Workshop on

"Computational Models in Biology and Medicine"

September 10th - 11th, 2015 - Leipzig, Germany

Abstracts

(2015-09-09)

Spatio-temporal modeling of infection by human-pathogenic fungi

Marc Thilo Figge

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

During the past two decades the frequency of life-threatening infections in humans that are caused by fungal pathogens has increased significantly. Our current research focuses on modeling the dynamics of infections caused by the two major human-pathogenic fungi, *i.e.* the filamentous fungus *Aspergillus fumigatus* and the yeast *Candida albicans*.

Since experimental investigations of biological systems are nowadays often routinely accompanied by microscopy experiments, we suggest exploiting the valuable information on dynamical, functional and morphological aspects of fungal pathogenicity that is contained in image data. This can be realized by an *image-based systems biology* approach, which seeks to take full advantage of the information contained in microscopy images and thereby establishes an essential connective link between experimental and theoretical examination of biological processes at a quantitative level. In general, *image-based systems biology* includes the following aspects [1]: (i) acquisition and automated analysis of image data for high-content and high-throughput screening, (ii) quantitative description of biological processes by appropriate characteristic measures, and (iii) construction of image-derived spatio-temporal models and predictive computer simulations.

In order to unravel the fungal infection dynamics by means of quantitative predictions derived from to-scale models, we apply and combine various approaches, such as differential equations for the spatio-temporal dynamics of continuous variables, state-based models to quantify the immune network and agent-based models to simulate the interaction between discrete individuals in space and time. In this talk, we describe models for the kinetics of *C. albicans* infection on oral epithelial cells and show that epithelial invasion outcompetes hypha development in this early stage of infection [2]. At the stage of bloodstream infections, we quantify innate effector mechanisms by state-based and agent-based models, and test hypotheses for the observation of *C. albicans* immune escape in human blood [3,4]. Regarding the human-pathogenic fungus *A. fumigatus*, we study the early immune response against inhaled conidia, by means of a hybrid agent-based model for the human alveolus [5,6]. A key readout of the associated *in silico* experiments is the first-passage-time of alveolar macrophages that are searching for conidia in the alveolus. Testing for various migration modes and parameter combinations of alveolar macrophage dynamics, we arrive at the prediction that these phagocytes must be guided by chemotactic signals in order to achieve the clearance in due time before germination of conidia.

On a general note, we want to encourage that investigations of biological systems, which are nowadays often routinely accompanied by microscopy experiments, should exploit the valuable information contained in image data by quantitative analyses and construction of to-scale models for spatio-temporal simulations.

[1] Medyukhina A, Timme S, Mokhtari Z, Figge MT* (2015) Image-based Systems Biology of Infection. *Cytometry A* 87, 462-470.
[2] Mech F, Wilson D, Lehnert T, Hube B, Figge MT* (2014) Epithelial invasion outcompetes hypha development during Candida albicans infection as revealed by an image-based systems biology approach. *Cytometry A* 85, 126-139.
[3] Hünniger K[+], Lehnert T[+], Bieber K, Martin R, Figge MT*, Kurzai O* (2014) A virtual infection model quantifies innate effector mechanisms and Candida albicans immune escape in human blood. *PLoS Comp Biol* 10(2), e1003479.
[4] Lehnert T[+], Timme S[+], Pollmächer J, Hünniger K, Kurzai O, Figge MT* (2015) Bottom-up modeling approach for the quantitative estimation of parameters in pathogen-host interactions. *Frontiers in Microbiology* 6 (608).
[5] Pollmächer J, Figge MT* (2014) Agent-Based Model of Human Alveoli Predicts Chemotactic Signaling by Epithelial Cells during Early *Aspergillus fumigatus* Infection. *PLOS ONE* 9(10), e111630.
[6] Pollmächer J, Figge MT* (2015) Deciphering chemokine properties by a hybrid agent-based model of *Aspergillus fumigatus* infection in human alveoli. *Frontiers in Microbiology* 6 (503).

Mechanistic modelling of the communication between alveolar macrophages and epithelial cells during Legionella pneumophila infection

Xin Lai

Laboratory of Systems Tumour Immunology - University Hospital Erlangen

Pneumonia is an acute inflammatory lung disease provoked by infection with different pathogens, including Legionella pneumophila (L. pneumophila). The invading of L. pneumophila into the lung triggers the response of resident alveolar macrophages, which produce pro-inflammatory cytokines, such as IL-1β. However, the mechanism by which the macrophages communicate with surrounding epithelial cells in the lung to keep a tight control of the local inflammatory response remains to be further elucidated. In this study, we combined experimental data with mathematical modelling to dissect the features of the NF-κB signalling mediated process underlying this mechanism.

We found that alveolar macrophages can cause the tolerance of lung epithelial cells via IL-1β. After recognising IL-1β, quick degradation of IRAK1 protein happens within the epithelial cells and blocks further stimulation by bacterial factors, such as flagellin. Moreover, we used the data-driven model to assess the influence of clinically relevant factors, such as single nucleotide polymorphisms (SNPs) within the IRAK1 gene altering its protein stability, on the lung inflammatory response induced by L. pneumophila.

Modelling lymphoma therapy and immune response

Markus Scholz
Institute for Medical Informatics, Statistics and Epidemiology - University of Leipzig

Moderate intensifications improved the outcome of lymphoma chemotherapy, but highly intense therapies are inferior. We hypothesise that the immune system has a key role in controlling residual tumour cells after treatment. More intense therapies result in a stronger depletion of immune cells allowing an early re-growth of the tumour.

To understand this process in more detail, we propose a differential equations based model of the dynamics and interactions of tumour and immune cells under chemotherapy. Major model features are an exponential tumour growth, a modulation of the production rate of effector cells by the presence of the tumour (immunogenicity) and mutual destruction of tumour and immune cells. Chemotherapy causes damage to both, immune and tumour cells. Immuno-therapy by the monoclonal antibody Rituximab is modelled by a direct cell kill and an intensified immune response. Growth rate, chemosensitivity, immunogenicity and initial size of the tumour are assumed to be patient-specific, resulting in heterogeneity regarding therapy outcome. Maximum-entropy distributions of these parameters were estimated on the basis of clinical survival data.

The resulting model can explain the outcome of eight different chemotherapeutic regimens with and without Rituximab and corresponding hazard-ratios. Estimated parameters are biologically plausible. We demonstrate how the model can be used to make predictions regarding yet untested therapy options.

We conclude that our model explains observed paradox effects in lymphoma therapy by the simple assumption of a relevant anti-tumour effect of the immune system. Heterogeneity of therapy outcomes can be traced back to heterogeneity of a few model parameters whose distribution can be estimated on the basis of clinical survival data. The model can be used to predict the performance of new therapy options.

Heterogeneity and cell fate control in mouse embryonic stem cells

Ingmar Glauche
Institute for Medical Informatics and Biometry - TU Dresden

Bistability is a characteristic feature of many molecular switches facilitating cell decision processes as they occur during differentiation and in response to external stimuli. Undifferentiated mouse embryonic stem cells (ESC) are a typical example, in which such a bistable situation is caught in an in vitro setting, thus contributing to a heterogeneous albeit dynamically stabilized cell culture. In particular, it is the expression of pluripotency factors like Nanog and Rex1 that obeys a bimodal distribution, which is reestablished after cell sorting.

We study heterogeneity and cell fate control in ESC using a variety of methods including cell culture experiments, flow cytometry, live-cell imaging, and quantitative image analysis. Complementary we developed a multiscale, spatial mathematical model of ESC growth, which allows comparing experimental results with our model predictions. The integrated, agent-based model serves as a framework to link population features, such as proliferation rates and spatial arrangements, to reveal potential transcription factor related cellular and intercellular mechanisms behind the emergence of observed patterns that cannot be derived from experimental data directly.

Modelling epigenetic control of HSC lineage specification

Jens Przybilla
Interdisciplinary Center for Bioinformatics - Leipzig University

Lineage specification of stem cells has been shown to be under control of both transcription factor (TF) networks and histone modification pattern. In particular, bivalent genes, that are associated with histones that carry both tri-methylation of H3 at lysine 4 (H3K4me3) and at lysine 27 (H3K27me3) have been implicated in this process. A mechanistic understanding of the regulation of these genes is currently missing. Here, we provide a computational model of hematopoietic stem cell (HSC) populations which describes their specification into the myeloid and lymphoid lineage based on epigenetic feedback on the transcriptional regulation of bivalent genes.

We assume that lineage specification depends on changes in the activity of epigenetic modifiers. Accordingly, we model the specification of HSCs into their differentiated myeloid and lymphoid progeny assuming characteristic changes of histone modifying and DNA methylating enzymes. Building on our multi-scale model of transcriptional regulation in HSCs, in our simulations, each cell contains an artificial genome. The expression of this genome depends on the cell's epigenome which is represented by three histone modifications (H3K4me3, H3K9me3 and H3K27me3) cross-talking with DNA methylating enzymes.

While the model is not capable of deciding on whether lineage specification is initiated by TFs or chromatin reorganization, it suggests that intimate feedback loops between these regulatory layers stabilize lineage specific transcriptional programs. Consequently, aberrant expression of epigenetic modifiers will impact cell fate decisions potentially leading to stem cell loss of function. As an example, we demonstrate how disturbed histone modification processes might result in a differentiation block as seen in acute myeloid leukemia. Moreover, we show how accelerated proliferation can destabilize histone modifications and therewith effects lineage specification processes.

The developing endoderm in the zebrafish embryo

Konstantin Thierbach
Institute for Medical Informatics and Biometry - TU Dresden

The zebrafish (ZF), danio rerio, is a popular model organism in developmental biology. Intensive image-based studies of the embryogenesis of wild-type and mutant embryos have been made over the last decades. However, former microscopy technologies only allowed for partial analysis of embryonic development due to several problems concerning image acquisition. Under these conditions the image-based studies could only qualitatively describe ZF embryogenesis or the quantitative analysis was limited to spatially confined regions, thus requiring further generalization.

The development of selective plane illumination microscopy (SPIM) has revolutionized the analysis of the ZF embryo and allows to investigate embryogenesis at a very high temporal resolution. This technology also enables recordings of 4D time series data of many embryos without interference. In turn, such time-series data builds the basis for quantitative analysis of the embryonic development on a cellular level.

Our goal is a quantitative, image based analysis of the ZF endoderm during gastrulation with a particular focus on cellular specification and patterning. Therefore we extract single cells and the flow of cell movement. Furthermore, the methodology for a thorough description of cell numbers, cell densities and cell motility development over time for a wild-type embryo is shown.

In a next step we apply a refined methodology for the single-cell based analysis of certain areas which appear more interesting in the structural patterns.
We used our established techniques to analyze a larger number of embryos in order to get insight in the variability among embryos. This approach provides a more detailed understanding of similarities and variabilities in endodermal development and helps to reveal principle mechanisms of ZF embryogenesis.

Visual Analytics for Relations in High Dimensional Data Spaces

Shenghui Cheng
Department of Computer Science - Stony Brook University

The growth of digital data is tremendous. These data come from many aspects of life and matter such as medical records, environment monitoring, business market, social networks etc. It is a challenge for humans to understand the intricate relations among data items and variables on such a large scale. Visual analytics can offer powerful mechanisms to assist humans in the exploration of these complex data, by mining the relations from the raw data and sculpting them as visualizations to help human gain insight. In this talk, I will first describe several techniques that can visualize the relations among the data items and variables in high dimensional data space. Then I will focus on contextual relationships and propose our tool called "The Data Context Map" which can fuse the data items and variables together and operate as a semantic map. By means of interaction, human can appreciate accurate contextual relations embedded in the high dimensional data space.

Pathway-based integration of time-series omics data using public database knowledge

Astrid Wachter
Department of Medical Statistics - University Medical Center Göttingen, Germany

The increased generation of omics data on different functional levels of the cell represents a constantly growing challenge for their analysis and interpretation. Time-series measurements add another dimension of complexity but likewise enable deeper characterization of biological processes. We developed a straightforward systems biology approach for the integrative analysis of time-series data from different high-throughput technologies based on pathway and interaction models from public databases. Implemented in our software tool 'pwOmics' this approach performs pathway-based level-specific data comparison of coupled human proteomic and genomic/transcriptomic data sets. Separate downstream and upstream analyses are performed on the functional levels of pathways, transcription factors and genes/transcripts and integrated in the cross-platform consensus analysis. Via network reconstruction and inference methods (steiner tree, dynamic bayesian network inference) consensus graphical networks provide detailed insight into dynamic regulatory processes.

With this approach we investigated a public data set comprising time-course mass-spectrometry and microarray data from EGF signaling in human mammary epithelial cells. Understanding the dynamics of the underlying physiological signaling mechanisms helps to characterize dysregulation processes in EGF signaling, which are observed in many human malignancies, e.g. cancer. Regulatory consensus profiles could be identified that help understanding complex pathway interdependencies and feedback mechanisms. Integration of coupled high-throughput time-series data enables a highly comprehensive interpretation of time-dependent signaling. Our approach exploits public database knowledge and cross-platform omics data in order to generate hypotheses on the succession of underlying pathway interplay mechanisms.

Probabilistic Models for Biological High Throughput Data

Holger Fröhlich
Institute for Computer Science, c/o Bonn-Aachen International Center for IT - University of Bonn

Modern biological high-throughput techniques allow for measuring large sets of molecular features in a massively parallel manner. Such omics data can cover different biological aspects (e.g. gene expression, DNA methylation) of a biological system and are of high relevance in modern biomedicine. Probabilistic models could help to integrate heterogeneous omics data, to decipher parts of the underlying complex biological system and to derive experimentally verifiable predictions. As a first example I will consider the question, how miRNA and gene expression data can be used together to infer the regulatory activity as well as the network dependencies of transcription factors and miRNAs. I propose biRte as a computationally attractive approach, which combines Bayesian inference with network reverse engineering.

Molecular features represent parts of a complex biological system. Classically, ordinary differential equations (ODEs) have been used to describe such systems mechanistically. However, a major difficulty lies in the fact that any biological sub-system is embedded into and influenced by the surrounding system. I will present a probabilistic approach, which aims to detect these latent influences based on time series data. Moreover, I will demonstrate that the method can be also used to estimate missing and wrong reactions in ODE systems.

My last example focuses on reverse engineering of biological networks from perturbation data. Besides molecular data there is an increasing interest in imaging based techniques for this purpose. I will present a probabilistic graphical model, which can be used to learn the structure of a biological network from such data.

Systematically evaluating interfaces for RNA-seq analysis from a life scientist perspective

Alicia Poplawski
Institute of Medical Biostatistics, Epidemiology and Informatics - University Medical Center Johannes Gutenberg, University Mainz, Germany

Massively parallel RNA-sequencing (RNA-seq) has become the instrument of choice for transcriptome analysis and a cornerstone of modern life science laboratories, but it is generally carried out by bioinformaticians. One of the reasons is the huge amount of data generated. For example, sequencing platforms can generate terabytes of data in a single sequencing run, where many datasets are sequenced in the scope of a single project. For extracting useful information out of this massive amount of data, substantial computational skills and resources are required. Nevertheless, some of the analysis steps are accessible to life science researchers without extensive knowledge of command line tools. With sequencing costs being constantly reduced and sequencing speed and efficiency rising exponentially, it might become more important to shift at least some analysis steps from core facilities to life scientists, using standardized tools with easy-to-use interfaces.

We performed a systematic search and evaluation of such interfaces to investigate to what extent these can indeed facilitate RNA-seq data analysis even for users without extensive computer-science background. We defined criteria for a detailed evaluation of more widely used interfaces. Central criteria were ease of configuration, documentation, usability, computational demand, and reporting.

We found a total of 29 open source interfaces, and 6 of the more widely used interfaces were evaluated in detail. No interface scored best in all of these criteria, indicating that the final choice will depend on the specific perspective of users and the corresponding weighting of criteria. Considerable technical hurdles had to be overcome in our evaluation. For many users this will diminish potential benefits compared to command line tools, leaving room for future improvement of interfaces.

Genome-wide eQTL analysis

Holger Kirsten
Institute for Medical Informatics, Statistics and Epidemiology - University of Leipzig

Dissecting the Genetics of the Human Transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding loci Markus Scholz Institute for Medical Informatics, Statistics and Epidemiology / The Leipzig Research Centre for Civilization Diseases - University of Leipzig Genetics of gene expression (eQTLs) has proved an indispensable tool for understanding biological pathways and pathomechanisms of trait associated SNPs.

However, power of most genome-wide eQTL studies is still limited. Therefore, our aim was to perform a large eQTL study in peripheral blood mononuclear cells of 2,112 individuals increasing the power to detect trans-effects genomewide. Furthermore, going beyond univariate SNP-transcript associations, we aimed to investigate relations of eQTLs to biological pathways, polygenetic effects of expression regulation, trans-clusters, and enrichment of co-localised functional elements.

We used Illumina HT-12-v4 Expression BeadChips and Affymetrix Axiom CEU SNP-chips imputed to HapMap2 CEU. After thorough pre-processing, 2,625,374 autosomal SNPs and 18,738 probes corresponding to 13,338 genes were included. We found eQTLs for about 85% of analysed genes, 18% of genes were trans-regulated. Local eSNPs were enriched up to 5 Mb to the transcript challenging typically implemented ranges of cisregulations. Pathway enrichment within regulated genes of GWAS-related eSNPs supported functional relevance of identified eQTLs.

We demonstrate that nearest genes of GWAS-SNPs might frequently be misleading functional candidates. We identified novel trans-clusters of potential functional relevance for GWAS-SNPs of phenotypes including obesity-related-traits, HDL-cholesterollevels, and haematological phenotypes. We used chromatin immunoprecipitation data for demonstrating biological effects. Yet, we show for strongly heritable transcripts that little transchromosomal heritability is explained by all identified trans-eSNPs, however, most cis-heritability of these transcripts seems explained. Dissection of co-localised functional elements indicated a prominent role of SNPs in loci of pseudogenes and non-coding RNAs for the regulation of coding genes. Our study substantially increases the catalogue of human eQTLs and improves our understanding of the complex genetic regulation of gene-expression, pathways and disease-related processes.

Hierarchical Modelling

Arnd Gross
Institute for Medical Informatics, Statistics and Epidemiology - University of Leipzig

To assess potential effects of variants in six lipid modulating genes (*SORT1, HMGCR*, *MLXIPL, FADS2, APOE* and *MAFB*) on early development of dyslipidemia independent of the degree of obesity in children, we investigated their association with total (TC), low density lipoprotein (LDL-C), high density lipoprotein (HDL-C) cholesterol and triglyceride (TG) levels in 594 children.

In genetic association studies adjusted for age, BMI SDS and sex, we identified significant associations for rs599839 near *SORT1* with TC and LDL-C and for rs4420638 near *APOE* with TC and LDL-C. We performed Bayesian modelling of the combined lipid phenotype of HDL-C, LDL-C and TG to identify potentially causal polygenic effects on this multi-dimensional phenotype and considering obesity, age and sex as a-priori modulating factors. This analysis confirmed that rs599839 and rs4420638 affect LDL-C.

In summary, we show that variants near *SORT1* and *APOE* influence lipid levels independent of obesity in children. Bayesian modelling suggests causal effects of these variants.

POSTER

Hoor Al-Hasani
Fraunhofer IZI

Estimations show that 1 in 6 men will develop Prostate Cancer (PCa). However, the cause of it is yet unknown. PCa in early stages has no symptoms. The advance states of PCa cause noticeable problems to the man, such as pain when urinating, or bleeding. These symptoms, however, could be mistaken with other prostate diseases (e.g. benign prostatic hyperplasia). Unfortunately for the patient, the diagnose is done through two methods: 1) Prostate-Specific Antigen (PSA), which is used to measure any indication for over-expression in prostate cancer tissues. 2) Digital Rectal Examination (DRE) or Transurethral resection of the prostate (TURP); both are an unpleasant procedures. The luck of a precise "painless" PCa detector, triggered the need to find a biomarker.

We hypothesize that aggressive tumour accumulates more mutations which might permit additional phenotypes and characteristics. For instance size, cell-cell interaction, invasive and metastatic ability, and therapy resistance. Therefore the intratumor heterogeneity could be used as an indicator of the tumour state.

Our observation points to the influence of the type of mutations on the Allelic Heterogeneity (AH) profile of a tissue and in the significant mutated genes (SMG). Once the inhered mutations and germline events were excluded, the AH profile of potential tumour specific mutations could clearly be distinguished.

POSTER

In-silico study of the urothelium with the GGH approach

Angelo Torelli[1], Philipp Erben[2], <u>Markus Gumbel[1]</u>
1Institute of Medical Informatics, University of Applied Sciences Mannheim www.hs-mannheim.de
2Clinic of Urology, Medical Faculty Mannheim at the University of Heidelberg www.umm.de

Bladder cancer is the sixth most common cancer in men [1]. Although cancer can spread from neighboring organs to invade and grow in the bladder, the most common form of bladder cancer originates within the bladder's epithelial walls, the urothelium. Novel models for cell differentiation and proliferation in the urothelium are presented. The models are simulated with the Glazier-Graner-Hogeweg technique [2] using the CompuCell3D framework [3]. Since little is known about the characteristics of each bladder cell type and the urothelial cell lineages [4, 5], various models were set up against each other to compare the biological facts with the hypotheses made in this work. From the variety of tested models, the *contact model* is the best candidate to explain cell proliferation in the healthy urothelium. Based on this model, four variations were compared to highlight the key variations that best fit real urothelium. All simulations were quantified by a fitness function designed for the requirements of the urothelium [6]. The findings suggest that adhesion and a nutrient dependent growth may play a crucial role in the maintenance of the urothelium. Aberrations in either adhesion or nutrient dependent growth led to the development of polyp-like formations. This work mimics the regeneration process and the steady state of the urothelium with a spatial and adhesion dependent approach for the first time and differentiates itself from other approaches applied on other tissues, such as the small intestine [7] and the epidermis [8].

References
[1] Ahmedin Jemal et. al. Global cancer statistics, March, April 2011.
[2] Ariel Balter et. al. The glazier-graner-hogeweg model: Extensions, future directions, and opportunities for further study. Single-Cell-Based Models in Biology and Medicine, 2007.
[3] Maciej H. Swat et. al. Multi-scale modeling of tissues using compucell3d. Computational Methods in Cell Biology, 2012.
[4] Philip Levy Ho et. al. Normal and neoplastic urothelial stem cells: getting to the root of the problem. Nature reviews. Urology, 9(10), 2012.
[5] Puneet Khandelwal et. al. Cell biology and physiology of the uroepithelium. Am J Physiol Renal Physiol, (297), 2009.
[6] Angelo Torelli et. al. Modeling of the urothelium with an agent based approach. Lecture Notes in Computer Science, 2015.
[7] Markus Gumbel et. al. Simulation and 3-d visualization of the intestinal crypt. SCS Publications, 1998.
[8] Xinshan Li et. al. Skin stem cell hypotheses and long term clone survival - explored using agent-based modelling. Sci. Rep., May 2013.

POSTER

Automated characterization of cell tracks and classification by using their position coordinates and velocity components

Shariful Islam
Department of Medical Statistics - University Medical Center Goettingen

  Cell motility is an important factor in many biological processes, such as tissue repair, metastatic potential, chemotaxis, or analysis of drug performance (Nath et al., 2006). During embryonic development, cells migrate to distant locations where they differentiate into different cell types (Zimmer et al., 2006). Cell migration is also crucial to the cells of the immune system. Immune cells migrate extensively during immunosurveillance to search for inflammatory stimuli. Upon encountering such stimuli they mount an immune response which leads to further recruitment of immune cells which also involves cell migration processes. Of note, different immune cell populations, in different milieus may exhibit distinct migratory patterns that are dependent from their functional status. Therefore, in order to decipher the complexity of these biological processes, an unbiased computational approach is required. On this poster we present some methods and ideas of how to cluster and classify cell tracks. We test the utility of these methods on data from intravital imaging of immune cells in the leptomeningeal blood vessels of the spinal cord of Lewis rats. By using three dimensional position-coordinates and velocity components of different kind of blood cells, we have computed several features (such as „volume asphericity", „confinement ratio", as described in Mokhtari et al., 2013) which are used successfully to characterize and classify the tracks.

POSTER

Yuri Kheifetz
Institute for Medical Informatics, Statistics and Epidemiology - University of Leipzig

The aim of this work is the individual parameters estimation for the complex biomathematical thrombopoiesis model (Scholz et al., 2009) using clinical data, combined with a prior information from different sources. Generally, individual parameters estimations are provided mostly for simplistic pharmacometrical and pharmacodynamical models. Individualization of the complex models is limited by the lack of data. In order to overcome this difficulty, 135 CHOP/CHOEP- treated patients with reach data were selected from 1515 patients of NHL study. The inclusion criteria was a sufficiently big number of observation (at least 4 cycles with at least 5 observations per cycle). 3 CHOEP-treated patients (Engel et al., 1999), each with 50 measurements of platelets as well as with 50 measurements of TPO, constituted a data source for the population model.

We analyzed Fisher Information Matrix for the overfitting estimation. It turned out that the above data were insufficient for the reliable parameters estimation of the basic model. We have revised the model according to resent scientific updates as well by improving the goodness of fitting and minimizing the overfitting. We used an average dynamics of TPO, platelets, megakaryocytes and megakaryocytes ploidy of rTPO treated healthy patients (Harker et al., 2000) as a prior data. We used also a platelets dynamics after labeled platelets transfusions to patients with different degrees of thrombocytopenia (Hanson, Slichter 1985). The individual and prior data were included into the weighted likelihood function for each patient during a simultaneous parameters estimation procedure. A few poor identified parameters were fixed from other studies. The weighting of likelihood terms was empirically determined by minimizing a bias in the fitting.

We have found an optimal tradeoff between goodness of fit and overfitting for most of the patients. We have proposed a possible explanation for the long-range effects of chemotherapy.

POSTER

Julia Perera-Bel
Medical Statistics Department - University Medical Center Goettingen

  The understanding of complex diseases, such as cancer, is becoming more comprehensive with the improvements of high-throughput (HT) technologies such as Next-Generartion Sequencing. However, advances in technology platforms and bioinformatic tools contrast with the scarce implementation of cancer genomics in clinical practice. The main reason of this situation is that pathologists and oncologists have to face thousands of genomic alterations and unravel their clinical relevance. Therefore, there is a need to develop novel strategies and settle a framework to prioritize and interpret the complexity of genomic findings in order to allow an appropriate decision which actually improves patient care.

  In the consortium project Genoperspective we want to overcome this bottleneck regarding the implementation of HT data in clinical routine. The consortium wants to address the ethical, social and legal aspects as well as data management, bioinformatic data processing and statistical aspects.

  From the bioinformatic subproject, we are developing a framework for genomic data interpretation and reporting. The major goals are 1) to report omics data in a validated and straightforward way allowing individual clinical decisions; 2) to properly evaluate the impact of omics-based therapy selection methods on clinical practice and, 3) to develop appropriate recommendations and strategies to deal with the aspects raised in Genoperspective.

POSTER

Nab Roshyara
Institute for Medical Informatics, Statistics and Epidemiology - University of Leipzig

   Modern analysis of high-dimensional SNP data requires a number of biometrical and statistical methods such as pre-processing, analysis of population structure, association analysis and genotype imputation. Software used for these purpose often use specific and incompatible input and output formats of data. Therefore data management with multiple format conversions is necessary during analyses.

   In order to support fast and efficient data management and computing summary statistics during SNP data analyses, we developed the software fcGENE using C++ object-oriented programming language. This software simplifies and automates the use of different existing GWA packages, especially when we go through the process of genotype imputation.

   fcGENE transforms SNP data including the imputation results into different formats necessary for GWA anaylsis. More precisely, fcGENE creates basic formats of SNP data required by commonly used tools for SNP data analysis such as PLINK, SNPTEST, HAPLOVIEW, EIGENSOFT, GenABEL and tools for genotype imputation such as MaCH, IMPUTE and others. fcGENE also supports data management and summary statistics. The basic means of quality control on a SNP-wise and sample-wise level for both, raw and imputed genotype data can be performed through this tool. This tool also generates templates of commands required to run specific software packages. We demonstrate the functionality of fcGENE with an example workflow of SNP data analyses and provide a comprehensive manual for commands and applications.

   We have developed a user-friendly open-source software fcGENE, which allows SNP data management, summary statistics and format conversion among different GWA software formats. This software also generates templates of commands necessary to run diverse software packages and supplementary files providing information on basic procedures of quality control. To the best of our knowledge, so far there is no software available with a similar comprehensiveness regarding data and analysis management of SNP data.

POSTER

Mehdi Sadeghi
Biophysics Department - Tarbiat Modares University

  Module detection is one of the most important methods for data mining and decreasing high throughput data dimension. Methods based on correlation analysis are mostly used method for this purpose. Here we used Weighted Gene Correlation Network Analysis (WGCNA) for detecting modules in different stages of prostate cancer, finding correlation between these modules and clinical and pathological factors and finally identifying key hub genes in the progression of prostate cancer. Our dataset comprised microarray gene expression data of prostate cancer in different stages (primary prostate cancer, metastatic prostate cancer and normal prostate tissue). In the first step differentially expressed genes were found in each stage of prostate cancer. With assigning 0.01 p-value and 0.5 fold change thresholds we respectively 504 and 1017 differentially expressed genes could be found for primary and metastatic prostate cancer. These genes were used for detecting modules for two stage of prostate cancer. WGCNA method uses the scale free topology property of biological networks for constructing high correlated gene modules. There are more modules for metastatic prostate cancer than primary tumor because in metastatic phase of cancer more pathways and genes involved in. As we supposed genes in each module has similar and related biological function. This can be seen with functional annotation of genes in each module. Prostate cancer metastatic stage modules show significant correlation with important clinical and pathological variables such as PSA level and tumor stage. Among important modules with high correlation to clinical variables we could find genes with high degree (connection to other genes in the network) and also high gene significance (correlation coefficient) for specific clinical variable. This handful number of genes may be supposed as important key hub genes in the progression of prostate cancer and also as target genes for designing new drugs.

POSTER

Dynamical modelling of the murine immune response to pneumococcal lung infection with and without antibiotic treatment

Sibylle Schirm
Institute for Medical Informatics, Statistics and Epidemiology - University of Leipzig

Pneumonia is considered to be one of the leading causes of death worldwide. The outcome depends on both, proper antibiotic treatment and the effectivity of the immune response of the host. However, due to the complexity of the immunologic cascade initiated during infection, the latter cannot be predicted easily.

We construct a biomathematical model of the murine immune response during infection with pneumococcus aiming at predicting the outcome of antibiotic treatment. The model consists on a number of non-linear ordinary differential equations describing dynamics of pneumococcal population, the inflammatory cytokine IL-6, neutrophils and macrophages fighting the infection and destruction of alveolar tissue due to pneumococcus. Equations were derived by translating known biological mechanisms and assuming certain response kinetics. Antibiotic therapy is modelled by a transient depletion of bacterials.

Unknown model parameters were determined by fitting the predictions of the model to data sets derived from mice experiments of pneumococcal lung infection with and without antibiotic treatment. Time series of pneumococcal population, debris, neutrophiles, activated epithelial cells, macrophages, monocytes and IL-6 serum concentrations were available for this purpose. The antibiotics Ampicillin and Moxifloxacin were considered.

Parameter fittings resulted in a good agreement of model and data for all experimental scenarios. Sensitivities of parameter estimates could be estimated. The model can be used to predict the performance of alternative schedules of antibiotic treatment.

We conclude that we established a biomathematical model of pneumococcal lung infection in mice allowing predictions regarding the outcome of different schedules of antibiotic treatment. We aim at translating the model to the human situation in the near future.

POSTER

Integrative analysis of a GWAS for amino acids and acylcarnitines in whole blood and gene-expression data identifies six novel loci and reveals insight into regulatory mechanisms

Markus Scholz
Institute for Medical Informatics, Statistics and Epidemiology - University of Leipzig

  Emerging data suggests that altered blood levels of amino acids and acylcarnitines are associated with common metabolic diseases in adults. Therefore, our aim was to identify common genetic determinants for blood metabolites to better understand pathways contributing to human physiology and common diseases.

  We applied a targeted mass-spectrometry-based method to analyze whole blood spots for 96 amino acids, acylcarnitines and metabolite ratios in a cohort of 2,107 adults. Using chip-based technology and HapMap-based imputation, we performed genome-wide association (GWA) to identify genetic modifiers of metabolite concentrations. Discovered hits were analyzed for replication in an independent cohort. Furthermore, we used Illumina HT-12 v4 Expression BeadChips to study gene expression in peripheral mononuclear cells in order to analyze causal relations between SNPs, gene-expression and metabolite levels.

  We discovered and replicated six novel loci associated with blood levels of total acylcarnitine, arginine, propionylcarnitine, 2-hydroxyisovalerylcarnitine, stearoylcarnitine, and aspartic acid traits. Several SNPs associated with blood metabolites in our study overlap with previously identified loci for human diseases suggesting a shared genetic basis or pathomechanisms involving metabolic alterations. Causal analyses resulted in following putative causative genes: SLC22A16 for total acylcarnitines, ARG1 for arginine, HLCS for 2-hydroxyisovalerylcarnitine, JAM3 for stearoylcarnitine via a trans-effect at chromosome 1, and PPP1R16A for aspartic acid traits.

  Additionally, we report replication and provide additional functional evidence for ten loci that have previously been published for metabolites in several tissues. The integrative analysis of SNP, gene-expression and metabolite data revealed novel insights into the genetic regulations of human metabolism. At several examples, we provide evidence for metabolite regulation via gene-expression and observed overlaps with GWAS loci for common diseases. In consequence, our findings provide strong candidates for future functional studies directed to understand human metabolism and pathogenesis of related diseases.