

Instructions for Using Cluster

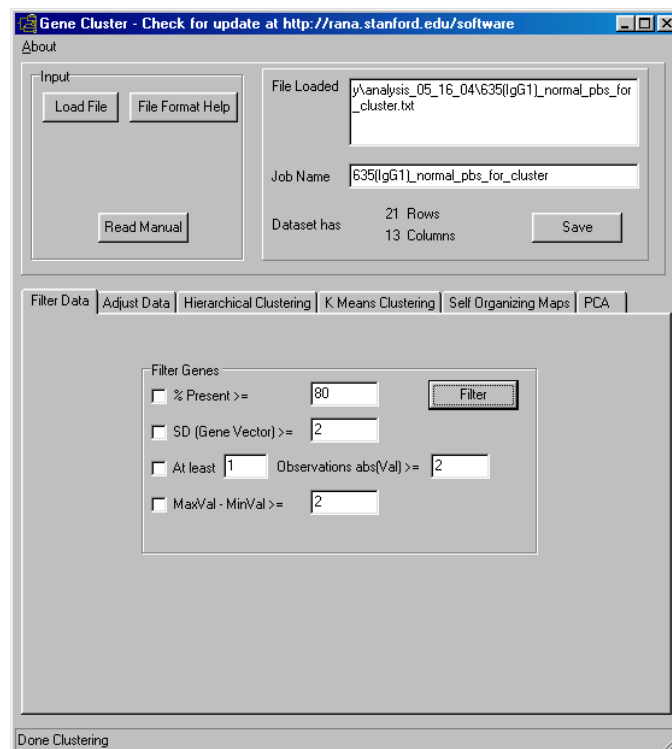
Introduction

Michael Eisen developed the Cluster program when he was at Stanford University. The purpose of this program is to “Perform a variety of types of cluster analysis and other types of processing on large microarray datasets.” Cluster is used in conjunction with TreeView for analyzing and visualizing microarray data respectively. The Cluster algorithm is available for Windows, MAC, Linux, and Unix. For more information, check out:

- <http://rana.lbl.gov/EisenSoftware.htm> (Eisen’s homepage includes the manual, software, and links to other sites)
- *ClusterTreeView.pdf* (Manual for Cluster and TreeView)
- *Eisen_PNAS_98_Cluster_Analysis.pdf* (Paper that describes the cluster algorithm, PNAS Vol. 95, pp. 14863–14868, Dec, 98)

Creating a Node-Map

Launch the *Cluster* (version 2.12*) program



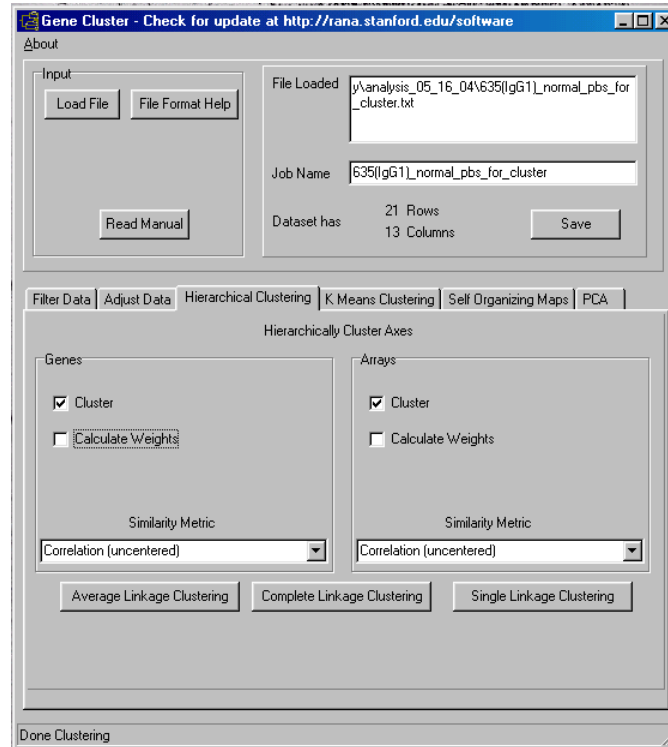
* The version is very important because the cluster nodes do not appear on other versions.

Select *Load File* and pick the *.txt file one wants to cluster

Select the *Hierarchical Clustering* tab

Check the *Cluster* "check boxes" under the headings Genes and Arrays (two check boxes in total)

Select the *Complete Linkage Clustering* button (when the clustering finishes, the lower left corner of the dialog window will say done clustering)



Complete Linkage Clustering creates a *.cdt file in the same folder as the *.txt folder. Note that the output of a node-map depends on the Cluster algorithm and will not be the same format, in terms of the slide name/number location and the antigen order, as the text file. In addition, the node-map will have a hierarchical set of nodes like an evolutionary tree that group samples or antigens based on how similar they are. We're done with making a node-map. Congratulations!

Creating a Heatmap (Map without Nodes)

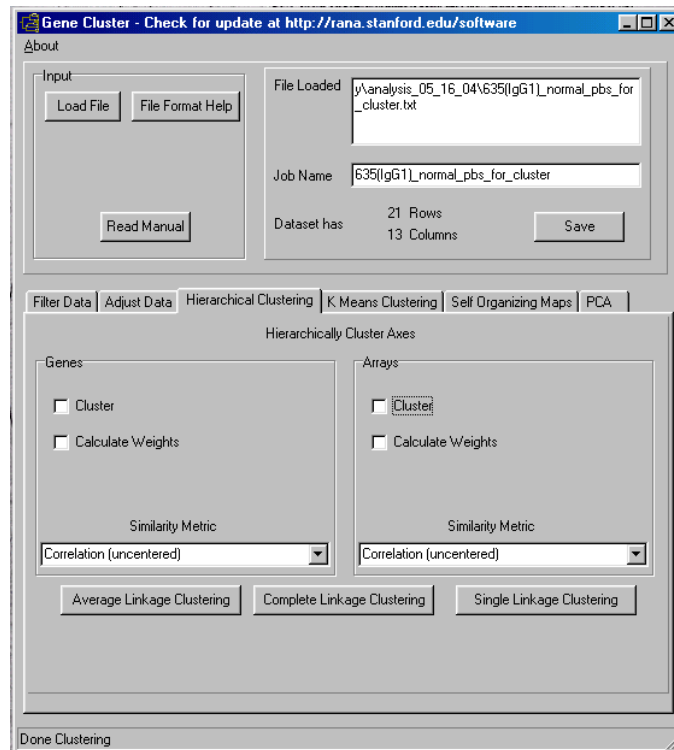
Launch the *Cluster* (version 2.12) program as before

Select *Load File* and pick the *.txt file one wants to cluster

Select the *Hierarchical Clustering* tab

Uncheck the *Cluster* "check boxes" under the headings Genes and Arrays (two check boxes in total)

Select the *Complete Linkage Clustering* button (when the clustering finishes, the lower left corner of the dialog window will say done clustering)



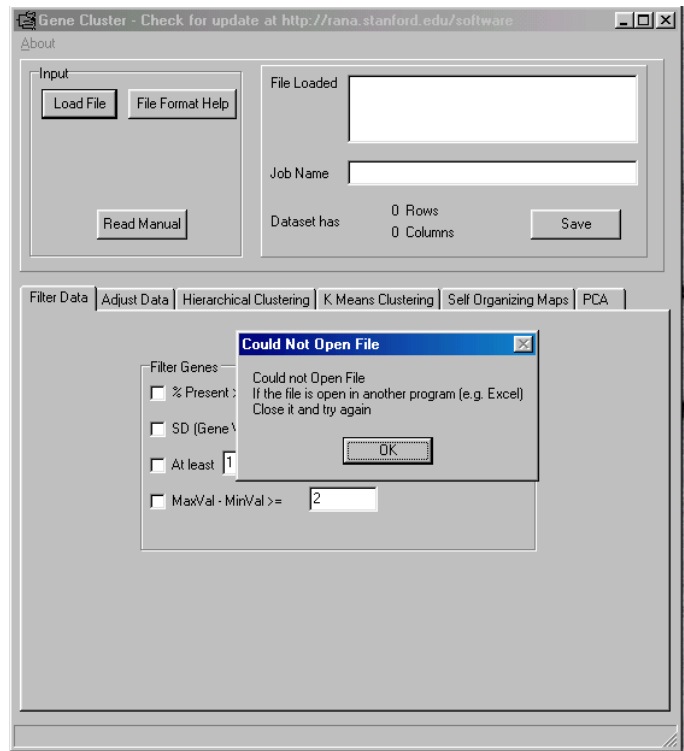
Complete Linkage Clustering creates a *.cdt file in the same folder as the *.txt folder. Note that the output of a heatmap is exactly the same format, in terms of the slide name/number location and the antigen order, as the text file. This feature distinguishes the heatmap from the node-map. In addition, the heatmap will not have any nodes in TreeView. We're done with making a heatmap. Congratulations!

Errors in Cluster

Cluster is a free program so it has some interesting quirks and errors that crop up for unknown reasons. When everything works well, it's like heaven and the analyses flow like water. When errors come up, however, it can be frustrating and difficult to get things back on track. Here is some advice to aid in those frustrating times.

A common error is for cluster to complain that it *Could Not Open File*. When this happens there might be a simple or a not so simple solution.

1. Simple solution – the simple solution is the text file is still open in Excel. In this case, one will have to close cluster, close excel and then start again.
2. Not so simple solution – everything is closed and the file format is correct. This is the strange case of who the hell knows what's going on. In my experience, the only way to get out of this jam is to copy the file format to a *new* excel file and save the file as a completely different name. One may have to do this a few times; it's very frustrating and bizarre.



Another common error is for cluster to complain that it *Could Not Open File* and an *Access Violation... Warning to appear*. When this happens it's likely that the program is still open in excel.

Finally, there are formatting errors that originate in the Cluster file format but only manifest in TreeView. These errors will appear as big gray bars in data map or with misspelled names in the antigen or sample list. To correct these errors, go back to the *for_cluster.xls* file and make sure everything is spelled correctly and the format is right.