

Bioinformatics for Genomics 2022 - Clustering

alex.button
alexanderluke.button@unil.ch

April 2022

1 PCA

Extract the GEO data (input/GEO data) and perform principal component analysis with three principal components.

2 PCA 2d plots

Generate a series of two dimensional plots by plotting each principal component against each other. Which two components best separate the data? How did the data cluster? How well does the PCA represent the sample labels?

3 PCA 3d plot

Generate a three dimensional plot of the three principal components. How does this compare to the two dimensional plots? How well separated is the three-dimensional data?

4 Hierarchical clustering

Perform hierarchical clustering on the gene expression data. How well did this cluster the data? Were the samples separated by sample label?

5 K-means

Perform k-means clustering with 2,3, and 6 clusters on the PCA transformed data. How well do the clusters represent the sample labels? How does the number of clusters influence the clustering?