Lu.Getz.Miska.Nature.June.2005.clustering.miGCM218

Module name: Lu.Getz.Miska.Nature.June.2005.clustering.miGCM218
Description: Hierarchical clustering of 218 samples from various tissue types
Author: Gad Getz (Broad Institute), gp-help@broad.mit.edu

Summary
Hierarchical clustering of 218 samples from various tissue types partitions samples according to their developmental origins.

Hierarchical clustering of 218 samples from various different tissue types

We clustered a collection of 218 samples, which include 140 human cancer specimens, 46 normal human tissues, and various cell lines. The collection of samples covers more than ten tissues and cancer types. This collection was referred to as miGCM (for miRNA Global Cancer Map). We performed hierarchical clustering on this data set.

Hierarchical clustering was performed with average linkage, unweighted average distance (UPGMA), using (1-Pearson correlation) as the distance measure. Before clustering, data were filtered to eliminate genes with expression lower than 7.25 (on a log2 scale) in all samples. Next, all features (miRNAs) were centered and normalized to a mean of 0 and a standard deviation of 1. Clustering was performed on both the samples and genes axes.

The main results from this clustering are that the resulting dendrogram parallels the developmental origins of the tissues. For example, samples of epithelial origin are located on a single branch of the dendrogram, whereas the other major branch is predominantly populated with samples of haematopoietic malignancies. In addition, among the epithelial samples, those of the gastrointestinal tract are of particular interest. Samples from colon, liver, pancreas and stomach all cluster together, reflecting their common derivation from tissues of embryonic endoderm.

Note that using (1-correlation) to measure distances among the samples does not reflect overall change in miRNA abundance and, therefore, does not strongly capture tumor/normal differences.

References:

- Lu, Getz, Miska, et al. “MicroRNA Expression Profiles Classify Human Cancers,” Nature 435, 834-838 (9 June 2005)
- Jain, A. K. & Dubes, R. C. Algorithms for clustering data. Prentice-Hall Inc., Upper Saddle River (1988).