Lecture 5

L5.1. PCA

a. Work with ***mice*** data from http://edu.sablab.net/data/txt/mice.txt. Perform PCA and identify outliers (or strangely behaving creatures) among mice population.

   Hint: Replace NA values in each column by the corresponding median value (for this column).

b. Acute lymphoblastic leukemia (ALL), is a form of leukemia characterized by excess lymphoblasts. ***ALL*** contains the results of transcriptome profiling for ALL patients and healthy donors using Affymetrix microarrays. The expression values in the table are in $\log_2$ scale. See data at http://edu.sablab.net/data/txt/all_data.txt

   - Perform exploratory analysis of ALL dataset using PCA

L5.2. Clustering

a. Work with ***ALL*** dataset.

   - Using t-test, identify top 100 genes with significantly different expression b/w ALL and normal condition.
   - Build a hierarchical clustering for these genes

L5.3. Classification

a. Data ***"infarction"*** was recently by the group of Y. Devaux (LIH). The diagnostic performance of high-sensitivity cardiac troponin T (hs-cTnT) and cardiac enriched microRNAs (miRNAs) in patients with myocardial infarction were compared. Data contain measurements for 3 groups: control, STEMI (acute ST segment elevation myocardial infarction) and non-STEMI. Classification b/w STEMI and NSTEMI is extremely important for selection of the therapy.

   Devaux Y., et al. Use of circulating microRNAs to diagnose acute myocardial infarction. Clin Chem. 2012

   - Compare individual prediction power for each of the markers
   - Try to classify patients using these markers
   - Think about the strategy to enhance the predictive power of the markers