Softwarewerkzeuge der Bioinformatik

Prof. Dr. Volkhard Helms — Saarland University
PD Dr. Michael Hutter, Markus Hollander, **Daria Gaidar** — Department of Computational Biology
Wintersemester 2019/2020

Projekt 3

as of 31. January 2020. Updated version

Pathway and Network Analysis of -Omics Data

UPD I offer each group a 1 hour consultation session to discuss for example, project and presentation drafts, what to present from the paper. Please send me an email with few days that suit you for me to pick from. Come prepared.

What will be graded:

- [60/100] A concise (scientific language, short but informative) well documented report filled with your notes, results (in text and figure forms) and their discussion. Here you should transmit an impression that you trully understand what is going on and how to read plots. Submit per email latest on 01.03.2020 at 01:30 a.m..
- [40/100] Group specific. Read a scientific paper assigned to you and conduct a short 25-30 min long presentation. Presetation session will be held on 21.02.2020 at 10:15-12:30.
- [-60/100] Plagiarism of any sort (e.g, figures, text) in your report.
- [-10/100] Missing/inconsistant formating (e.g, front page, group members, figure and page numbers, figure subscripts, citations, sub-sectioning, table of contents).

How to:

- Work in the group of 3. The default group assignment is as stated earlier. Cases of migration shall be self managed and well labeled with the names of the authors on submission files.
- Submit one report from the group.
- Submissions are accepted in electronic (daria.gaidar(at)bionformatik.uni-saarland.de) and/or printed versions.
- Language in the reports' text is german or english.
- Deadline is 01.03.2020 at 01:30 in the morning.

Practical work. Report. [60/100]

This year we will focus on the pathway and network analysis of omics data. Continuing the practice of bringing to you the recent material to learn from the best, we will follow the course material from Canadian Bioinformatics Workshop that took place in June 2016.

Please follow this link to find throughout documented lectures and tutorials.

http://bioinformatics-ca.github.io/pathway_and_network_analysis_of_omics_data_2016/
https://www.youtube.com/playlist?list=PL3izGL6oi0S9amh6CzalbUo4ICcvZIWFO

Your task: Process the material under Day 1 to Day 3 and optionally the pre-workshop tutorials (they wont be graded).

Module 1: Introduction to Pathway and Network Analysis

Lecture

Module 2: Finding Over-Represented Pathways in Gene Lists

Lecture
Lab practical: Enrichment-Based Analysis - Performing ORA

(a) [4/60] Exercise 1: Over-Representation Analysis (ORA) using GSEA.

(b) [4/60] Exercise 2: Over-Representation Analysis (ORA) using g:Profiler.

Module 3: Network Visualization and Analysis with Cytoscape

Lecture Part 1 and 2
Lab practical: Cytoscape Demo, Enrichment Map

(a) [8/100] Excercise 1. Create an EnrichmentMap from GSEA results and navigate through the network. Add drug target gene-sets to the network. Expand the network by adding an extra layer of information. Autoannotate. Create an automatically generated cluster labels to the network.

(b) [4/100] Excercise 2. Create an EnrichmentMap from g:Profiler results.

(c) [6/100] Excercise 3. Integrated assignment: g:Profiler/EnrichmentMap.

Module 4: More Depth on Pathway and Network Analysis

Lecture Part 1 and 2
Lab practical: De Novo Subnetwork Clustering Analysis in Reactome
This exercise will provide you with an opportunity to perform and network analysis using the Reactome Functional Interaction (FI) and the ReactomeFIViz app.

(a) [8/60] Exercise 1: Analyze somatic mutation data to identify biology that contributes to ovarian cancer.

Module 5: Gene Function Prediction

Lecture
Lab practical: GeneMANIA (web version)
Create GeneMANIA networks starting from a single gene to predict its function or starting from a gene list. Explore and understand the main output features of GeneMANIA such as the network composition or the enriched functions.

(a) [6/60] Exercise 1: Imagine that you are interested in exploring the function of the human GRN gene: GRN returned as the strongest hit from your omics experiment but not many information about this gene is available in functional databases. Use GeneMANIA to identify its predicted function as well as potential interaction partners.

(b) [6/160] Exercise 2: You are working with a list of 30 prostate cancer genes. During this exercise, you will explore the types of networks that have been used to create the GeneMANIA network from the prostate cancer gene list and you will see how changing input parameters can affect the results.

(c) [4/60] Exercise 3: Integrated assignment: g:Profiler/EnrichmentMap. All *optional* question *are obligatory* to do.

Module 6: Regulatory Network Analysis

Lecture

Lab practical: iRegulon

Import a Cytoscape network and apply iRegulon on all the selected nodes. Explore and understand the main output features of iRegulon such as the Transcription target view. Learn how to display predicted targets of a specific transcription factor by creating its metatargetome.

(a) [6/60] Exercise 1. Detect regulons from co-expressed genes.
In this exercise, we will continue to use the genes from the prostate cancer list from the GeneMANIA assignment. iRegulon requires a network from the start, and we will use the GeneMANIA network that we previously saved for this purpose. Using iRegulon, we will look for transcription factors (TFs) that may regulate a set of genes in this network.

(b) [4/60] Exercise 2. Create a metatargetome using iRegulon and merge 2 networks in Cytoscape.
This exercise will teach you to use the metatargetome function of iRegulon. This function displays a list of potential targets for a specific TF. We will create the metatargetome of two TFs, that we found as potential coregulators of the prostate cancer genes (Exercise 1): MTF1 and LARP4. We will then learn how to use Cytoscape to merge two networks and visualize nodes in common.

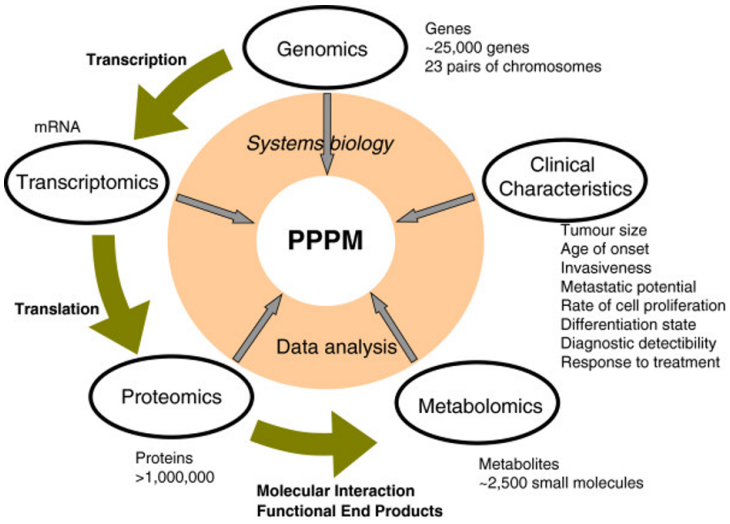

Abbildung 1: Multi-parameter systematic strategies for predictive, preventive and personalised medicine in cancer. From: https://doi.org/10.1186/1878-5085-4-2

Theoretical part. Working with scientific paper. Conducting presentation. [40/100]

In the following task you should get a well rounded overview on the capabilities of the Systems Biology and Bioinformatics. Papers assigned span from the application paper partially touched in the previous part to the review papers for the -omics mentioned below.

Genomics is a discipline in genetics concerned with the study of the genomes of organisms. The field includes efforts to determine the entire DNA sequence of organisms and fine-scale genetic mapping. The field also includes studies of intragenomic phenomena such as heterosis, epistasis, pleiotropy and other interactions between loci and alleles within the genome. In contrast, the investigation of the roles and functions of single genes is a primary focus of molecular biology or genetics and is a common topic of modern medical and biological research. Research of single genes does not fall into the definition of genomics unless the aim of this genetic, pathway, and functional information analysis is to elucidate its effect on, place in, and response to the entire genome's networks.

Proteomics is the large-scale study of proteins, particularly their structures and functions. The term p*roteomics* was first coined in 1997 to make an analogy with genomics, the study of the genes. The word proteomeïs a blend of *protein* and *genome*, and was coined by Marc Wilkins in 1994 while working on the concept as a PhD student. The proteome is the entire complement of proteins, including the modifications made to a particular set of proteins, produced by an organism or system. This will vary with time and distinct requirements, or stresses, that a cell or organism undergoes.

Finally, **metabolomics** is the scientific study of chemical processes involving metabolites. Specifically, metabolomics is the *systematic study of the unique chemical fingerprints that specific cellular processes leave behind*, the study of their small-molecule metabolite profiles. The metabolome represents the collection of all metabolites in a biological cell, tissue, organ or organism, which are the end products of cellular processes. Thus, while mRNA gene expression data and proteomic analyses do not tell the whole story of what might be happening in a cell, metabolic profiling can give an instantaneous snapshot of the physiology of that cell. One of the challenges of systems biology and functional genomics is to integrate proteomic, transcriptomic, and metabolomic information to give a more complete picture of living organisms. From: https://openwetware.org/wiki/GPM

Group-paper assignment

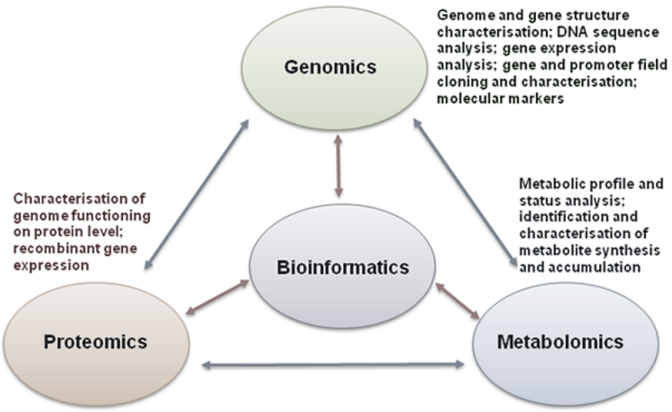

Abbildung 2: Grand schema. From: https://openwetware.org/wiki/GPM

(a) Group 1 - Verfaillie, Annelien, Hana Imrichova, Bram Van de Sande, Laura Standaert, Valerie Christiaens, Gert Hulselmans, Koen Herten et al. **iRegulon: from a gene list to a gene regulatory network using large motif and track collections.** PLoS computational biology 10, no. 7 (2014).

(b) Group 2 - Sanchez-Vega, Francisco, Marco Mina, Joshua Armenia, Walid K. Chatila, Augustin Luna, Konnor C. La, Sofia Dimitriadoy et al. **Oncogenic signaling pathways in the cancer genome atlas.** Cell 173, no. 2 (2018): 321-337.

(c) Group 3 - Keskin, Ozlem, Nurcan Tuncbag, and Attila Gursoy. **Predicting protein,Äìprotein interactions from the molecular to the proteome level.** Chemical reviews 116, no. 8 (2016): 4884-4909.

(d) Group 4 - Mitra, Koyel, Anne-Ruxandra Carvunis, Sanath Kumar Ramesh, and Trey Ideker. **Integrative approaches for finding modular structure in biological networks.** Nature Reviews Genetics 14, no. 10 (2013): 719-732.

(e) Group 5 - Bordbar, Aarash, Jonathan M. Monk, Zachary A. King, and Bernhard O. Palsson. **Constraint-based models predict metabolic and associated cellular functions.** Nature Reviews Genetics 15, no. 2 (2014): 107-120.

Have fun!