Syllabus

# 1. Course Description

## Title of a Course

Systems Biology and Personalized Medicine

## Pre-requisites

To master this course, students must possess the knowledge and competence of the following disciplines:

* Introduction to molecular biology
* Modern methods of data analysis

Given the wide range of backgrounds among students this course tries to avoid heavy usage of mathematics. The "officially supported" programming language is R, but problems can be done in any language (Python, etc).

## Course Type

This course is compulsory.

## Abstract

This course provides an introduction to systems biology with an emphasis on analysis of -omics data and applications in medicine. Accent is made on analysis of expression data, heavy usage of graphs and networks and other methods to enhance routine statistical analysis with biological knowledge. Special lecture is devoted to introduction to oncogenomics. This course is designed for students with strong backgrounds in either molecular biology or computer science, but not necessarily both. The open R+BioConductor proramming environment —which is widely employed for bioinformatics and computational biology — is used to illustrate the applications.

# 2. Learning Objectives

Systems biology is an inter-disciplinary field that makes attempt to understand the function of living systems by combining multivariate -omics data with structured biological prior knowledge (molecular interactions, gene functions, and pathway information, etc) into predictive models. A key modeling framework for systems biology is the network (including such concepts as pathway, and geneset) – a powerful approach to describe interactions between different molecular entities and incorporate biological knowledge. In order to develop these predictive models, systems biology makes heavy use of the tools of graph theory, machine learning and statistics. The goal of this course is to give students an overview of various modern methods of systems biology and analysis of omics data with an emphasis on application in medicine.

# 3. Learning Outcomes

At the end of this course, the student will be able to:

* Describe the main principles and the classes of problems that system biology can be applied to.
* Understand principles of modeling of biological processes at various levels
* Identify the main types of biological network models used in systems biology.
* Use publicly available resources that are relevant to a given biological application domain.
* Generate biologically meaningful hypotheses based on the analysis of high-throughput omics data.
* Conduct a high-level system-biological analysis of various types of omics data
* Explain how systems approaches can be used for purposes of personalized medicine.

# 4. Course Plan

**Lecture 1.** Definition and main problems of system biology. Properties of biological systems. Omics data and omics technologies, including proteomics and metabolomics. GWAS analysis. The main repositories of omics data. Common software tools system biology. Reproducibility problem.

**Lecture 2.** Modeling of biological and biochemical systems. Classification of models. Simulation based on differential equations: the law of effective masses, Michaelis-Menten kinetics. Compartmental models and pharmacokinetics. Parameter problem. Boolean networks, analysis of metabolic fluxes, Petri nets. Agent modeling approach. Description of models: SBML, BioPAX, SBGN. Software tools: CellDesigner, COPASI.

**Lecture 3.** System-biological analysis of transcriptome data. A simple overrepresentation test. GSEA method. Gene Ontology Database, MSigDB. Online resources for enrichment analysis (DAVID, WebGestalt, G: profiler). R packages: clusterProfiler, Enrichment-Browser. The peculiarities of Gene Ontology. Visualization of the enrichment results (Rev-iGO). GSEA modifications: GSVA, GAGE, STEPath algorithms. Adaptation for RNASeq data.

**Lecture 4.** Problems arising in the analysis of omics data. Principal Componene Analysis and Independent Component Analysis to overcome the “curse of dimensionality”. Missing data. Outliers. Batch effects identification and removal (ComBat algorithm). Meta-analysis and integration of omics data (PARADIGM, iCluster, MixOmics, MINT, YuGene). Analysis of prioritized gene lists (RankProd). Visualizing omics data.

**Lecture 5.** Biological networks and their properties. Typical network analysis problems. Signalling and metabolic pathways. Databases: KEGG, Reactome, PathwayCommons. Application of text-mining analysis to pathway creation. Software for pathway analysis. Cytoscape and its plugins (ClueGO, GeneMANIA, ReactomeFIViz).

**Lecture 6.** The use of networks for the analysis of omics data. Accounting for topology in the analysis of overrepresentation: SPIA, DEAP algorithms. Search for gene regulators: iRegulon, SNEA, master regulators. Subnetwork identification: jActiveModules, BioNet. Patient subtyping and identification of driver subnetworks. Prioritizing a list of genes. Identification of complexes in proteomics.

**Lecture 7.** Introduction to oncogenomics. Oncogenes and tumor suppressors. Cancer genome. TCGA resource. Identification of novel cancer-associated genes. Identification of molecular tumor subtypes.

Individual selection of optimal therapy. Clinically significant mutations. Mutation filtering and prioritization. The overall scheme of analysis of cancer genomes.

**Lecture 8.** Systems medicine. Predicting the risk of developing diseases according to omics data: RiskOGram, POGO. Metagenomic data and their use in medicine. Omics data and aging. Epigenetic clock. Microbiome analysis. iPOP.

# 5. Reading List

## Required

“Elements of Computational Systems Biology”, Stephen H. Muggleton, Huma M. Lodhi; John Wiley & Sons, 2010, ISBN: 9780470180938

“Multi-omic Data Integration“, Paolo Tieri, Christine Nardini, Jennifer Elizabeth Dent, Frontiers Media, 2015, ISBN: 978-2-88919-648-7

“Computational Modeling of Genetic and Biochemical Networks.” James M. Bower and Hamid Bolouri. The MIT Press, Cambridge, Massachusetts, London, England, 2001, ISBN: 978-0-262-02481-5

## Optional

Systems Biology: A Textbook, 2nd Edition, E. Klipp, W. Liebermeister, C. Wierling, A.Kowald, May 2016, Wiley-Blackwell, 504 pages ISBN: 978-3-527-33636-4

Computational Systems Biology of Cancer, E. Barillot, L. Calzone, P. Hupe, J.-P. Vert, A. Zinovyev, 2012, CRC Press, 461 pages, ISBN 9781439831441

Systems Biology in Practice. Concepts, Implementation and Application, E. Klipp, R. Herwig, A. Kowald, C. Wierling, H. Lehrach, 2005 Wiley-VCH Verlag GmbH & Co. KGaA, ISBN 9783527310784

Systems Biology in Biotech& Pharma. A Changing Paradigm. A. Prokop, S., Michelson, SpringerBriefs in Pharmaceutical Science & Drug Development, 2012, ISBN 9789400728493

# 6. Grading System

A 10-point grading scale is used in this course.

# 7. Guidelines for Knowledge Assessment

There will be several homework assignments usually with 2-week deadline. Homework assignments that are turned in late will be penalized by multiplying the grade by 0.66. The final grade is evaluated as average across all homeworks. Additional grades may be earned by making a presentation of recently published article on topic. The final exam will be held at the end of the course. The final grade is estimated as 0.7\*homework + 0.3\*exam.

# 8. Methods of Instruction

* Lectures: Introductory lectures are given to provide an overview of the main topics in the course.
* Computer Lab: Practical exercises and homework assignments throughout the whole semester where the students will attempt to solve various systems biology tasks/
* Compulsory Assignment: Students are to make a presentation covering the recently published article on systems biology.

# 9. Special Equipment and Software Support (if required)

Throughout the course only open-source and freely available software will be used:

R/Bioconductor, Cytoscape, Python.