POWER Och!DOK programme for lecture:

Introduction to bioinformatics (Winter term 2019/2020)

1. Dr Marek Kochańczyk - 3 x 1.5 h: 7.10.2019, 14.10.2019, 21.10.2019

Overview of data analysis tools.

As a PhD student you will likely have to generate and analyze data. Arguably, the most typical first-choice way of handling experimental data are spreadsheets that poorly document reasoning of their creators, are challenging to automate and extend, and notoriously produce sub-standard quality graphs. Initial lectures of the course are intended to be workshop-style introduction to general-purpose tools for data analysis and visualization, with focus on automation of typical tasks and generation of publication-quality plots. First week will provide a quick introduction to python programming language. During second week, students will have hands-on experience with the standard open-source “data science” stack based on Jupyter notebooks. During the third lecture, two alternatives will be presented: R and Matlab.

2. Dr Tomasz Zieliński - 2 x 1.5 h: 28.10.2019, 4.11.201

Theoretical and practical introduction to COMSOL program

3. Prof. Bartlomiej Wilczynski - 3 x 1.5 h: 18.11.2019, 25.11.2019, 2.12.2019

Global approaches to gene regulation modeling

Mathematical modeling of gene regulation has been done for almost 50 years and there is a long history and rich literature on the subject. In particular, a number of models have been proposed to address the question of what the expected level the regulatory state of the gene product of the cell should be given. This state of the cell can be very simple, as in the logical models (1,2) where the cell state is a binary vector of regulator gene's states, or much more complex, including quantitative gene expression levels and the occupancy state of transcription factor binding sites in gene promoters as in more quantitative models (3). In any case, it is assumed that the output of the gene of interest is a function of the state of the cell, even if some form of stochastic element is frequently included in this function.

From experiments, we know that, at least in the case of multi-cellular eukaryotes, the process of gene regulation depends on many events. Some of these events, such as transcription initiation or elongation, are mostly local with respect to a single gene and could be modeled with gene-oriented variables. However, there are other events that have implications for gene expression, that cannot be solely associated with a single gene. Two such phenomena are distal regulatory elements and global chromatin changes. In case of distal regulatory elements, it has been observed that concurrent binding of several transcription factors to a single DNA element can trigger a transcriptional response of one or more target genes(4). While distal regulatory elements are not strictly gene-associated, and are known to reside sometimes very far from their target genes, they are still localized, in the sense, that any regulatory element has a xed position in the genome. However, there exist more global chromatin changes, mostly related to developmental processes, but also very well described in case of the heat shock stress, where large parts of chromosomes change their state to either enable or disable transcriptional activation of many genes.

In terms of mathematical modeling of both these phenomena - i.e. regulation of genes by multiple distal regulatory elements and larger scale chromatin changes - is still an active area of research, however, there are already some quite clear trends. In our lectures, we will discuss several studies that have dealt with these issues. In particular, we will discuss approaches for identification and activity prediction of regulatory elements (5,6) based on different experimental measurements of the chromatin state, as well as the more global approaches to modelling the chromatin state - both in the supervised manner using Bayesian Networks(7) and unsupervised with Hidden Markov Models (8)

References

[1] Kauffman, Stuart (11 October 1969). "Homeostasis and Differentiation in Random Genetic Control Networks". Nature. 224(5215): 177-178.
[2] Thomas, R (1973). "Boolean formalization of genetic control circuits". Journal of Theoretical Biology. 42: 563-85.
[3] Janssens, H., Hou, S., Jaeger, J., Kim, A. R., Myasnikova, E., Sharp, D., & Reinitz, J. (2006). Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. Nature genetics, 38(10), 1159.
[4] Wilczynski, B., & Furlong, E. E. (2010). Challenges for modeling global gene regulatory networks during development: insights from Drosophila. Developmental biology, 340(2), 161-169.
[5] Bonn, Stefan, Robert P. Zinzen, Charles Girardot, E. Hilary Gustafson, Alexis Perez-Gonzalez, Nicolas Delhomme, Yad Ghavi-Helm, Bartek Wilczynski, Andrew Riddell, and Eileen EM Furlong. "Tissue-specific analysis of chromatin state identies temporal signatures of enhancer activity during embryonic development." Nature genetics 44, no. 2 (2012): 148.
[6] Podsiadło, A., Wrzesien, M., Paja, W., Rudnicki, W., & Wilczynski, B. (2013). Active enhancer positions can be accurately predicted from chromatin marks and collective sequence motif data. BMC systems biology, 7(6), S16.
[7] Wilczynski, B., Liu, Y. H., Yeo, Z. X., & Furlong, E. E. (2012). Predicting spatial and temporal gene expression using an integrative model of transcription factor occupancy and chromatin state. PLoS computational biology, 8(12), e1002798.
[8] Ernst, J., & Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nature methods, 9(3), 215.

4. Dr Pawel Kocieniewski - 3 x 1.5 h: 9.12.2019, 16.12.2019, 13.01.2020

Modeling Combinatorial Complexity - Rule-based Modeling.

Signal transduction pathways often display vast combinatorial complexity due to the number of possible protein-protein interactions and post-translational modifications of individual proteins. This poses serious difficulties in formulating and simulating realistic models of such systems due to the number and complication of the corresponding equations. Rule-based modeling is an approach that alleviates this problem by utilizing generalized rules to automatically generate networks of possible reactions and the corresponding system of equations. We will explore the origins and consequences of combinatorial complexity of biological systems (especially in signal transduction) and familiarize ourselves with some modeling tools used to address it (BioNetGen)

5. Dr Paweł Górecki - 2 x 1.5 h: 20.01.2020, 27.01.2020

Algorithmic and practical aspects of phylogenetic tree inference.

Inferring phylogenetic trees from molecular sequences is one of the most important problems in modern computational biology and bioinformatics. In this lecture, I will present several classical methods for tree inference such as maximum likelihood, neighbor-joining, maximum parsimony and hierarchical clustering with a special focus on algorithmic aspects of tree reconstruction. Apart from the purely computational

features, I will discuss known limitations and broader practical contexts that are often avoided or ignored when using in practice tools based on these methods. In particular, this lecture will also cover the problems of error correction of a tree, tree binarization, rooting an unrooted tree and non-parametric bootstrap. Depending on time, I will also present the problems of tree simulation procedures for testing pipelines used in bioinformatics.

6. Dr Jan Poleszczuk - 2 x 1.5 h: 3.02.2020, 10.02.2020

Mathematical modeling of cancer development and treatment: Introduction to mathematical oncology

Modern oncology is facing many challenges, most of which are related to the constantly increasing number of approved drugs and the amount of available clinical data. Back in 2014 there were about 1000 new anti-cancer drugs under investigation in the United States alone. Moreover, new drugs are rarely used alone and every drug combination needs to be clinically tested before it can be used in practice. As an example, in 2016 in the United States alone there were about 500 clinical trials in which new immune checkpoint inhibitors were tested with other, already available therapies. Some of the data gathered during those trials (and during routine clinical practice) is stored in a large public repositories, maintenance and analysis of which requires a lot of resources (both human and computational). In "The Cancer Genome Atlas" alone there is almost 2.5 petabytes of data available to researchers. Finally, there ia a constant development of new diagnostic methods which are generating new type of not fully understood data.

Mathematical methods are invaluable tools that could be used to solve some problems of the modern oncology. This has been already recognized by the scientific and medical communities with the establishment of the eld named "mathematical oncology" [1] The idea behind it is to create models of two types: 1) those that can test the assumptions and protocols of clinical trials before they start, and 2) those that can help to predict patient's response to a given treatment. The former type of models can be compared to the virtual crash test that are nowadays performed by the car industry before building rst prototypes. The second type of the models can be compared to so-called "spaghetti models" that are nowadays used to predict path of hurricanes.

During the lectures I will present exemplary mathematical oncology models and show how can they be applied to specific problems. Many different types of the models will be presented, starting from relatively simple equations describing cellular response to radiation [2], through elaborated agent-based models describing interactions of multiple populations [3, 4], finishing with models based on ordinary or partial di erential equations [5, 6]. I will show di cul-ties associated with building models as well as problems with comparing them to the experimental and clinical data. Therefore, during the lecture I will show also some statistical and optimization methods. The lectures will finish with the presentation of some open problems of the mathematical oncology.

References

[1] Anderson ARA, Quaranta V. Integrative mathematical oncology. Nature Reviews Cancer. 2008;8(3):227-234.

[2] Sachs RK, Hahnfeld P, Brenner DJ. The link between low-LET dose-response relations and the underlying kinetics of damage production/repair/misrepair. International Journal of Radiation Biology. 1997;72(4):351-374.

[3] Kather JN, Poleszczuk J, Suarez-Carmona M, et al. In silico modeling of immunotherapy and stroma-targeting therapies in human colorectal cancer. Cancer Research. 2017;77(22):6442-6452.

[4] Kather JN, Charoentong P, Suarez-Carmona M, et al. High-throughput screening of combinatorial immunotherapies with patient-speci c in silico models of metastatic colorectal cancer. Cancer Research. 2018;78(17):5155-5163.

[5] Kuznetsov VA, Makalkin IA, Taylor MA, Perelson AS. Nonlinear dynamics of immunogenic tumors: Parameter

estimation and global bifurcation analysis. Bulletin of Mathematical Biology. 1994;56:295-321.
[6] Enderling H, Chaplain MAJ, Anderson ARA, Vaidya JS. A mathematical model of breast cancer development, local treatment and recurrence. Journal of Theoretical Biology. 2007 may;246(2):245-259