HierarchicalClustering Documentation

Module name: HierarchicalClustering
Description: Hierarchical clustering of rows or columns
Author: Gad Getz (Broad Institute), gp-help@broad.mit.edu

Summary: Given a set of items to be clustered (items can be either genes or chips/experiments), agglomerative hierarchical clustering (HC) recursively merges items with other items, or with the result of previous merges, according to their pair-wise "distance" (with the closest item pairs being merged first). As a result, it produces a tree structure, referred to as dendogram, whose nodes correspond to: i) the original items (these are the leaves of the tree); and ii) the merging of other nodes (these are the internal nodes of the tree). If k clusters are required (k 1), the merging proceeds until k nodes are left. HierarchicalClustering will produce a gct file which contains the original data, but reordered to reflect the clustering. In addition to the gct file, an odf file that desribes the dendrogram is produced.

Parameters:

Name	Description
input.filename	input filename - .gct
normalize.type	normalization applied before clustering
merge.type	merge type
distance.measure	distance measure to use
cluster.by	cluster by rows or columns
reordered.dataset.file	reordered gct output file name - .gct
dendrogram.file	dendrogram output file name - .odf

Return Value:
     1. reordered gct
     2. odf dendrogram file

Platform dependencies:

Task type:	Clustering
CPU type:	x86
OS:	Windows
Language:	MATLAB