

Block 3: Clustering

Apply hierarchical clustering, and apply K-means clustering to the dataset containing skin tissue (normal and tumor data). Perform clustering separately for the genes (based on their expression in different samples) and for the samples (based on the expression levels of the genes).

Address the following issues related to clustering the genes:

1. When clustering genes, test using different definitions of distance. What do you observe?
2. Visualize the gene expression levels in the different samples from genes in at least two clusters. Use a visualization that enables to compare the similarity between expression levels of genes that are clustered together. Comment on what you observe.

Address the following issues related to clustering the samples:

3. When clustering samples, compare clustering results obtained with different approaches with the known labels for these samples. What do you learn from this?
4. Can you say something on how many clusters there are in this dataset?