

# GenePattern

## HierarchicalClustering Documentation

**Module name:** HierarchicalClustering  
**Description:** Agglomerative hierarchical clustering of genes/experiments  
**Author:** Stefano Monti (Broad Institute), [gp-help@broad.mit.edu](mailto:gp-help@broad.mit.edu)  
**Date:** 4/16/2003  
**Release:** 1.0

**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 dendrogram, 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 ( $k \geq 1$ ), the merging proceeds until k nodes are left. The ) the Euclidean distance measure is used.

HierarchicalClustering will produce a cdt file which contains the original data, but reordered to reflect the clustering. In addition to the cdt file, either an atr file or a gtr file is created. An atr file is produced if you clustered by samples (columns) and a gtr file is produced if you clustered by genes (rows), is created. For a more detailed description of the format of the output files see <http://genome-www5.stanford.edu/help/formats.shtml>.

### References:

- M.B. Eisen, et al. "Cluster Analysis and Display of Genome-Wide Expression Patterns," PNAS, 14863-14868 (1998).

### Parameters:

Name	Description	Choices
input.filename	RES, GCT, or ODF file	
output.stub	output stub	
merge.type	merge type=average, single, complete	average;single;complete
normalize.type	row-wise, col-wise, both	row-wise;col-wise;both
num.iter	row/column normalize data 'num iter' times (overrides normalize_type) Note: num iter, if used, should be about 10	
cluster.by	cluster by rows or columns	rows;columns

### Return Value:

1. <output.name>.cdt see description in Summary.
2. <output.name>.atr if clustered by columns or <output.name>.gtr if clustered by rows. See description in Summary.

# GenePattern

## Platform dependencies:

<b>Task type:</b>	Clustering
<b>CPU type:</b>	any
<b>OS:</b>	any
<b>Java JVM level:</b>	1.4
<b>Language:</b>	Java