

*Prof. Dr. Knut Reinert,
Julianus Pfeuffer,
Xiao Liang*

October 22, 2014

Computational proteomics and metabolomics

WS 2014/15

Exercises 0

1. Omics and systems biology

One challenge in systems biology is the integration of multiple omics layers. Please explain how the individual omics layers genomics, transcriptomics, proteomics, and metabolomics are related to each other.

2. Connect the terms

Connect the processes/compounds of the second group with the Omics terms in the first group.

a Lipidomics

b Proteomics

c Metabolomics

d Epigenomics

1 Phosphorylation

2 Gluconeogenesis

3 Methylation site elucidation

4 Arachidonic acid

3. Protein isoforms

Use the protein database UniProt to investigate how many different splice isoforms are reported for the human cyclin-dependent kinase 10 (CDK10).

4. PCA in KNIME

In this assignment we construct a basic KNIME workflow. Note: Understanding all mathematical details of the PCA is not required as our focus here lies on workflow construction. Principal Component Analysis is a frequently used method to reveal underlying structure in data. It can be used to reduce the dimensionality of high dimensional data set e.g. for easier visualization. Without going into the mathematical details, PCA determines orthogonal vectors (=principal components) pointing into the directions the data spreads. The first principal component points into the direction of the main spread of the data and retains most variance if data is projected on it.

Please acquire the famous iris data set from <https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data> (information on it: http://en.wikipedia.org/wiki/Iris_flower_data_set).

1. Construct a KNIME workflow that performs a principal component analysis on this data set.
2. Visualize the original data and the result after reduction to the two biggest principal components in KNIME .

Present your workflow in the exercises.

Hints: You probably might want to take a look at following nodes: CSV Reader, PCA Compute/Apply, Color Manager and Scatter Plot

5. Correlation and regression in KNIME

As you made a PCA with the Iris dataset, of course you asked yourself, to which extent petal length and sepal length correlate.

So you can easily reuse the read in of the dataset and calculate Pearson's correlation coefficient.

You can also create a linear regression(learner) of petal length on sepal length and give the R^2 of that regression here.