corresponding experiments. This additional information included osteoblasts count change during multi-cyclic chemotherapy (Li et al. 2015), average dynamics of TPO, platelets and megakaryocytes of recombinant-TPO treated healthy patients (Harker et al. 2000) as well as a platelets dynamic after labeled platelets transfusions to patients with different degrees of thrombocytopenia (Hanson and Slichter 1985).

Results

Several new biological insights were discovered and modeled. We have hypothesized a bi-phasic stimulation of thrombopoiesis and we described detailed megakaryocytes populations of ploidies 2-128 as well as proplatelets. We proposed complex TPO-mediated regulation of megakaryocytes’ commitment to either endomitosis or to proplatelets formation, or to the hypothesized dormant

state. The cumulative decrease in average platelets level during multi-cyclic chemotherapy was attributed to interactions between quiescent and active stem cells compartments as well as to the accumulating injures of bone-supporting osteoblasts. We found an optimal tradeoff between goodness of fit and overfitting for most of the patients.

Conclusions

We established a model of individual thrombopoiesis response to chemotherapy. Heterogeneity between patients can be traced back to heterogeneity of a few model parameters. These long-term systems changes during multi-cyclic poly-chemotherapies explain increasing thrombocytopenia severity in the later treatment cycles comparing to that of the early ones. This model is being used currently for a posteriori next-cycle dosage adjustment and we are studying its predictive potential.

References

Engel, C.; Loeffler, M.; Franke, H.; Schmitz, S. (1999): Endogenous thrombopoietin serum levels during multicycle chemotherapy. In: British journal of haematology 105 (3), S. 832–838.

Hanson, S. R.; Slichter, S. J. (1985): Platelet kinetics in patients with bone marrow hypoplasia: evidence for a fixed platelet requirement. In: Blood 66 (5), S. 1105–1109.

Harker, L. A.; Roskos, L. K.; Marzec, U. M.; Carter, R. A.; Cherry, J. K.; Sundell, B. et al. (2000): Effects of megakaryocyte growth and development factor on platelet production, platelet life span, and platelet function in healthy human volunteers. In: Blood 95 (8), S. 2514–2522.

Li, Sidan; Zou, Dehui; Li, Changhong; Meng, Hengxing; Sui, Weiwei; Feng, Sizhou et al. (2015): Targeting stem cell niche can protect hematopoietic stem cells from chemotherapy and G-CSF treatment. In: Stem cell research & therapy 6, S. 175. DOI: 10.1186/s13287-015-0164-4.

Pfreundschuh, Michael; Trumper, Lorenz; Kloess, Marita; Schmits, Rudolf; Feller, Alfred C.; Rube, Christian et al. (2004a): Two-weekly or 3-weekly CHOP chemotherapy with or without etoposide for the treatment of elderly patients with aggressive lymphomas: results of the NHL-B2 trial of the DSHNHL. In: Blood 104 (3), S. 634–641. DOI: 10.1182/blood-2003-06-2095.

Pfreundschuh, Michael; Trumper, Lorenz; Kloess, Marita; Schmits, Rudolf; Feller, Alfred C.; Rudolph, Christian et al. (2004b): Two-weekly or 3-weekly CHOP chemotherapy with or without etoposide for the treatment of young patients with good-prognosis (normal LDH) aggressive lymphomas: results of the NHL-B1 trial of the DSHNHL. In: Blood 104 (3), S. 626–633. DOI: 10.1182/blood-2003-06-2094.

Scholz, Markus; Gross, Arnd; Loeffler, Markus (2010): A biomathematical model of human thrombopoiesis under chemotherapy. In: Journal of theoretical biology 264 (2), S. 287–300. DOI: 10.1016/j.jtbi.2009.12.032.

Ziepert, M.; Schmits, R.; Trumper, L.; Pfreundschuh, M.; Loeffler, M. (2008): Prognostic factors for hematotoxicity of chemotherapy in aggressive non-Hodgkin's lymphoma. In Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 19 (4), pp. 752–762. DOI: 10.1093/annonc/mdm54.

A mathematical model of LPS tolerance involving indoleamine 2,3- dioxygenase activation along the kynurenine pathway in pigs.

Dana Kleimeier

Institut of bioinformatics University medicine Greifswald

dana.kleimeier@web.de

In general, tolerance towards recurrent bacterial infection is mediated by different mechanisms in the cell ranging from changes in chromatin profiles over alterations in the expression of genomic regulatory elements to changes in the cell metabolism, such as the activation of the kynurenine pathway. This pathway is activated by different cytokines, espacially TNF-α and IFN-γ and inflammatory stimuli like lipopolysaccharides (LPS). All can activate the rate limiting protein Indoleamine 2,3-dioxygenase (IDO), which is upregulated in placental compartements as well as in tumor cells. To understand activation of IDO and connections between the proteins of the kynurenine pathway, different cytokines and immune cells, we developed a mathematical model. We use this model to address different questions regarding the development of tolerance and its effect
as well as IDO-up-regulation and down-regulation (via 1-methyl-L-tryptophan).

Meta-Analysis of West-Nile-Virus Infection Induced Gene Expression Changes

Robin Kosch[1], Julien Delarocque[1], Stefanie Becker[2], Klaus Jung[1]

1) Institute for Animal Breeding and Genetics, 2) Institute for Paracitology
University of Veterinary Medicine Hannover

robin.kosch@tiho-hannover.de

Meta-analysis is frequently used to combine two or more independent datasets to obtain more precise estimates in a statistical model and thus to increase the level of scientific evidence. Therefore we used meta-analysis to reanalyze microarray data from public repositories (GEO, ArrayExpress), with the intention to find differential expressed genes. Our focus lay on the upcoming West-Nile-Virus (WNV), which can cause severe neurological diseases in humans and horses[1].
With the help of a systematic databank search, we found two comparable datasets of WNV-infected mice and performed a meta-analysis with them. After merging the raw data, they were processed by quantile-normalizing and removal of the batch effects. A differential analysis resulted in 63 differentially expressed genes, which we compared to the results of a third WNV study[2] with RNAseq data. 17 genes intersected with the genes identified by the performed meta-analysis. By studying the overall of three experiments, we achieved a higher confidence.

[1] Hayes, E. B., & Gubler, D. J. (2006). West Nile virus: epidemiology and clinical features of an emerging epidemic in the United States. Annu. Rev. Med., 57, 181-194.

[2] Kumar, M., Belcaid, M., & Nerurkar, V. R. (2016). Identification of host genes leading to West Nile virus encephalitis in mice brain using RNA-seq analysis. Scientific reports, 6.

Infeering clonal mutations from in silico model of intratumour heterogeneity

Luka Opasic, Benjamin Werner, David Dingli, Arne Traulsen

Max-Planck-Institute for Evolutionary Biology

opasic@evolbio.mpg.de

The discovery of cancer heterogeneity and clonal evolution added a further layer of complexity upon biology of cancer and its treatment. Thus, ability to correctly distinct clonal from subclonal mutations became important for making the right therapeutic choice.

Here, we investigate the effect of biopsy size on the ability to identify truly clonal alteration from multi-region profiling of tumours. We simulated neoplastic growth and dynamics of tumour heterogeneity in structured and unstructured populations in order to generate mutational profiles. Different sampling strategies were then compared on generated data.

We found that biopsy size affects the ability to correctly estimate the sizes of two populations after the first bifurcation in phylogenetic history of each tumour, which the crucial parameter for classification of clonal mutations. In all simulated scenarios accuracy of estimation decreased or stayed the same as the biopsy size increased. Our model suggests to reduce the biopsy size in order to reduce the probability to misclassify truly clonal alterations.

Longitudinal analysis of cytokine profiles in community-acquired pneumonia

Maciej Rosolowski, Holger Kirsten, Peter Ahnert and Markus Scholz

IMISE, University Leipzig

maciej.rosolowski@imise.uni-leipzig.de

Factors influencing the clinical course of community-acquired pneumonia (CAP) are still incompletely known. Some patients recover quickly, whereas others require intensive care unit admission or uccumb to the illness. Inflamatory response of the host is one of the factors linked to the severity of CAP.

In the PROGRESS project, multiple measurements in a cohort of ~400 CAP patients were collected over 4 days after admission to the hospital. The measurements include, among others, parameters reflecting the state of multiple organs, summarized in the SOFA score and levels of 10 cytokines.
We investigate two questions: 1) what are the associations among the cytokines and between the cytokines and the performance parameters of the patients on the same day and shifted in time, 2) are the levels of the cytokines predictive of the clinical course of CAP. An answer to the first question could be used for better modeling of the immune response in CAP. The second question is clinically relevant.

We present first results of the analysis and report a number of associations which are supported by the recent literature on CAP and also some which are less known. The multivariate longitudinal nature of the data suggests a possible extension of the analyses to the exploration of lagged associations which we plan to pursue as the next step.

KALIS – a web-based system for patient-specific risk assessment of drugs

Alban Shoshi

Bielefeld University

alshoshi@techfak.uni-bielefeld.de

In Germany, it is well-known and discussed by several studies that adverse drug reactions cause 3-5% of all hospitalizations which in turn lead to more than 16,000 drug-related deaths per year. Approximately one in two deaths could be avoided by an appropriate system for health professionals. For this purpose, we have developed a web-based system named KALIS for patient-specific risk assessment of drugs. The core of KALIS represents a data warehouse which integrates several drug-related pharmacological and biomolecular databases. Based on this comprehensive knowledge, KALIS provides efficient modules to analyze multiple drugs and make an evaluation in terms of risks. KALIS is meant to be a web-based system for health professionals and researchers in order to reduce risks of adverse drug reactions and to achieve improvements in patient safety.

Predicting the influence of combination therapies in signaling networks

Maren Sitte

University Medical Center Göttingen

maren.sitte@med.uni-goettingen.de

A german-wide consortium named “Molecular Mechanisms in Malignant Lymphomas - emonstrators of Personalized Medicine” compound of research groups of biologists, bioinformaticians and doctors propose to develop prognostic and diagnostic platforms that guide treatment decisions and that support the process of therapeutic target identification in diffuse large B-cell lymphomas (DLBCL). The focus lies on the DLBCL microenvironment as prognostic relevance, which is the foundation of the diagnostic platforms the consortium will establish. The communication of the cell microenvironment with the tumour cells will be the target for the novel therapeutic strategies the consortium wants to investigate.
In our subproject, we aim to investigate hybrid-models, which will integrate signaling data with existing gene expression data to predict how lymphomas translate signaling stimuli in expression phenotypes. For this approach we will integrate pathway knowledge and experimental data and implement previously developed network reconstruction methodology. These existing approaches
as Deterministic Effects Propagation Networks (Bender et al., 2011) and Nested Effects Models (Fröhlich et al., 2008; Markowetz et al., 2005) are based on Bayesian networks. This is the ground line of my research and shall be adapted, so that measurements from phosphoproteomic experiments and prior pathway knowledge can be combined.

Decoding cellular dynamics of EGF signaling by pathway-based integration of proteomics and transcriptomics time-series data

Astrid Wachter, Tim Beissbarth

Department of Medical Statistics, Georg-August-University Göttingen, Germany

Astrid Wachter@med.uni-goettingen.de

Epidermal growth factor (EGF) signaling is studied extensively as its dysregulation is associated with many human malignancies. Systems biology aims to get a more complete picture of such regulatory signaling through the integration of omics data generated on different cellular levels. Time-series information can facilitate an even deeper characterization of biological processes as causal relationships can be inferred.

We investigated a public data set comprising time-course mass-spectrometry and microarray data from EGF signaling in human mammary epithelial cells. Our approach for the integrative analysis of data from different high-throughput platforms is based on pathway and interaction models from public databases. It is implemented in the R software tool pwOmics that enables a pathway-based level-specific data comparison of coupled human proteomic and genomic/transcriptomic data sets. Separate analyses on the functional levels of pathways, transcription factors and genes/transcripts allow an integrated cross-platform consensus analysis. Via network reconstruction and inference methods consensus graphical networks are generated that provide detailed insight into dynamic regulatory processes.

Temporal profiles of EGF signaling consensus molecules could show previously known activity patterns but also hint to further direct or indirect auto-feedback regulations. Time profile clustering identified four dynamic co-regulation patterns that determine EGF signaling. Via dynamic consensus analysis we could furthermore generate a probabilistic network that reflects activating and inhibiting relationships between consensus molecules.

We demonstrate that integration of paired high-throughput time-series data with public database knowledge enables a comprehensive interpretation of time-dependent signaling and provides a basis to generate hypotheses on the succession of underlying regulatory mechanisms.

A tool for guided therapy adaptation to control for haematotoxic side-effects of multicycle chemotherapy

Kheifetz Yuri

IMISE (Institute for Medical Informatics, Statistics and Epidemiology)

yuri.kheifetz@imise.uni-leipzig.de

Question

Reduced blood cells counts is a major dose-limiting side effect of many dose-intense cancer chemotherapies. Understanding haemathopoiesis during cytotoxic chemotherapy is crucial for the control of thrombocytes, granulocytes or erythrocytes nadir by chemotherapy dose adjustments, therapy postponement, platelet transfusion or growth factors applications. Since current chemotherapy regimens apply drug treatments in few cycles, a posteriori dose adjustment for the next drug applications is very important for controlling next-cycle side effects. A posteriori dose adjustment can be either a prescribed rule of thumb, or a Bayesian predictor based on statistical models of different complexities. Current statistical model cannot identify patients in the beginning of the treatment, which display sever thrombocytopenia only at the late cycles (Ziepert et al. 2008). Existing model-based myelosuppression -guided dose adaptation tool (Wallin et al. 2010, 2009) considers an inter-occasion variability (IOV) of blood cells response to multi-cyclic chemotherapy as a purely random effect, which makes the adaptation method less successful.

Methods & Results

In order to incorporate a huge amount of current biological knowledge to the dose adjustments, we have constructed a standalone GUI application which make Bayesian prediction a posteriori next-cycles dose adjustment by simulating individualized complex mechanistic models of thrombopoiesis and granulopoiesis developed so far by our group (Scholz et al. 2012; Scholz et al. 2010, Kheifetz et al in preparation). Our recent mechanistic thrombopoiesis mechanistic model fills a gap between simple phenomenological predictive population models and complex biomathematical modeling of averaged data. The model assumes long-term systems changes during multi-cyclic poly-chemotherapies, which explain gradual increase of thrombocytopenia severity as a deterministic effect. Heterogeneity between patients is traced back to heterogeneity of a 12 model parameters.

The GUI application utilizes clinical data and data from the literature as prior information to guarantee that the predictions are consistent with current knowledge. The tool is intended to be usable by physicians without modelling background. The user can choose specific study and patient data from few databases and perform simulations and predictions thereon. Analyses of single patients or groups of patients can be made. The tool supports visualization of individual time courses of platelets during multicycle chemotherapy as well as predictions regarding possible modifications of the thera-

py during the next chemotherapy cycle. A posteriori dose adjustment is based on the user's judgment of few simulated next-cycle treatment scenarios. The tool can estimate the maximal dose controlling a prescribed thrombocytopenia or neutropenia grade in the next treatment cycle.

We validated the predictive power of our tool for thrombocytopenia on the basis of 135 patients treated with CHOP or CHOEP chemotherapy. After 3-4 cycles our predictions are highly reliable and 40-200% more precise than those provided by other tools. Simulations of averaged parameters values predict better individual platelets dynamics during cycles 2-6 of chemotherapy than simplistic model –based predictor calibrated on the results of the first cycle.

References

Kheifetz, Yuri; Scholz, Markus; Loeffler, Markus (2010): Modeling individual time courses of thrombopoiesis during multi-cyclic chemotherapy. In preparation.

Scholz, Markus; Gross, Arnd; Loeffler, Markus (2010): A biomathematical model of human thrombopoiesis under chemotherapy. In: Journal of theoretical biology 264 (2), S. 287–300. DOI: 10.1016/j.jtbi.2009.12.032.

Scholz, Markus; Schirm, Sibylle; Wetzler, Marcus; Engel, Christoph; Loeffler, Markus (2012): Pharmacokinetic and -dynamic modelling of G-CSF derivatives in humans. In: Theoretical biology & medical modelling 9, S. 32. DOI: 10.1186/1742-4682-9-32.

Wallin, Johan E.; Friberg, Lena E.; Karlsson, Mats O. (2009): A tool for neutrophil guided dose adaptation in chemotherapy. In: Computer methods and programs in biomedicine 93 (3), S. 283–291. DOI: 10.1016/j.cmpb.2008.10.011.

Wallin, Johan E.; Friberg, Lena E.; Karlsson, Mats O. (2010): Model-based neutrophil-guided dose adaptation in chemotherapy: evaluation of predicted outcome with different types and amounts of information. In: Basic & clinical pharmacology & toxicology 106 (3), S. 234–242. DOI: 10.1111/j.1742-7843.2009.00520.x.

Ziepert, M.; Schmits, R.; Trumper, L.; Pfreundschuh, M.; Loeffler, M. (2008): Prognostic factors for hematotoxicity of chemotherapy in aggressive non-Hodgkin's lymphoma. In Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 19 (4), pp. 752–762. DOI: 10.1093/annonc/mdm54.

Phenotype Prediction using Prior Knowledge-based Causal Networks

Daniel Ziemek

Pfizer Deutschland GmbH

daniel.ziemek@pfizer.com

The quest for omics-based biomarkers in complex diseases outside of cancer has proven difficult. We have recently developed a novel method that leverages a network of prior causal relationships to improve phenotype prediction and biomarker discovery. The method compares favorably to other methods on several clinical datasets and, maybe most importantly, shows better generalization performance to completely independent study data. In this talk, I will outline the general problem and our proposed approach to a solution.