Course outline

Analysis of complex datasets requires the adoption of standardised data nomenclatures and formats and the application of sophisticated mathematical and statistical approaches to mine meaningful information from a sea of numbers. Models can be constructed from the data allowing the characteristics of the system to be systematically tested in a way that is too time consuming to achieve experimentally. Models are therefore an important source of new hypotheses. Thus, in order to gain an understanding of systems biology, it is necessary to have an idea of the underlying principles for automated analysis of large datasets and visualisation techniques. In addition, expertise has to be created for the integration of various datasets into biological models.

Wageningen University, together with BioWise, is developing a Systems biology course, which will deal with the principles underlying the handling and integration of large datasets derived from the different ~omics disciplines.
The course is in the first place meant for Dutch Masters-students but will be adjustable for HBO students and company people. The course will be divided into three modules, each focussing on an aspect of Systems Biology. Each module will contain background information, a number of simple, illustrative examples and a number of relevant exercises using (trimmed down versions) of real data sets.

In the systems biology course we will

- Teach how different methods can be used to validate and analyse complex biological expression data for the elucidation of specific gene functions and protein relationships.
- Show different approaches to integrate data sets from various sources in comprehensive virtual biological networks.
- Show how these integrative networks can be used to extract pathways and yield statistics.

To reach this goal, the course will be divided into several modules each focusing on an aspect of systems biology.

1. **Data generation**: a brief introduction in the various –omics techniques that are used to produce the transcriptome, proteome and metabolome data. The focus will be on dealing with sources of technical and biological variation in data generation (pre-processing, normalisation, experimental design).
2. **Data analysis**: the principles of different statistical approaches in data analysis will be explained and exercised (t-test, ANOVA, PCA, clustering). For the interpretation of large data sets, these statistical analysis methods are very often linked to visualisation techniques. Therefore, **data visualisation** techniques will be integrated in the data analysis module
3. **Data integration**: different approaches in the development of virtual metabolic pathways or regulatory or genetic networks will be discussed and exercised (KEGG, Boolean networks).

Each module will contain background information, a number of simple, illustrative examples and a number of relevant exercises using (trimmed down versions) of real data sets. Rather then getting lost on technical details of specific programs the modules will focus on the underlying principles used. The course is concluded by a final case study in which all the different aspects of the course can be applied. The candidate discusses this case study in a 15-min. presentation with 5-min discussion time.

For further information or ideas please contact Dr. I. Rienties or Dr. P. Schaap