Gene Pilot v1.07b

Row Count:0 Col Count:0 Chip Type:Other

GenePilot V1.07b August 28, 2003

Table of Contents

Chapter 1: Getting Started 1

Installation 2

Windows Instructions: 2

Mac OS Classic (8.1 or greater) Instructions: 2

Mac OS X Instructions: 2

Solaris Instructions: 3

Linux Instructions: 3

HP-UX Instructions: 3

Generic Unix Instructions: 3

All Other Platforms Instructions: 4

Tutorials 5

Creating Dataset Templates 6

Preparing Data 8

Chapter 2: Main Window 12

Menu Choices 13

Data Adjustment Settings 15

Filtering Options 15

Data Adjustment Options 15

Missing Data Imputation Options 16

Dataset Info 16

Interface - Row Info Import 17

Interface - Replicate Column Combine Panel 18

Chapter 3: Dataset Display 21

Parts - Heatmap Panel 22

Dataset Display- Menu Choices 23

Preferences - Row Info Display 26

Preferences - Gene Ontology 27

Interface - Copy to Clipboard 29

Interface - Copy to File 29

Interface - Search 30

Common Windows - Row Information Window 31

Common Windows - Palette 31

Common Page - Row Information Page 32
Common Interfaces - Url Lookup Editor 33

Chapter 4: Hierarchical Clustering 35

Run Settings 36
Results - Parts - Left Panel 37
Results - Parts - Right Panel 39
Results - Menu Choices 41
Interface - Create Dataset 43
Preferences - Row Info Display 44
Preferences - Gene Ontology 45
Interface - Copy to Clipboard 47
Interface - Copy to File 47
Interface - Search 48

Chapter 5: K-Means Clustering 49

Run Settings 50
Results - Parts – Graph and Button Panel 50
Results - Parts – Heatmap Panel 52
Results - Menu Choices 53
Windows - Create Dataset 56
Preferences - Row Info Display 57
Preferences - Gene Ontology 58
Interface - Copy to Clipboard 60
Interface - Copy to File 60
Interface - Search 61

Chapter 6: SOM 62

Run Settings 63
Results - Parts – Node Panel 64
Results - Parts – Heatmap Panel 65
Results - Menu Choices 67
Windows - Create Dataset 70
Preferences - Row Info Display 71
Preferences - Gene Ontology 72
Interface - Copy to Clipboard 74
Interface - Copy to File 74

Interface - Search 75

Chapter 7: SAM 76

Run Settings 77

Results - Parts – FDR Graph 78

Results - Parts – Score Graph 78

Results - Parts – Heatmap Panel 79

Results - Menu Choices 81

Windows - Create Dataset 84

Preferences - Row Info Display 85

Preferences - Gene Ontology 86

Interface - Copy to Clipboard 88

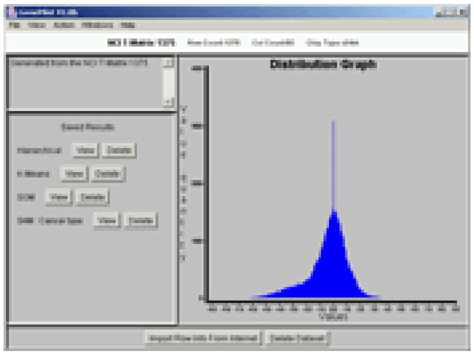
Interface - Copy to File 88

Interface - Search 89

References 90

Index 91

Chapter 1



Getting Started

Description

GenePilot™ is a stand-alone application designed to provide an intuitive and comprehensive interface for analyzing MicroArray Data. It combines the most popular and widely used tools with features that greatly enhance the ability of the user to datamine their information. It is simple enough to use for any user to feel comfortable in using most of it's features within a half hour. This contrasts with most of the popular tools that are currently available, which are quite confusing for even veterans at MicroArray Analysis.

To get started, it is recommended that the user start with one of the sample Datasets. The first is a sample of a cDNA Dataset, from the NCI. It is based on the NCI T-Matrix Dataset, and has very interesting results. It can be found in the 'Sample Data' folder, and it's name is 'NCI_T_Matrix.txt'.

Installation

Windows Instructions:

Instructions

After downloading, double-click **GenePilot_v1b.exe**

Notes

If you do not have a Java virtual machine installed, be sure to download the package above which includes one

Mac OS Classic (8.1 or greater) Instructions:

Instructions

After downloading, double-click **GenePilot_v1b.exe**

Notes

Requires PowerPC and Mac OS 8.1 or later

You may need to install Mac OS Runtime for Java (MRJ) 2.2 or later before using this package.

The installer is MacBinary encoded and should be automatically decoded after downloading. If it is not automatically decoded, you can decode it using StuffIt Expander 4.5 or later

Mac OS X Instructions:

Instructions

After downloading, double-click **GenePilot_v1b.exe**

Notes

Requires Mac OS X 10.0 or later

The compressed installer should be recognized by Stuffit Expander and should automatically be expanded after downloading. If it is not expanded, you can expand it manually using StuffIt Expander 6.0 or later.

If you have any problems launching the installer once it has been expanded, make sure that the compressed installer was expanded using Stuffit Expander. If you continue to have problems, please contact technical support

Solaris Instructions:

Instructions

After downloading open a shell and, **cd** to the directory where you downloaded the installer.
At the prompt type: **sh ./GenePilot_v1b.bin**

Notes

You need to install a Java 1.1.8 (or later) virtual machine. You can download one from Sun's Java web site or contact your OS manufacturer

Linux Instructions:

Instructions

After downloading open a shell and, **cd** to the directory where you downloaded the installer.
At the prompt type: **sh ./GenePilot_v1b.bin**

Notes

You need to install a Java 1.1.8 (or later) virtual machine. You can download one from Sun's Java web site or contact your OS manufacturer

HP-UX Instructions:

Instructions

After downloading open a shell and, **cd** to the directory where you downloaded the installer.
At the prompt type: **sh ./GenePilot_v1b.bin**

Notes

You need to install a Java 1.1.8 (or later) virtual machine. You can download one from Sun's Java web site or contact your OS manufacturer

Generic Unix Instructions:

Instructions

After downloading open a shell and, **cd** to the directory where you downloaded the installer.
At the prompt type: **sh ./GenePilot_v1b.bin**

Notes

You need to install a Java 1.1.8 (or later) virtual machine. You can download one from Sun's Java web site or contact your OS manufacturer

All Other Platforms Instructions:

Instruction(Unix or Unix-like operating systems)

For Java 2, after downloading, type

java -jar GenePilot_v1b.jar

If that does not work, try

java -classpath [path to]**classes.zip:GenePilot_v1b.jar install**

If that does not work either, on sh-like shells, try

cd [to directory where GenePilot_v1b.jar is located]
CLASSPATH=GenePilot_v1b.jar
export CLASSPATH
java install

Or for csh-like shells, try

cd [to directory where GenePilot_v1b.jar is located]
setenv CLASSPATH GenePilot_v1b.jar
java install

Instructions (for other platforms)

Be sure you have Java 1.1.8 or later installed. You can download Java from Sun's site

In a console window, change to the directory where you downloaded **GenePilot_v1b.jar** to before running the installer

Your operating system may invoke Java in a different way. To start the installer, add **GenePilot_v1b.jar** to your **CLASSPATH**, then start the main class of the installer named **install**

Tutorials

Sample Data

We highly recommend that you go through the tutorials, especially the Sample Data tutorial. Within this tutorial you will be walked through the whole process of bringing a Dataset into GenePilot and stepping through most of the features that GenePilot has to offer. You can access this tutorial with Help->Tutorials->Sample Data, once you have launched GenePilot. We estimate that the average user can get through the comprehensive tutorial in about two hours unless they get side-tracked by the very interesting (and actual) data that is used in the Tutorial. This sample data is the NCI T-Matrix1375 data which contains gene expression results from various kinds of cancer.

Your Data

When you have completed the Sample Data tutorial (or choose to skip the tutorial) and are ready to import your own data, we highly recommend that you use this tutorial to walk you through the process of importing your own data for the very first time.

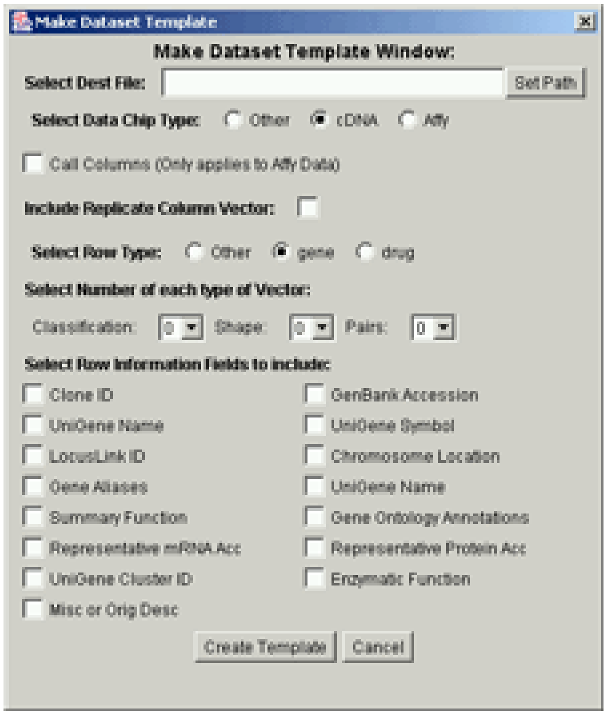
Creating Dataset Templates

Description

The very first step which needs to be taken in order to analyze data in GenePilot is to export a template which is specific to the Dataset. By exporting a template, it can be customized to exactly fit the needs of the dataset, avoiding the need to add or delete rows or columns. Once the template is created, the information can then be filled in (See Preparing Data) and loaded into GenePilot. The Make Dataset Template window can be opened by accessing the menu item; File->Dataset->Create Template.

Select Dest File

This field sets the name and location of the template file. Click on the 'Set Path' button to select the location and file name.



Make Dataset Template Window:

Select Dest File:

Select Data Chip Type: ○ Other ● cDNA ○ Affy

□ Call Columns (Only applies to Affy Data)

Include Replicate Column Vector: □

Select Row Type: ○ Other ● gene ○ drug

Select Number of each type of Vector:

Classification: 0 Shape: 0 Pairs: 0

Select Row Information Fields to include:

□ Clone ID
□ UniGene Name
□ LocusLink ID
□ Gene Aliases
□ Summary Function
□ Representative mRNA Acc
□ UniGene Cluster ID
□ Misc or Orig Desc
□ GenBank Accession
□ UniGene Symbol
□ Chromosome Location
□ UniGene Name
□ Gene Ontology Annotations
□ Representative Protein Acc
□ Enzymatic Function

Make Dataset Template

Select Data Chip Type

This selection sets the type of chip used to generate the data. The following types are available:

Other - This is for Chip Types that aren't listed.
cDNA - This is for data that is produced using cDNA type chips.
Affy - This is for data that is produced using Affymetrics chips.

There is one additional field in this selection, in the form of a checkbox, Call Columns. This only applies to Affy data and should be checked if there is columns of Call Data included after each Data Column.

Include Replicate Column Vector

This selection adds in a Vector for specifying replicate columns for each experiement. This Vector can then be used by GenePilot to do a sophisticated combine of the replicate columns. This information can also be used in sorting the columns in the result displays for all of the analytic programs.

Select Row Type

This selection sets the type of data that is represented in each row. The following types are available:

Other - This is for types of data which is different from the listed types.
gene - This is for when the rows are genes.
drug - This is for when the rows are drugs.

Select Number of each type of Vector:

This selection sets the quantity of each type of vector that will be included with the Dataset. It is recommended that each Dataset have at least one Classification vector included in order to fully utilize the power of GenePilot. The following types of Vectors are as follows:

Classification - This vector defines the Class (i.e. Breast, Lung for Cancers) of each column of data.
Shape - This vector type defines a shape or vector (i.e. expression values of a row).
Pairs - This vector type matches column pairs (1, -1, 2, -2, etc) good for before/after data.

Select Row Information Fields to include

This selection sets the type of Row Information that will be included in the Dataset. Note that GenePilot provides an interface for importing data from an outside source by using some of the ID's (Fields with <Query Field> can be used for importing additional information). This allows for importing minimal information then populating the dataset with a significant amount of information through the Data import process (fields with <Import Item> are fields that can be imported). GenePilot supports the following information:

Clone ID - <Query Field><Import Item> Commonly used in NCI Data (CLID)
GenBank Accession - <Query Field><Import Item>
UniGene Name - <Query Field><Import Item>
UniGene Symbol - <Query Field><Import Item>
LocusLink ID - <Query Field><Import Item>
Chromosome Location - <Import Item>
Gene Aliases - <Import Item>
UniGene Name - <Import Item>
Summary Function - <Import Item>
Gene Ontology Annotations - <Import Item>
Representative mRNA Acc - <Import Item>
Representative Protein Acc - <Import Item>
UniGene Cluster ID - <Query Field><Import Item>
Enzymatic Function - <Import Item>
Misc or Orig Desc - Suggested field for description information included with Original data.

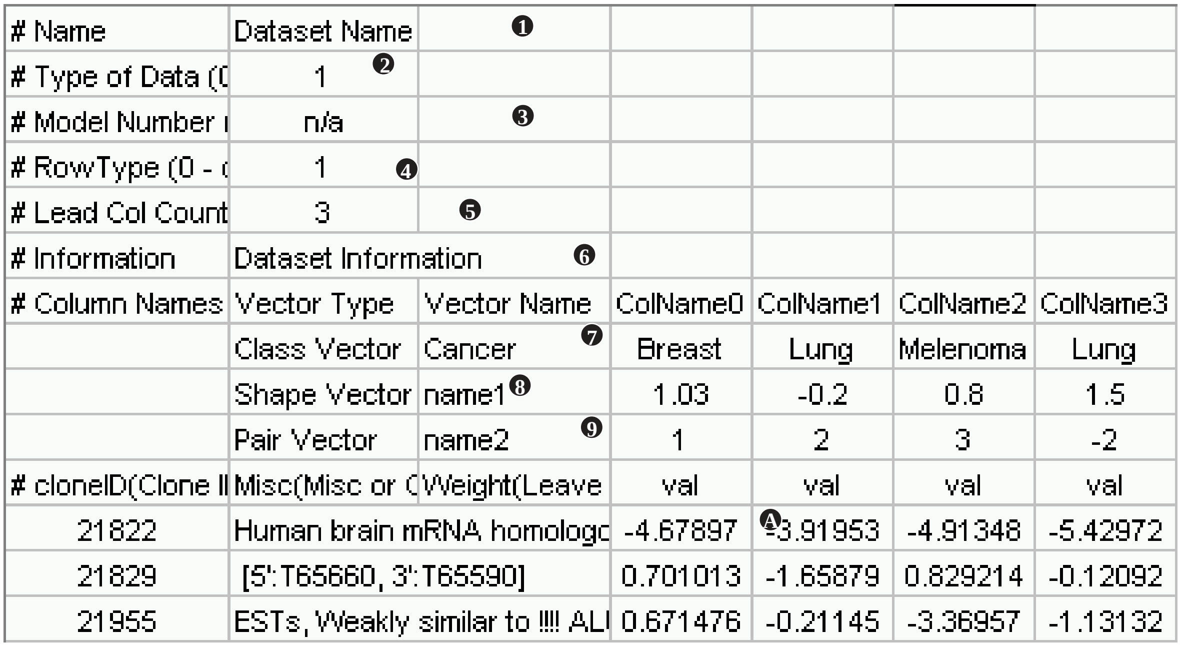
Button - Create Template

This button creates the template.

Button - Cancel

This button closes the window without creating the template.

Preparing Data



# Name	Dataset Name	❶				
# Type of Data ((	1 ❷					
# Model Number r	n/a	❸				
# RowType (0 - c	1 ❹					
# Lead Col Count	3	❺				
# Information	Dataset Information	❻				
# Column Names	Vector Type	Vector Name	ColName0	ColName1	ColName2	ColName3
	Class Vector	Cancer ❼	Breast	Lung	Melenoma	Lung
	Shape Vector	name1❽	1.03	-0.2	0.8	1.5
	Pair Vector	name2 ❾	1	2	3	-2
# cloneID(Clone I	Misc(Misc or (	Weight(Leave	val	val	val	val
21822	Human brain mRNA homologc		-4.67897	Ⓐ-3.91953	-4.91348	-5.42972
21829	[5':T65660, 3':T65590]		0.701013	-1.65879	0.829214	-0.12092
21955	ESTs, Weakly similar to !!!! ALI		0.671476	-0.21145	-3.36957	-1.13132

Input Data: 1-Dataset Name, 2-Type of Data, 3-Model Number, 4-Lead Columns before Data, 5-Dataset Information, 6-Column Name Row, 7-Classifidcation Vector, 8-Shape Vector, 9-Pair Vector, A-Data Row

Description

To simplify the process of data importation, GenePilot uses a template format. This structured template uses keywords and tab-delimiting for creating an intuitive template to provide GenePilot with the information necessary for the best results while analyzing your data. In the description of the fields below, required information is marked with **<required>**. Each required row must be in the order as laid out in the template. It is recommended that you use a spreadsheet to create this data, then save it as 'tab-delimited text'.

A very important part of the input data is the Vectors Section. In this section, you can define Classification Vectors, Shape Vectors and Pair Vectors. These vectors can be used as supervising vectors in the Supervised Analysis (i.e. SAM). The Classification vectors are used by all the tools to indicate classes (i.e. Breast Cancer, Lung Cancer, Melanoma, etc.), which is extremely useful when viewing the results.

Input - Name <Required>

This entry will be the Name that this Dataset is referred to, throughout GenePilot. It must be located in the second column of the fist row, and the first column must have '# Dataset Name'.

Input - Type of Data (0 - Other, 1 - cDNA, 2 - Affy) <Required>

This entry tells the program what type of chip was used to produce the data. It can be one of three possibilities:
0 - Other Data - This applies to all data that isn't defined
1 - cDNA Data - This applies to data created with cDNA-style chips.
2 - Affy Data - This applies to data created with Affymetrics chips.
The number that defines the chiptype for the current Dataset must be located in the second column, the first column must have, at the start '# Type of Data'.

Input - Model Number <Optional>
This entry tells the program the model number of the chip, if applicable. In future releases, this information can be used to import additional data about the rows of information, such as accession numbers and gene names. The model number must be located in the second column, with the first column starting with '# Model Number'.

Input - Lead Col Count <Required - Do Not Edit!>
This entry provides information to the import code to indicate where the Data Columns start. Do Not Edit!

Input - Information <Optional - Recommended>
This entry provides information about the dataset, that is displayed with the Dataset Information. This information is especially useful, when several versions of the same Dataset are being used, and specific entries are made to explain the differences. The Information must be located in the second column, with the first column starting with '# Information'.

Input - Column Names <Required>
Column names define the names for each column that are displayed in the result windows. It is very important to choose names that will make sense when viewing the results. Each name should be placed above the corresponding column of data, starting in the 5th column. The first column must start with '# Column Names'.

Input - Vectors - Replicates Vector <Optional>
This entry ties the replicate columns together which represent the same experiment. Starting with 0 (zero), assign the same number to each replicate for an experiment. If there are three columns of data for each experiment then there should be three 0's, three 1's, three 2's, etc. The following describes the entries needed:
Column 1 - <leave blank>
Column 2 - Vector - Type - Must begin with 'Replicates'
Column 3 - Vector Name - Must be 'Replicates'.
First Data Column to last - Number representing which column group this column is part of.

Input - Vectors - Classification Vector <Optional - Recommended>
This entry defines the class which each column is associated with. With a Dataset with Cancer data, this could be Breast, Lung, Melanoma, etc. This class information can be selected and displayed under each column name for every Analytical result, which can prove extremely useful in understanding the results. You can define and include as many different classification vectors as you wish. All Vectors must be in the rows immediately following the Column Names row. The following describes the entries needed:
Column 1 - <leave blank>
Column 2 - Vector - Type - Must begin with 'Class'
Column 3 - Vector Name - Defines the name of the vector used throughout GenePilot.
First Data Column to last - Name of the Class - Make sure it's identical to others of same type (i.e. Breast, Lung, etc.)

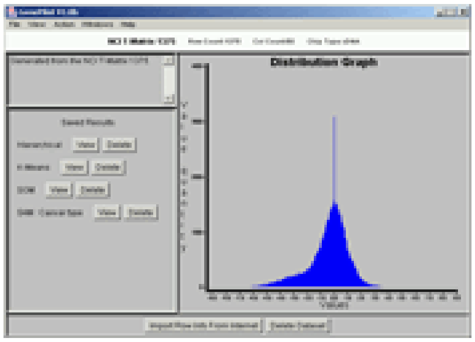
Input - Vectors - Shape Vector <Optional>
This entry defines the Shape Vector associated with this Dataset. This can be a vector from one of the rows or a cluster of rows, it can be the time, or any other set of numbers that define a shape that may be used to analyze the data. You can define and include as many different Shape vectors as you wish. All Vectors must be in the rows immediately following the Column Names row. The following describes the entries needed:
Column 1 - <leave blank>
Column 2 - Vector - Type - Must begin with 'Shape'
Column 3 - Vector Name - Defines the name of the vector used throughout GenePilot.
Column 5 - <leave blank>
Column 5 to n - Column Value - Each column must have a number entry.

Input - Vectors - Pair Data <Optional>
This entry defines the Pair Data associated with this Dataset. This type of vector defines the pairing in Datasets that contain pairs of data. Each pair of columns will have it's own pair of numbers, starting with '1,-1', with one of the pair having a positive value and the other having a negative value. Values must start with 1 (-1) and progress up with no missing gaps. You can define and include as many different Pair vectors as you wish. All Vectors must be in the rows immediately following the Column Names row. The following describes the entries needed:
Column 1 - <leave blank>
Column 2 - Vector - Type - Must begin with 'Pair'
Column 3 - Vector Name - Defines the name of the vector used throughout GenePilot.
Column 4 - <leave blank>
Column 5 to n - Column Value - Each column must have a matching column whose value is matched and opposite.

Input - Data Header <Required>
This entry is used to define the start of the 'Data' portion of the input form. It contains the headers that can be used as a column guide. The first column must begin with '# <Name of first Row Information Field>'

Input - Data Rows <Required>
These rows contain the Data. The following describes the entries needed:
Column 1 - Row Name - Defines the name of this row (i.e. <gene name>). Shown in result windows, with Information.
Column 2 - Identifier - Defines the Identifying number, preferably the Accession number for this row.
Column 3 - Information - Defines the information that you wish to see displayed in the results window, next to each row.
Column 4 - Weight - If no specific weight, set to 1 or leave blank.
Column 5 to n - Column Value - Any value, but **don’t use scientific notation**!

Chapter 2



Main Window

Description

The main window is the control window for GenePilot™. From this window, Datasets can be uploaded, selected, closed or saved.   Preprocessing settings can be set, information for datasets can be viewed, and analytical runs can be configured. The purpose of this window is to control the process of Analyzing MicroArray Data. Following is a description on how to navigate through the process of Analyzing data with GenePilot™.

Menu Choices

Archive Dataset **File->Dataset->Create Archive**
This is used to Save the Dataset, the settings used for preprocessing and the analytics, and the saved results. All of this information is stored in pkzip format, to save room for storage. This format is also very useful for sending the results of a Dataset analysis to a colleague. Upon importing this archive, another user can view exactly the same results as the sender.

Close Dataset **File->Close Dataset->'Dataset Name'**
This selection is used to close an open Dataset. If the Dataset being closed is the current Dataset and another Dataset is open, then the other Dataset will become the current Dataset.

Export Dataset **File->Dataset->Export Dataset**
This Selection is used to export the current dataset out to the same format that is used for loading a Dataset into GenePilot. This can be very useful when a new dataset is created out of a result window.

Load Dataset Archive **File->Dataset->Load Archive**
This is used to load in a Dataset that was Archived with 'Create Archive' (Archive Dataset). This will include the Dataset information, run settings and Saved Results.

Make Template **File->Dataset->Make Template**
This is used to create a template for importing Data. This is likely the very first step that a new user will do, as it is the necessary first step in bringing data into GenePilot.

New Dataset **File->New**
This is used to load a new Dataset into GenePilot. The dataset file should be in the format described in 'Import Format', is needs to be in ASCII format with columns of information separated by a Tab. Once GenePilot has imported the Dataset, the main window will display the 'Dataset Info' screen.

Open Dataset **File->Open->'Dataset Name'**
This selection is used to open a Dataset that has already been loaded into the system. Once GenePilot has opened the Dataset, the main window will display the 'Dataset Info' screen.

Pre-Process Settings **Action->Pre-processing**
This selection is used to view the preprocessing settings for this Dataset. This is where the Filtering, Data Adjustment and Missing Data Imputation selections are made. Go to Pre-Processing section to find out more about the options of this interface.

Quit **File->Quit**
This selection is used to quit out of GenePilot.

Base Interface Menu Choices

Row Info Display **File->Preferences->Row Info Display**
This selection is used to set the Row Information fields that will be displayed as Row Information to the right of the Result Heatmap.

Run Hierarchical **Action->Hierachical**
This selection is used to bring up the interface for selecting the Hierarchical Clustering settings and running the analytical program. Go to Hierachical Clustering to find out more about the options of this interface.

Run K-Means **Action->K-Means**
This selection is used to bring up the interface for selecting the K-Means Clustering settings and running the analytical program. Go to K-Means Clustering to find out more about the options of this interface.

Run SOM **Action->SOM**
This selection is used to bring up the interface for selecting the SOM Clustering settings and running the analytical program. Go to SOM Clustering to find out more about the options of this interface.

Run SAM **Action->SAM**
This selection is used to bring up the interface for selecting the SAM settings and running the analytical program. Go to SAM to find out more about the options of this interface.

Save Dataset **File->Save Dataset**
This selection is used to save the current Dataset settings. This includes preprocessing settings and settings used by Analytic programs run on the Dataset.

Select Open Dataset **File->Select Dataset->'Dataset Name'**
This selection is used to select the current Dataset to use, when more than one Dataset is open. Upon selecting a new Dataset, the main window will display the 'Dataset Info' screen.

Save Settings **File->Save Setup**
This selection is used to save settings that were used for preprocessing or Analytic runs. These settings will become the default settings for future new Datasets. For preprocessing, the settings will be stored according to the Data Type.

Url Targets for Row Fields Interface **File->Preferences->Field Info Urls**
This selection launches the preferences window for Field Info Target Url.

View Dataset Heatmap **View->Current Dataset**
This selection is used to launch a window containing a heatmap of the current Dataset, along with row and column information.

View Dataset Info **View->Datset Info**

This selection is used to view the information that is displayed when a Dataset is opened or imported.

Data Adjustment Settings

Filtering Options

Percent Present >= n (cDNA or Other)

This option filters out rows of data with less than n of present (not missing) values.

Less than n Negative Values < p (Affymetrix)

This option filters out rows that have at least n values less than p.

Standard Deviation (Row Vector) >= n (All)

This option filters out rows of data that have a standard deviation of less than n.

At least n Observations abs(Val) >= p (All)

This option filters out rows that have less than n values which have an absolute value greater than p.

MaxVal – MinVal >= n (All)

This option filters out rows whose maximum value minus the minimum value is less than n.

Calculate Remaining Button

This Button runs the current filter settings to show how many rows will remain after filtering.

Data Adjustment Options

No Data Adjustment

This option results in no data adjustment.

Mean Centering

This option mean centers each row of data.

Median Centering

This option median centers each row of data.

Linear Calibration

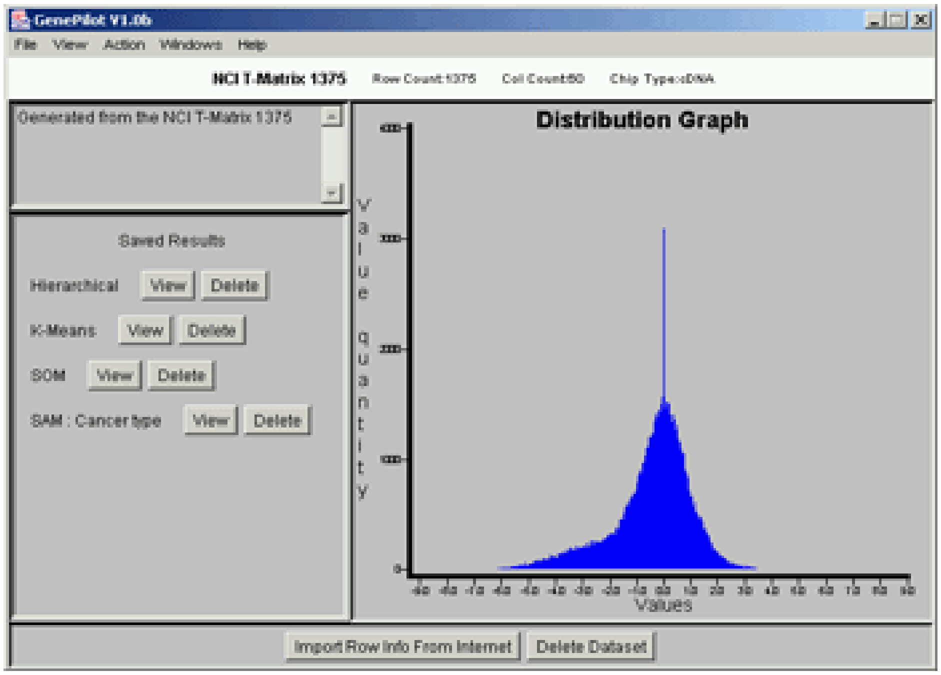
This option is specific for Affymetrix data and Highly recommended by leading experts in Statistics as applied to DNA data.

Missing Data Imputation Options

Nearest Neighbor

This option is currently the only missing-data imputation method offered with this software package.

Dataset Info



NCI T-Matrix 1375 Row Count:1375 Col Count:60 Chip Type:cDNA

Generated from the NCI T-Matrix 1375

Dataset Info Panel

Description

This display shows information about the current Dataset and allows the user to quickly access saved results and delete Dataset and saved results.

Information Panel

This Panel displays the information about this dataset that was either contained in the input template 'Information' field, or it was entered into the Information field when a sub-Dataset was created from a result window.

Results Panel

This Panel lists the saved results from analytic runs. From this panel, a saved result can be launched for viewing or can be deleted.

Graph Panel

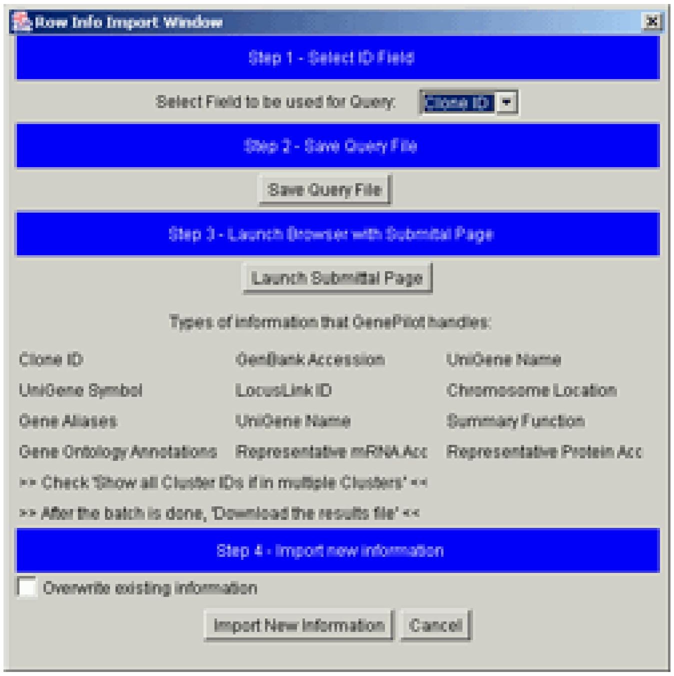
This panel shows a graphical representation the values contained in a Dataset. It shows the distribution of values.

Button Panel
This panel contains buttons for acting on the Dataset. These are the current Buttons:

Import Row Info From Internet - Launches interface for Row Info Import.
Delete Dataset - Deletes the current Dataset.

Interface - Row Info Import

Description
This interface provides the steps to import additional Row Information from the Stanford Source Website, which can greatly enhance the information that can easily be displayed in GenePilot. There is a large selection of information that GenePilot can handle including Gene Ontology information. The only requirement is that the dataset already contains one of the fields that can be used for the query.



Row Info Import Window

Step 1 - Select ID Field

Select Field to be used for Query: Clone ID

Step 2 - Save Query File

Step 3 - Launch Browser with Submittal Page

Types of information that GenePilot handles:

Clone ID	GenBank Accession	UniGene Name
UniGene Symbol	LocusLink ID	Chromosome Location
Gene Aliases	UniGene Name	Summary Function
Gene Ontology Annotations	Representative mRNA Acc	Representative Protein Acc

>> Check 'Show all Cluster IDs if in multiple Clusters' <<

>> After the batch is done, 'Download the results file' <<

Step 4 - Import new information

☐ Overwrite existing information

Row Info Import Interface

Step 1 - Select ID Field
In this section, the Row Info Field that will be used for the query needs to be selected.

Step 2 - Save Query File
In this section, the query file is exported to the File Path that is selected.

Step 3 - Launch Browser with Submittal Page
In this section, GenePilot launches a browser window with the URL set to the batch query submittal. The File that was saved in Step 2 must be specified, then the types of data that are desired must be selected. After the Batch is done running, the 'Download the results file' link needs to be selected and the resulting web page saved.

Step 4 - Import new information
In this section, the new Row Information is imported. Note that if a Dataset has been reloaded into GenePilot and the first three steps have already been taken and a result page is still present, then the second and third steps can be skipped (The Field used in the earlier query must be selected) and the earlier file can be imported. If the 'Overwrite existing information' checkbox is selected then all Row info fields will be overwritten where the incoming data has valid data (if empty field in incoming data, then original data, if present will remain). If the checkbox is not selected, then existing fields with valid data will not be overwritten, but where these fields are empty, valid incoming data will be filled in.

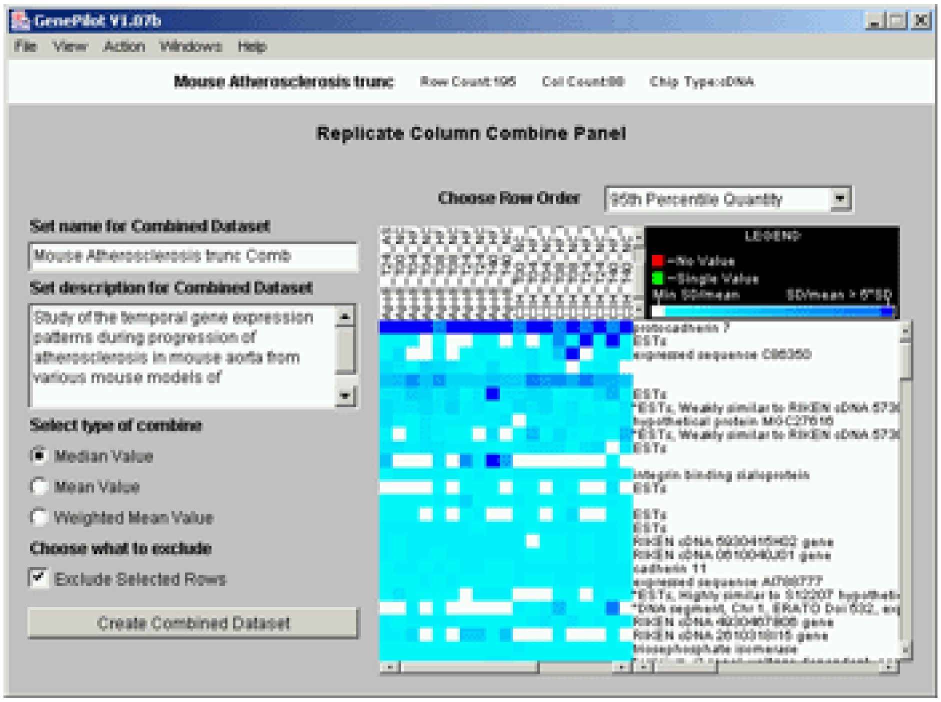
Button - Import New Information

This Button imports the data into GenePilot.

Button - Cancel

This button closes the interface without any additional actions.

Interface - Replicate Column Combine Panel



GenePilot V1.07b
File View Action Windows Help
Replicate Column Combine Panel
LEGEND
=No Value
=Single Value
Min SD/mean
SD/mean > 6*SD
Create Combined Dataset
protocadherin 7
ESTs
expressed sequence C86350
ESTs
*ESTs, Weakly similar to RIKEN cDNA 573
hypothetical protein MGC27616
*ESTs, Weakly similar to RIKEN cDNA 573
ESTs
integrin binding sialoprotein
ESTs
ESTs
ESTs
RIKEN cDNA 5930415H02 gene
RIKEN cDNA 0610040J01 gene
cadherin 11
expressed sequence AI788777
*ESTs, Highly similar to S12207 hypothetic
*DNA segment, Chr 1, ERATO Doi 532, ex
RIKEN cDNA 4930407I05 gene
RIKEN cDNA 2610318I15 gene
triosephosphate isomerase

Mouse Atherosclerosis trunc Row Count:195 Col Count:80 Chip Type:cDNA

Choose Row Order: 95th Percentile Quantity

Set name for Combined Dataset: Mouse Atherosclerosis trunc Comb

Set description for Combined Dataset: Study of the temporal gene expression patterns during progression of atherosclerosis in mouse aorta from various mouse models of

Select type of combine

● Median Value

○ Mean Value

○ Weighted Mean Value

Choose what to exclude

☑ Exclude Selected Rows

Replicate Column Combine Panel

Description

This interface allows for combining replicate Columns using a variety of methods and to also exclude the Rows with bad Data from the resulting Dataset. A display showing the quality of the combined data (measured by SD/Mean of the combined data for each row/column combination). Row information is included in order to see which rows have the bad (or good) data. Those rows can be sorted in two ways and a block of rows can be selected to exclude.

Column and Row Statistics Display

On the right hand side of the interface is a display which shows the quality of the replicates in relation to the other columns. The combined column Name (Name of first column in each set) is displayed at the top. To the right of that is the legend defining the colors used in the Heatmap. The Heatmap utilizes the colors to indicate the quality of the combined values. To select rows, click and drag. If more rows than can be displayed at once need to be selected, page down to ending row and select that row while holding down the ,<shift> key, this will extend the selection down to the selected row. To the right of the Heatmap is the Row Information. Clicking on a row in this section will bring up an information window with all loaded information about that Row.

Set Name for Combined Dataset

This field is for setting the new name of the combined Dataset. By default, 'Comb' is added to the name of the current Dataset. Not that each Dataset name must be unique.

Set Description for Combined Dataset

This field is for setting the description of the combined Dataset. By default, the description of the current dataset is placed in this field. We recommend that you add a description of what selections you used to produce the combined dataset.

Select Type of Combine

These radio buttons offer the selection of the type of method that will be used to combine the values in the replicate columns. If there is only one value then that value is used. If there are two values then the Average of those two values are used. If there are three or more values then the choices are:

Median Value - The median value is used .
Mean Value - The Mean (or Average) value is used.
Weighted Mean Value - The weighted mean value where the weight is equal to 1/variance.

Choose what to exclude

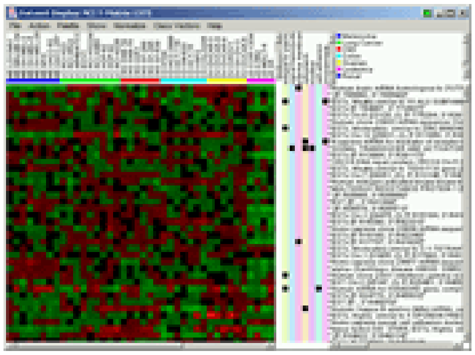
Currently there is one choice of what to exclude:
Exclude Selected Rows - This allows for selecting the worst rows in the dataset to exclude.

Choose Row Order

There are currently two ways to sort the statistics information in the Column and Row Statistics Display. They are:
95th Percentile Quantity - The row order is determined by the count of cels in Row where the SD/Mean of the Combined Data shows a value in the top 5 percent as the primary count and the average value for SD/Mean (With a maximum cut-off value) as a secondary sort.
Average StandDev over Mean - The row order is determined by the average value for SD/Mean (With a maximum cut-off value) as a secondary sort.

Button - Create Combined Dataset
This button executes the creation of the new Dataset with the combined rows.

Chapter 3

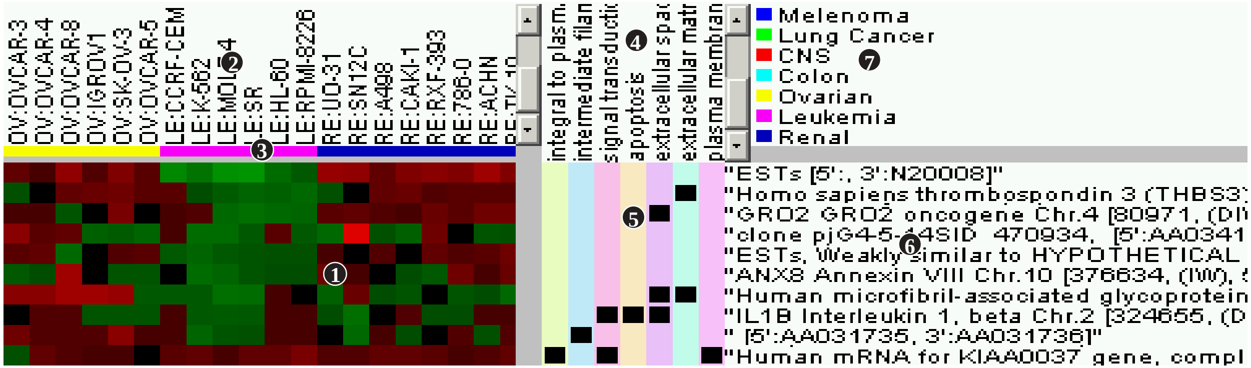


Dataset Display

Description

The Dataset Display provides a means to view a Dataset prior to running analytics on it. It can display all rows, only filtered rows or the rows that remain after filtering. If one or more Classification Vectors are loaded, the rows are sorted by classification.

Parts - Heatmap Panel



Dataset Display Heatmap: 1-Heatmap, 2-Column Information, 3-Class Indication (when avail), 4-Gene Ontology Names (when avail), 5-Gene Ontology Indication (when avail), 6-Row Information, 7-Class Information (when avail)

Description

This panel contains information about the rows. The central part is the Heatmap, which contains colors from the Palette to indicate values of each row (or cluster). When Classification Vector(s) are available a Class Indication row which, indicates the classes of each column, above the Average Heatmap. At the top is the Column Info, in the form of text rotated counterclockwise by 90 degrees. To the right of the Heatmap, is the Gene Ontology indication when Gene Ontology information has been imported into GenePilot for this Dataset. To the right is the Row Information, which contains the description of the row. In the top right corner, the legend for the Class Indication will define the colors for each class, when Classification Vector(s) are available.

Heatmap – Bottom Left

The heatmap contains rectangles color-coded in the Palette colors to represent the value of each cel (juncture of 1 row and 1 column). By default, these colors are Red (highest value) and Green (lowest value). When No-Normalize is selected, the value range is defined by the min and max values in the entire Dataset (Rows remaining after filtering), with zero being the medium value. When Normalize is selected, then the value range for each row is defined by the min and max values in that row, with the mean value being the medium value.

Class Indication – Above Heatmap

If there is Classification Vector(s), there will be Class Indicators above the Average Heatmap. This is a row of rectangles, which are color-coded to indicate the class of each column as defined by the Class Information block in the Top Right corner.

Column Names – Upper Left

The Column Information contains the column names, rotated counterclockwise by 90 degrees, above each column in the Heatmap.

Gene Ontology Names - Middle Right (If Gene Ontology Information)
The Gene Ontology Names is a variable-column display that displays the Gene Ontology Names that are currently chosen either automatically (Most common in dataset or most common in selected rows) or by the User. Using File->Preferences->Gene Ontology Display, these different choices can be set. If specific Gene Ontologies are selected by the user, the settings can be easily changed back to automatic selection by clicking on this (Gene Ontology Names) display.

Gene Ontology Indication - Bottom Middle
The Gene Ontology Indication indicates the Gene Ontologies that are associated with each row. The Gene Ontologies are chosen either automatically (Most common in dataset or most common in selected rows) or by the User. Using File->Preferences->Gene Ontology Display, these different choices can be set.

Row Information – Bottom Right
Row information contains the row name and row information to the right of each row in the heatmap.

Class Information – Top Right (If Class Vector)
If there is Classification Vector(s), there will be Class Information which defines the color coding for the Class Indication row.

Actions:

1. Clicking and dragging on the bitmap will select those rows.
2. Holding down the shift key and selecting a row will extend the row from the previous selection down to the selected row. This can be used to select rows in a group that is larger than those seen on the screen.

Dataset Display- Menu Choices

About **Help->About**
This selection shows the about screen for GenePilot.

Bitmap from Selected **Action->Save Sel to Bitmap**
This selection creates a gif bitmap containing all of the information specified by the selected rows.

Bitmap from All **Action->Save All to Bitmap**
This selection creates a gif bitmap containing all of the information in the Left Panel.

Copy to Clip **Action->Copy to Clip**
This selection launches an interface for copying all or selected rows to the clipboard. This interface also allows for selecting the Row Information Fields which the user is interested in.

Copy to File **Action->Copy to File**
This selection launches an interface for copying all or selected rows to a file. This interface also allows for selecting the Row Information Fields which the user is interested in.

Create Vector **Action->Create Vector**
This selection launches the Create Vector dialog box along with the mean vector of the currently selected rows.

Gene Ontology Display **File->Preferences->Gene Ontology Display**
This selection launches the preferences window for Gene Ontologies. Within this preference window the following can be set; set number of columns, set automatic preferences for selection of Gene Ontologies, select specific Gene Ontologies to display.

Generate Gene Ontology Statistics **Action->Generate GO Statistics**
This selection launches an html page that contains the statistics for the Gene Ontologies for the rows in the following order of precedence: 1. Selectd rows . 2. All rows being displayed.

Help **Help->Help**
This selection shows the Help for the Dataset Display.

Help-PDF **Help->Help-PDF**
This selection launches an html browser with this User manual.

Launch Info Page **Action->Launch Info Page**
This selection launches an information page containing a matrix of the genes vs. Gene Ontologies (if available) and a listing of all of the Row Information Fields along with links to websites which have further information.

Make Dataset **Action->Make Sub-Dataset**
Not currently implemented

Quit **File->Quit**
This selection closes the Result Window.

Row Info Display **File->Preferences->Row Info Display**
This selection launches the preferences window for Row Information. This preferences window is used to define the Row Information that is displayed to the right of the heatmap.

Search **Action->Search**
This selection launches the search interface, which allows the user to search the rows for a String. See ‘Search Interface’ for more information on the Search Interface.

Search Again **Action->Search Again**
This selection searches for the next case of the search string using the settings from the previous search.

Select Classification Vector (If Avail) **Classes-> 'Class Vector Name'**
If there is one or more Classification Vectors, then one of those vectors can be selected for class indication of columns in either the Column Dendigram (if column clustering was selected) or the Class Indicator, in the Right Panel.

Select Heatmap Palette – Red/Green **Palette->Red-Green**
This selection selects the traditional colors of Red and Green for the heatmap. Red is positive or Correlated and Green is negative or Anti-Correlated.

Select Heatmap Palette – Yellow/Blue **Palete-> Yellow -Blue (Default)**
This selection selects the traditional colors of Yellow and Blue for the heatmap. Yellow is positive or Correlated and Blue is negative or Anti-Correlated.

Select Heatmap Palette – Gray Scale **Palette->Gray Scale**
This selection selects shades of gray for the heatmap. Light Gray is positive or Correlated and Dark Gray is negative or Anti-Correlated.

Show Palette **Palette->Show Palette**
This selection Launches the Palette Window, which shows the color palette used by the heatmap.

Show Rows Normalized **Normalize->Norm**
This selection changes the heatmap to show each row normalized. That means that the color corresponding to the highest value in the palette will be applied to the highest value in the row, the color corresponding to the lowest value in the palette will be applied to the lowest value in the row, and all other values will be scaled accordingly.

Show Rows Un-Normalized **Normalize->NoNorm**
This selection displays the traditional heatmap display where color selections from the palette are determine from the highest and lowest value among the data (not just in a row).

View Original Data **Show->Original Data**
This selection shows the original data that was imported into GenePilot.

View Adjusted Data **Show->Adjusted Data**
This selection shows the adjusted data (if preprocessing has been done on the Dataset). This means that rows that have been filtered out are not included, missing data has been filled in through calculations and the Data has been adjusted.

View Filtered Out Data **Show->Filtered Out Data**
This selection shows the Rows of data that have been filtered out during preprocessing.

Url Targets for Row Fields Interface **File->Preferences->Field Info Urls**
This selection launches the preferences window for Field Info Target Url.

Window Selection **Windows-><Window Name>**
This selection brings the selected window up to the front.

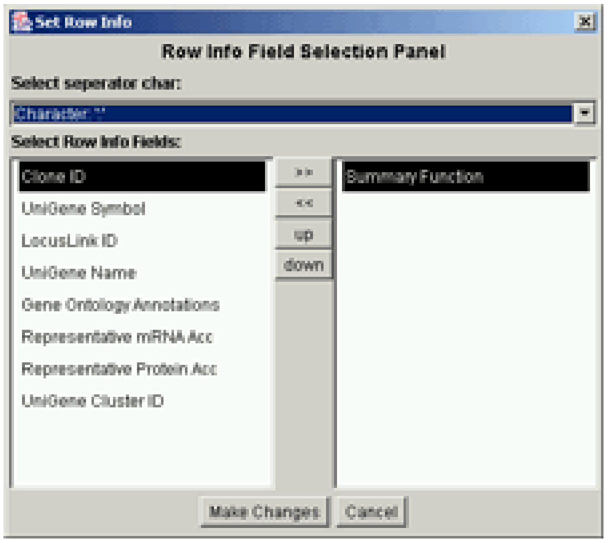
Preferences - Row Info Display

Description

The Row Info Display Preferences Window is used to specify the Row Information fields that will be displayed to the right of the Heatmap. In addition to selecting the fields, the field order can be specified along with the character separating the contents of each field. Fields are selected for inclusion by moving them from the left column (Available) over to the right column (Included). The order of display is set by their row order, with the first row being the first display field.

Select separator char

This selection sets the character that will separate the information from each field.



Row Info Field Selection Panel

Select Row Info Fields:

This section allows for selection of the specific fields that will be displayed in the row info and the order of these fields. It has the following buttons:

>> - This button moves the currently hilited field in the left column over to the right column.
<< - This button moves the currently hilited field in the right column over to the left column.
up - This button moves the currently hilited field in the right column up one row.
down - This button moves the currently hilited field in the right column down one row.

Button - Make Changes

This button must be clicked in order for the changes to be made.

Button - Cancel

This button cancels any changes that were made and closes the window.

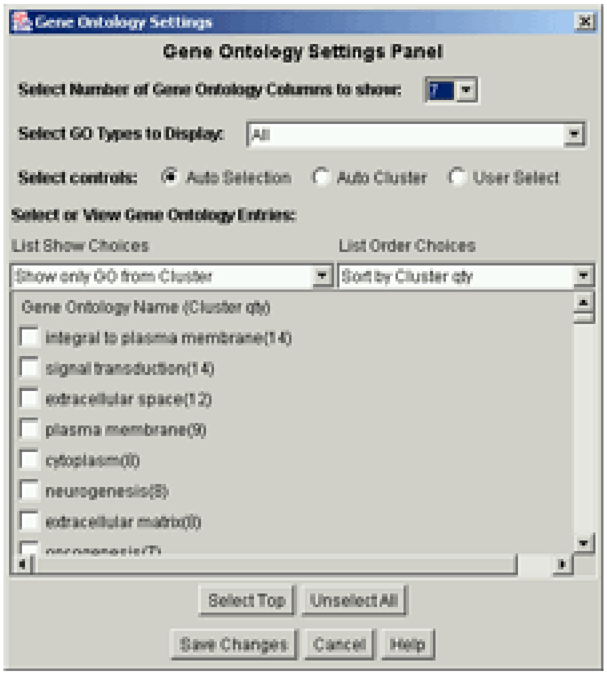
Preferences - Gene Ontology

Description

This interface is used to set the Gene Ontology preferences for this screen. The choices include automatic selection of the most common Gene Ontology categories based upon either the currently selected rows or for all rows. If the selected rows have preference and there aren't enough categories to fill the number of columns, then the most common categories across all rows are used to fill in the categories. Specific Gene Ontology categories can also be specified, this is aided by the List Show Choices and List Order Choices for quickly finding specific Gene Ontologies.

Select Number of Gene Ontology Columns to show

This selection sets the number of columns that will be displayed.



Gene Ontology Settings

Gene Ontology Settings Panel

Select Number of Gene Ontology Columns to show: 7

Select GO Types to Display: All

Select controls: ● Auto Selection ○ Auto Cluster ○ User Select

Select or View Gene Ontology Entries:

List Show Choices: Show only GO from Cluster

List Order Choices: Sort by Cluster qty

Gene Ontology Name (Cluster qty)

□ integral to plasma membrane(14)
□ signal transduction(14)
□ extracellular space(12)
□ plasma membrane(9)
□ cytoplasm(8)
□ neurogenesis(8)
□ extracellular matrix(8)

Gene Ontology Settings Panel

Select GO Types in Display

This selection sets the Gene Ontology Categories to Display with the following options:

Biological Process - Shows only Biological Process
Cellular Component - Shows only Cellular Component
Biological Process & Cellular Component - Shows both Categories
Molecular Function - Shows only Molecular Function
Biological Process & Molecular Function - Shows both Categories
Cellular Component & Molecular Function - Shows both Categories
All - Shows all three Categories

Select control for Gene Ontology Columns

This selection sets the way that Gene Ontology Columns are selected with the following options:

Auto Selection - Chooses most common Gene Ontologies among selected rows (if avail) then fills in remaining, if necessary, from most common among remaining rows.
Auto Set - Chooses most common Gene Ontologies among displayed rows.
User Select - Displays rows that the user has specifically selected. This gets turned off when the user clicks on the Gene Ontology Names.

Select or View Gene Ontology Entries
This section is for viewing or selecting specific Gene Ontologies by checking the checkbox either manually or using the ‘Select Top’ button under the list.

List Show Choices
This selection sets the Gene Ontologies that will be displayed in the list for selection. It has the following choices:

Show All GO - Displays all of the Gene Ontologies that are associated with the current Dataset.
Show only GO from Set - Displays the Gene Ontologies that are associated with the currently displayed genes.
Show only GO from Selected - Displays only the Gene Ontologies that are associated with the currently selected genes.

List Order Choices
This selection sets the order that the Gene Ontologies are displayed in the list for selection. It has the following choices:

Sort by name - Sorts the Gene Ontologies by their name.
Sort by Set qty - Sorts by the number of times that each Gene Ontology is associated to a gene in the currently displayed genes.
Sort by Selected qty - Sorts by the number of times that each Gene Ontology is associated to a gene in the currently selected genes.

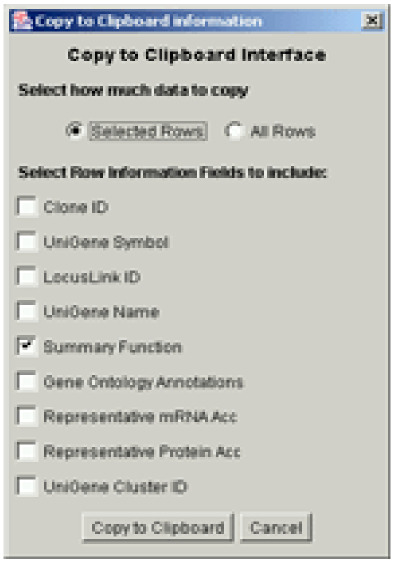
Button - Select Top
This button selects the Gene Ontologies that are at the top of the list, in the quantity specified.

Button - Unselect All
This button unselects all of the currently selected Gene Ontologies.

Button - Save Changes
This button sets the currently selected settings and closes the window.

Button - Cancel
This button closes the window without setting any of the changes made.

Interface - Copy to Clipboard


Copy to Clipboard information

Copy to Clipboard Interface

Select how much data to copy

● Selected Rows ○ All Rows

Select Row Information Fields to include:

□ Clone ID
□ UniGene Symbol
□ LocusLink ID
□ UniGene Name
☑ Summary Function
□ Gene Ontology Annotations
□ Representative mRNA Acc
□ Representative Protein Acc
□ UniGene Cluster ID

Copy to Clipboard Interface

Description

This Interface allows the user to select which Row Information Fields to include in the Copy to Clipboard and to also select whether they want to copy only the selected rows or all of the rows.

Select how much data to copy

This selection allows the user to select whether the rows copied include only selected rows or all rows.

Select Row Information Fields to include

This section allows the user to specify which rows they'd like to include in the Copy to Clipboard. Each Row Information Field that is checked will be included in the copy.

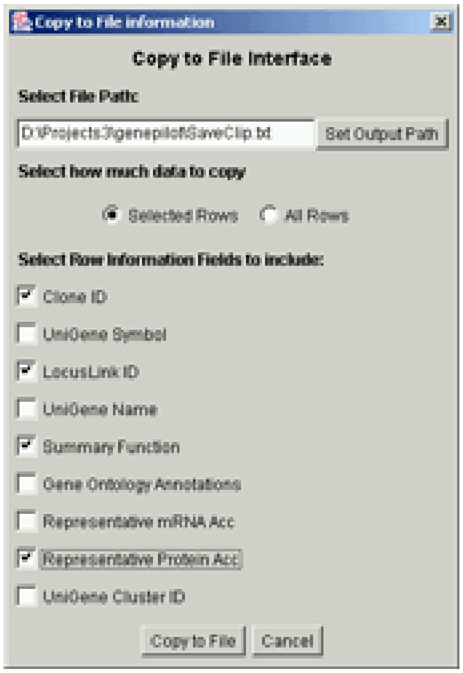
Button - Copy to Clipboard

This Button Initializes the copy and closes the window

Button - Cancel

This Button closes the window without doing the copy.

Interface - Copy to File


Copy to File information

Copy to File Interface

Select File Path:

D:\Projects3\genepilot\SaveClip.txt

Select how much data to copy

● Selected Rows ○ All Rows

Select Row Information Fields to include:

☑ Clone ID
□ UniGene Symbol
☑ LocusLink ID
□ UniGene Name
☑ Summary Function
□ Gene Ontology Annotations
□ Representative mRNA Acc
☑ Representative Protein Acc
□ UniGene Cluster ID

Copy to File Interface

Description

This Interface allows the user to select which Row Information Fields to include in the Copy to File and to also select whether they want to copy only the selected rows or all of the rows.

Select File Path

This selection is used to set the path to the output file. Click on the 'Set Output Path' to set the file. Traditional copy and paste methods work in this field.

Select how much data to copy

This selection allows the user to select whether the rows copied include only selected rows or all rows.

Select Row Information Fields to include

This section allows the user to specify which rows they'd like to include in the Copy to File . Each Row Information Field that is checked will be included in the copy.

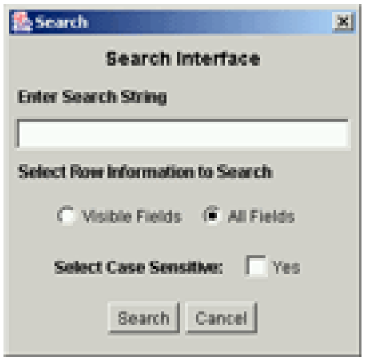
Button - Copy to File
This Button Initializes the copy and closes the window

Button - Cancel
This Button closes the window without doing the copy.

Interface - Search

Description
This interface is used to set the search term and the settings for the search. The search can be restricted to just the visible Row Information or to all Row Information. It can also be set to be case sensitive.

Enter Search String
The search string is entered into this field. Traditional copy and paste methods work on this field.



○ Visible Fields ● All Fields

Select Case Sensitive: □ Yes

Search Interface

Select Row Information to Search
Selecting ‘Visible Fields’ restricts the search to just the information that is visible, as set in the Preferences->Row Info Display. Selecting All Fields causes the search to include all Row Info Fields, whether visible or not.

Select Case Sensitive
Checking this box will cause the search to be case sensitive.

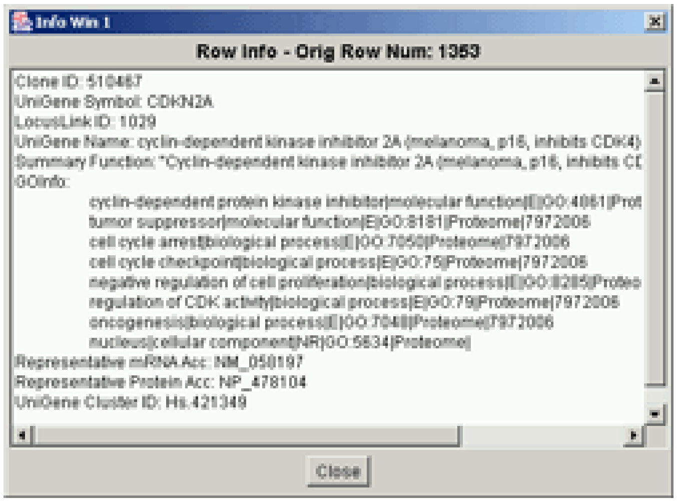
Button - Search
This Button activates the search.

Button - Cancel
This button closes the interface without searching.

Common Windows - Row Information Window

Description

The Row Information Window displays all of the information that GenePilot knows about the current row of information. This window is launched by clicking on a row within the Row Info section of a result screen. It is a single-row version of the Info Page that contains information about one or more rows of information, a hashtable showing the relationships between the selected genes and the Gene Ontology Categories (when Gene Ontology information is imported or loaded) and links to additional information.



Row Info - Orig Row Num: 1353

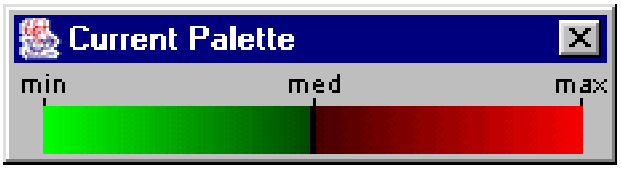
```
Clone ID: 510467
UniGene Symbol: CDKN2A
LocusLink ID: 1029
UniGene Name: cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)
Summary Function: "Cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CI
GOInfo:
            cyclin-dependent protein kinase inhibitor|molecular function|E|GO:4861|Prot
            tumor suppressor|molecular function|E|GO:8181|Proteome|7972006
            cell cycle arrest|biological process|E|GO:7050|Proteome|7972006
            cell cycle checkpoint|biological process|E|GO:75|Proteome|7972006
            negative regulation of cell proliferation|biological process|E|GO:8285|Proteo
            regulation of CDK activity|biological process|E|GO:79|Proteome|7972006
            oncogenesis|biological process|E|GO:7048|Proteome|7972006
            nucleus|cellular component|NR|GO:5634|Proteome|
Representative mRNA Acc: NM_058197
Representative Protein Acc: NP_478104
UniGene Cluster ID: Hs.421349
```

Row Information Window

Additional Information

When a Row Information window is launched, a reference gets added to the Window menu. Multiple instances of this window can be launched at the same time, which each new one getting a unique number (in order). Selecting this window from the Window menu will bring it to the front. Traditional copy methods should work on all platforms (i.e. <ctrl>c on Windows) to allow for copying and pasting into another program.

Common Windows - Palette



Heatmap Palette Window

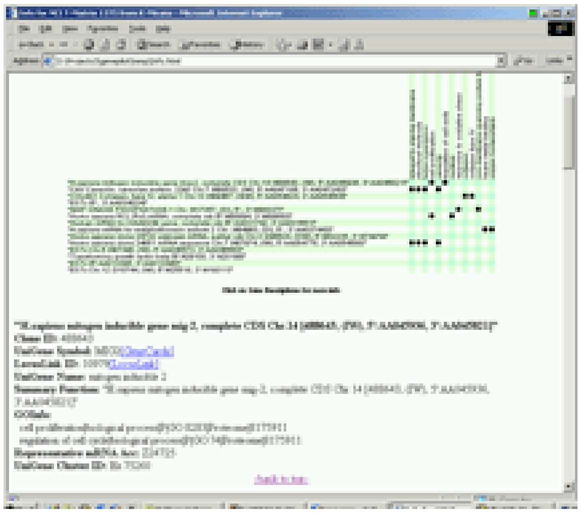
Description

The Palette Window displays the palette that is being used by the Heatmap. When Normalize is selected, the min and max colors represent the min and max values for each row. When Normalize is not selected, then the medium value is 0 and the min and max values are determined by the Absolute maximum value on either side of zero (if the lowest value is -3 and the largest value is 5, then min will be equal to -5 and max will be equal to 5).

Common Page - Row Information Page

Description

The Row Information Page is an html page that is generated by GenePilot. It includes a hashtable showing the relationships between the selected genes and the Gene Ontology Categories (when Gene Ontology information is imported or loaded), all of the Row information for each of the currently selected rows and links to more information from the fields that are configured to have links (using Preferences->Set Links). The purpose of this page is to provide maximum information for selected genes in a format that can be easily emailed to a colleague.



Row Information Page

Gene Ontology Hashtable

This image, which has the same prefix as the webpage, contains a hashtable of up to 50 of the most common Gene Ontologies with circles in the rows that they are included in. Clicking on the row description (same description as set by the user for Row Information) will take the user to the full description of known information contained within GenePilot.

Row Information

Each Row among the selected rows will have it’s own information section in this html page. Within this section will be displayed all of the information that has been loaded into GenePilot for this Row. Also within this section will be links to additional information if there is information that has a valid link set.

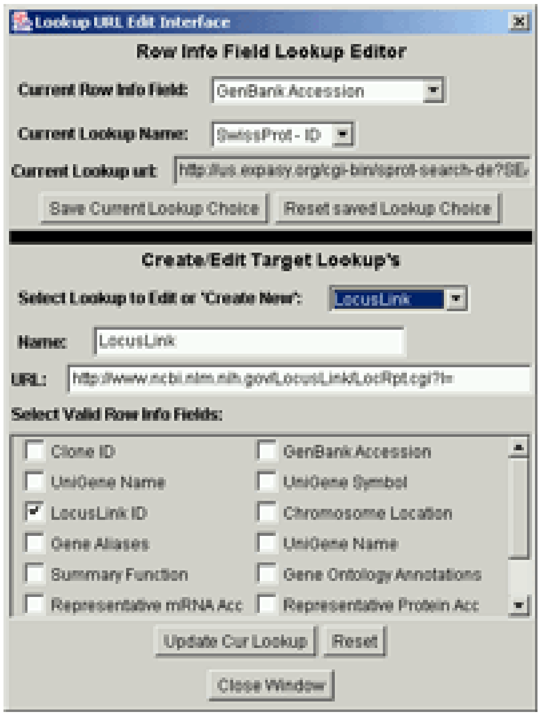
Additional information

The image at the top (when there is Gene Ontology Information) has the same name (ending in .gif) as the html page. This is to make it easy to either save the page for later viewing or to send the pair to a colleague for their viewing.

Common Interfaces - Url Lookup Editor

Description

The Url Lookup Editor provides a means to set Lookup Urls for Row Info Fields. It is broken into two parts: The Row Info Field Lookup Editor which sets a Lookup Url to a Row Info Field. The Create Edit Lookup Url's section which adds new Urls that can then be assigned to a Row Info Field.


Lookup URL Edit Interface
Row Info Field Lookup Editor
Save Current Lookup Choice
Reset saved Lookup Choice
Create/Edit Target Lookup's
Select Valid Row Info Fields:
Update Cur Lookup
Reset
Close Window

Current Row Info Field: GenBank Accession

Current Lookup Name: SwissProt - ID

Current Lookup url: http://us.expasy.org/cgi-bin/sprot-search-de?SE

Select Lookup to Edit or 'Create New': LocusLink

Name: LocusLink

URL: http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l=

☐ Clone ID	☐ GenBank Accession
☐ UniGene Name	☐ UniGene Symbol
☑ LocusLink ID	☐ Chromosome Location
☐ Gene Aliases	☐ UniGene Name
☐ Summary Function	☐ Gene Ontology Annotations
☐ Representative mRNA Acc	☐ Representative Protein Acc

Url Lookup Editor

Current Row Info Field

This selection sets the current Row Info Field.

Current Lookup name

This selection contains a list of the Lookup's that are available for the current Row Info Field. When a Row Info Field is first set, it displays the current Lookup selected for the current Row Info Field. To Change the set Lookup for a Row Info Field, select a different Lookup then click on the 'Save Current Lookup Choice' button.

Current Lookup Path

This text field displays the actual Url for the currently selected Lookup.

Button - Save Current Lookup Choice

This Button saves the Currently selected Lookup as the Lookup for the current Row Info Field.

Button - 'Reset saved Lookup Choice

This Button restores the currently saved Lookup for the current Row Info Filed.

Select Lookup to Edit or 'Create New'

This selection is for selecting a Lookup to edit or choosing to create a new Lookup.

Name

This selection displays the current name for a Lookup. To update a name for an existing Lookup, change the name in this field then click on the 'Update Cur Lookup' button. When creating a new Lookup, the name of that Lookup will be entered here.

URL

This selection displays the current Url for a Lookup. To update a Url for an existing Lookup, change the Url in this field then click on the 'Update Cur Lookup' button. When creating a new Lookup, the Url will be entered here.

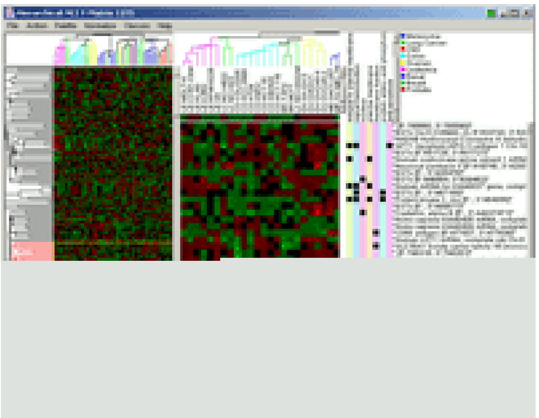
Select Valid Row Info Fields
This section with it’s check boxes sets the Row Info Fields which can use this Lookup.

Button - Save Cur Lookup
This button saves the new Lookup.

Button - Update Cur Lookup
This button updates the current Lookup.

Button - Reset
This button resets the fields back to their saved settings for an existing Lookup and clears the fields for a new Lookup.

Chapter 4



Hierarchical Clustering

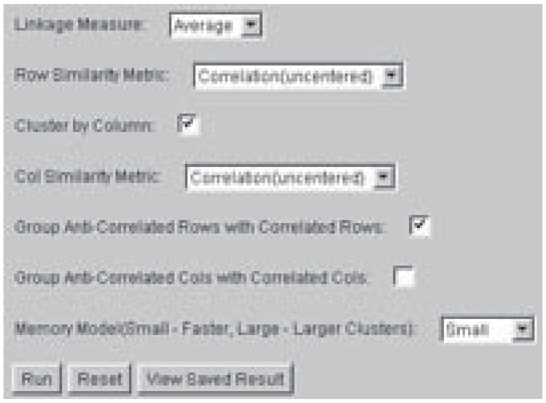
Description

Hierarchical Clustering is one of the most commonly used unsupervised methods for analyzing MicroArray Data. It is a bottom-up Clustering method, that starts at individual rows, clustering rows/clusters together until all of the rows and clusters are represented by a single cluster. The columns can also be clustered in an identical manner. This method provides a very good 'first look' at the Dataset. Patterns can be very easily spotted, in the heatmap, with the data clustered this way. Utilizing the color coded column dendigram, the quality of the samples can quickly be discerned by how well classes group together. For more information on Hierarchical Clustering, read more about it in 'Cluster Analysis and Display of Genome-Wide Expression Patterns'(Reference 3 - MB Eisen, et al.).

Run Settings

Description

The Run Settings provide a means to alter the way that Hierarchical Clustering analyzes the data. It is recommended that the default settings be used, prior to gaining more in-depth knowledge on the ramifications of each setting. In the vast majority of cases, the default settings will be more than adequate.



Linkage Measure: Average

Row Similarity Metric: Correlation(uncentered)

Cluster by Column: ☑

Col Similarity Metric: Correlation(uncentered)

Group Anti-Correlated Rows with Correlated Rows: ☑

Group Anti-Correlated Cols with Correlated Cols: ☐

Memory Model(Small - Faster, Large - Larger Clusters): Small

Linkage Measure

This setting determines which part of a cluster is used to determine the distance from another cluster or row. The choices are:

1. Average – The average vector, for the cluster is calculated, then used for calculating distances.
2. Single – The closest row vector, in a cluster, to another row vector or cluster is used for calculating distances.

Row Similarity Metric

This setting determines how Pearson's correlation will be applied to the row clustering. There are two choices:

1. Correlation(centered) – The row is mean centered before calculating.
2. Correlation(uncentered) – The row is not centered before calculating.

Cluster by Column Checkbox

This Checkbox determines whether the columns of this dataset will be clustered. If the box is checked, the columns will be clustered.

Col Similarity Metric

This setting determines how Pearson's correlation will be applied to the column clustering. There are two choices:

1. Correlation(centered) – The row is mean centered before calculating.
2. Correlation(uncentered) – The row is not centered before calculating.

Group Anti-correlated Rows with Correlated Rows Checkbox

This setting determines whether the value Pearson's Correlation will be made an absolute value before subtracting it from 1, in row clustering distance calculations. If the absolute value is used, then two vectors that are mirrors (vertically) of each other will cluster close to each other. If the absolute value isn't used, then two vectors that are mirrors of each other will not be clustered together.

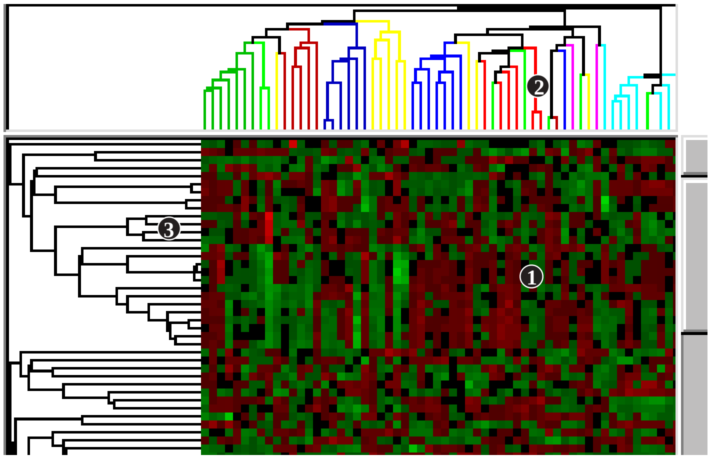
Group Anti-correlated Cols with Correlated Cols Checkbox

This setting determines whether the value Pearson's Correlation will be made an absolute value before subtracting it from 1, in col clustering distance calculations. If the absolute value is used, then two vectors that are mirrors (vertically) of each other will cluster close to each other. If the absolute value isn't used, then two vectors that are mirrors of each other will not be clustered together.

Memory Model (Small – Faster, Large- Larger Clusters)

This setting determines how many stored distances are used for row clustering. A simple model of Hierarchical clustering stores every distance between every row. What this means is that if there are n rows, then the size of memory required to run will be (n) x (n – 1) x 4, where 4 is the number of bytes necessary to hold the information. This means that for a 10k row dataset, 400megs of ram will be used, just to hold the distance calculations. The alternative is to calculate the distances every time, leading to prohibitive run times. The solution that GenePilot uses is to only store a percentage of the closest distances. This is more complicated in the traditional method, but it runs faster and uses far less memory than the traditional method. There are 3 choices (small, medium, large), with larger clusters. For most uses, the Small memory model should be sufficient. The smallest using less memory and running faster. The Largest uses more memory and results in

Results - Parts - Left Panel



Left Panel Parts: 1-Heatmap, 2-Column Dendigram, 3-Row Dendigram

Description

This panel displays a small cel version of the heatmap, with a row dendigram on the left, and a column dendigram, above the heatmap, when the columns have been clustered. The top dendigram is color coded, if Classification Vector(s) are available. Clusters can be selected by either clicking on a node, in the left dendigram or selecting rows in the heatmap. Selecting these clusters will cause the associated rows to be displayed in the right panel.

Heatmap – Center

The heatmap contains rectangles color-coded in the Palette colors to represent the value of each cel (juncture of 1 row and 1 column). By default, these colors are Red (highest value) and Green (lowest value). When No-Normalize is selected, the value range is defined by the min and max values in the entire Dataset (Rows remaining after filtering), with zero being the medium value. When Normalize is selected, then the value range for each row is defined by the min and max values in that row, with the mean value being the medium value.

Column Dendigram – Top (Column Clustering selected)

If Column clustering has been selected, there will be a dendigram, at the top. If there is a Category Vector, for this dataset, the top dendigram will be color-coded, by class. The column dendigram displays the cluster tree of the columns. Similar columns are clustered together with either another column or a cluster, depending on which is closer. Distances can be roughly determined by the height of the node that binds the two entities (columns or clusters) together.

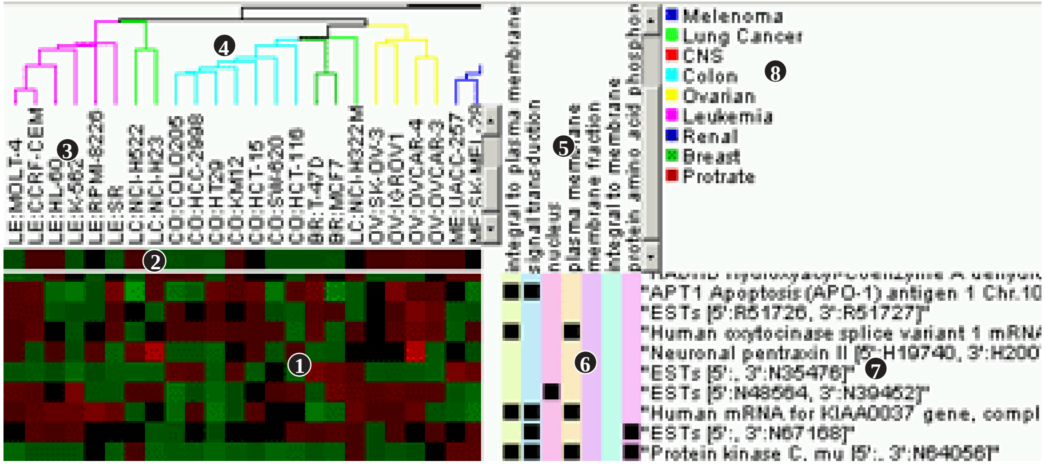
Row Dendigram –Left

To the left of the heatmap, is the row dendigram. The Row Dendigram displays the cluster tree of the rows. Similar rows are clustered together with either another row or a cluster, depending on which is closer. Distances can be roughly determined by the distance of the node that binds the two entities (row or cluster), from the heatmap on the right.

Actions:

1. Clicking on a node point of a cluster will select the cluster represented by that node. The selected rows will be displayed in the right panel.
2. Clicking on the heatmap and dragging the mouse, while keeping the left mouse button depressed will select those rows. Note that more rows than originally chosen will usually be selected, as the program will look for a common node to the selected rows, then select all rows owned by the cluster represented by a node. The selected rows will be displayed in the right panel.
3. Holding down the shift key and selecting a row will extend the row from the previous selection down to the selected row. This can be used to select rows in a group that is larger than those seen on the screen. The selected rows will be displayed in the right panel.

Results - Parts - Right Panel



Results -Right Panel: 1-Heatmap, 2-Average Heatmap, 3-Column Information, 4-Column Dendigram, 5-Gene Ontology Descipriptions, 6-Gene Ontology Indications, 7-Row Information, 8-Class Information, <Not Shown>-Class Indication

Description

This panel contains the specific information, for the current cluster that has been selected in the Left Panel. The central part is the Heatmap, which contains colors from the Palette to indicate values of each row (or cluster). Above the Heatmap is the Average Heatmap, a single-row heatmap that indicates the average vector of the current cluster. When Classification Vector(s) are available and there was no column clustering, a Class Indication row which, indicates the classes of each column, is above the Average Heatmap. Above all of this is the Column Info, in the form of text rotated counter-clockwise by 90 degrees. If the columns have been clustered, the Column Info is topped by a dendigram of the column clustering, which is either black or color-coded, depending on if Classification Vector(s) are available. To the right of the Heatmap, is the Gene Ontology indication when Gene Ontology information has been imported into GenePilot for this Dataset.To the right is the Row Information, which contains the description of the row. In the top right corner, the legend for the Class Indication will define the colors for each class, when Classification Vector(s) are available.

Heatmap – Bottom Left

The heatmap contains rectangles color-coded in the Palette colors to represent the value of each cel (juncture of 1 row and 1 column). By default, these colors are Red (highest value) and Green (lowest value). When No-Normalize is selected, the value range is defined by the min and max values in the entire Dataset (Rows remaining after filtering), with zero being the medium value. When Normalize is selected, then the value range for each row is defined by the min and max values in that row, with the mean value being the medium value.

Average Heatmap – Above Heatmap

The Average Heatmap is a single-rowed version of the Heatmap. It represents the average vector for the current cluster.

Class Indication – Above Average Heatmap
If there is Classification Vector(s), there will be Class Indicators above the Average Heatmap. This is a row of rectangles, which are color-coded to indicate the class of each column as defined by the Class Information block in the Top Right corner.

Column Names – Upper Left
The Column Information contains the column names, rotated counter-clockwise by 90 degrees, above each column in the Heatmap.

Column Dendigram – Top Left (If Column Clustering Selected)
If Column clustering has been selected, there will be a dendigram, at the top. If there is a Category Vector, for this dataset, the top dendigram will be color-coded, by class. The column dendigram displays the cluster tree of the columns. Similar columns are clustered together with either another column or a cluster, depending on which is closer. Distances can be roughly determined by the height of the node that binds the two entities (columns or clusters) together.

Gene Ontology Names - Middle Right (If Gene Ontology Information)
The Gene Ontology Names is a variable-column display that displays the Gene Ontology Names that are currently chosen either automatically (Most common in dataset or most common in selected rows) or by the User. Using File->Preferences->Gene Ontology Display, these different choices can be set. If specific Gene Ontologies are selected by the user, the settings can be easily changed back to automatic selection by clicking on this (Gene Ontology Names) display.

Gene Ontology Indication - Bottom Middle
The Gene Ontology Indication indicates the Gene Ontologies that are associated with each row. The Gene Ontologies are chosen either automatically (Most common in dataset or most common in selected rows) or by the User. Using File->Preferences->Gene Ontology Display, these different choices can be set.

Row Information – Bottom Right
Row information contains the row name and row information to the right of each row in the heatmap.

Class Information – Top Right (If Class Vector)
If there is Classification Vector(s), there will be Class Information which defines the color coding for the Class Indication row.

Actions:

1. Clicking and dragging on the bitmap will select those rows.
2. Holding down the shift key and selecting a row will extend the row from the previous selection down to the selected row. This can be used to select rows in a group that is larger than those seen on the screen.
3. When user selected gene ontologies are displayed, the Gene Ontology displays can be reverted back to automatic mode by clicking on the Gene Ontology Names.

Results - Menu Choices

Bitmap from Selected **Action->Save Sel to Bitmap**
This selection creates a gif bitmap containing all of the information in the right Panel. If no rows are selected, in the Left Panel, then it will not produce a Bitmap. If rows are selected in the Right Panel, then those rows will be included. If rows are selected in the Left Panel, but none are selected in the Right Panel, then all rows in the Right Panel will be included in the Bitmap.

Bitmap from All **Action->Save All to Bitmap**
This selection creates a gif bitmap containing all of the information in the Left Panel.

Copy to Clip **Action->Copy to Clip**
This selection launches an interface for copying all or selected rows to the clipboard. This interface also allows for selecting the Row Information Fields which the user is interested in.

Copy to File **Action->Copy to File**
This selection launches an interface for copying all or selected rows to a file. This interface also allows for selecting the Row Information Fields which the user is interested in.

Create Vector **Action->Create Vector**
This selection launches the Create Vector dialog box along with the mean vector of the currently selected rows.

Gene Ontology Display **File->Preferences->Gene Ontology Display**
This selection launches the preferences window for Gene Ontologies. Within this preference window the following can be set; set number of columns, set automatic preferences for selection of Gene Ontologies, select specific Gene Ontologies to display.

Generate Gene Ontology Statistics **Action->Generate GO Statistics**
This selection launches an html page that contains the statistics for the Gene Ontologies for the rows in the following order of precedence: 1. Selectd rows in Right Panel. 2. Rows of currently selected cluster in left panel. 3. All rows in left panel.

Launch Info Page **Action->Launch Info Page**
This selection launches an information page containing a matrix of the genes vs. Gene Ontologies (if available) and a listing of all of the Row Information Fields along with links to websites which have further information.

Hierarchical Clustering Results - Menu Choices

Make Dataset **Action->Make Sub-Dataset**

This selection launches and interface that let's the user define criteria for creating a new Dataset from rows of the current Dataset. For Hierarchical Clustering, there will need to be a cluster selected in the Left Panel. Then those rows or the selected rows, among those rows can be used as the rows to use in a new Dataset, or they can be excluded from a new Dataset

Quit **File->Quit**

This selection closes the Result Window.

Save Results **File->Save Results**

This selection is used to save the current result. This will save the results of an Analytic run so that the results can be viewed immediately. Once an analytic result is saved, for a Dataset, a button will be added to the Dataset Information screen to quickly view the saved result.

Search **Action->Search**

This selection launches the search interface, which allows the user to search the rows for a String. See 'Search Interface' for more information on the Search Interface. For Hierarchical Clustering, the search will search the currently selected cluster (from left Panel) first, it will then search beyond that and start at the beginning, when it reaches the end of the Dataset.

Search Again **Action->Search Again**

This selection searches for the next case of the search string.

Select Classification Vector (If Avail) **Classes-> 'Class Vector Name'**

If there is one or more Classification Vectors, then one of those vectors can be selected for class indication of columns in either the Column Dendigram (if column clustering was selected) or the Class Indicator, in the Right Panel.

Select Heatmap Palette – Red/Green **Palette->Red-Green**

This selection selects the traditional colors of Red and Green for the heatmap. Red is positive or Correlated and Green is negative or Anti-Correlated.

Select Heatmap Palette – Yellow/Blue **Palete-> Yellow -Blue (Default)**

This selection selects the traditional colors of Yellow and Blue for the heatmap. Yellow is positive or Correlated and Blue is negative or Anti-Correlated.

Select Heatmap Palette – Gray Scale **Palette->Gray Scale**

This selection selects shades of gray for the heatmap. Light Gray is positive or Correlated and Dark Gray is negative or Anti-Correlated.

Show Palette **Palette->Show Palette**

This selection Launches the Palette Window, which shows the color palette used by the heatmap.

Show Rows Normalized **Normalize->Norm**

This selection changes the heatmap to show each row normalized. That means that the color corresponding to the highest value in the palette will be applied to the highest value in the row, the color corresponding to the lowest value in the palette will be applied to the lowest value in the row, and all other values will be scaled accordingly.

Show Rows Un-Normalized **Normalize->NoNorm**

This selection displays the traditional heatmap display where color selections from the palette are determine from the highest and lowest value among the data (not just in a row).

Url Targets for Row Fields Interface **File->Preferences->Field Info Urls**

This selection launches the preferences window for Field Info Target Url.

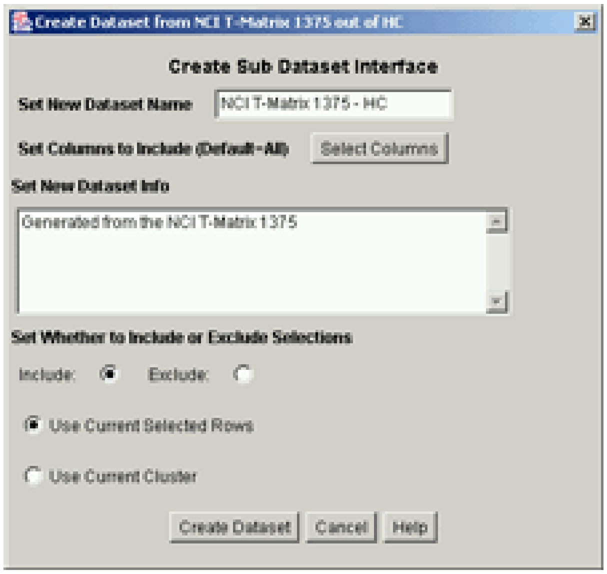
Interface - Create Dataset

Description

This Interface is used to create a new Dataset by using a cluster or selected text within a cluster for defining either the rows in the new dataset, or the rows to be excluded in the new Dataset.

Set New Dataset Name:

This text field is the name that the new Dataset will receive. It is seeded with the result Dataset Name along with '- HC' to indicate that it was created using the results from a Hierarchical Clustering result screen. That name can bet set to anything, as long as the name is not already in the system.



Create Dataset from NCI T-Matrix 1375 out of HC
Create Sub Dataset Interface
Select Columns
Create Dataset
Cancel
Help

Set New Dataset Name: NCI T-Matrix 1375 - HC

Set Columns to Include (Default=All)

Set New Dataset Info: Generated from the NCI T-Matrix 1375

Set Whether to Include or Exclude Selections

Include: ● Exclude: ○

● Use Current Selected Rows

○ Use Current Cluster

Create Dataset Settings Window

Select Columns to Include (Default=All)

The button (Select Columns) launches a window that allows for selection of the columns to be included in the new Dataset. If no column selection is made then all columns will be used.

Set New Dataset Info:

This text field is the Dataset information field. It is seeded from the result Dataset Information. Additional information should be added, here to provide a pedigree for the new Dataset.

Set Whether to Include or Exclude Selections

When Include is selected, only the rows designated will be used in the new Dataset. When Exclude is selected, the rows designated will be subtracted from the rows that made it through the filtering process (if any), to create a new Dataset.

Use Current Selected Rows and Use Current Cluster Radio Buttons

These radio selections only show up if rows are selected in the right panel, otherwise the entire cluster selected in the Left Panel will be used. If ‘Use Current Selected Rows’ is selected, then the rows selected in the Right Panel Heatmap will be used. If ‘Use Current Cluster’ is selected, then the rows contained in the currently selected cluster will be used.

Button - Create Dataset

This button creates the new Dataset.

Button - Cancel

This button closes the interface window without any further action.

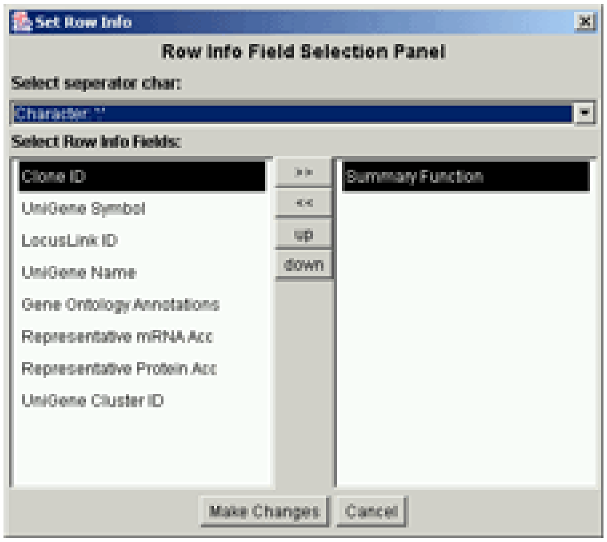
Preferences - Row Info Display

Description

The Row Info Display Preferences Window is used to specify the Row Information fields that will be displayed to the right of the Heatmap. In addition to selecting the fields, the field order can be specified along with the character seperating the contents of each field. Fields are selected for inclusion by moving them from the left column (Available) over to the right column (Included). The order of display is set by their row order, with the first row being the first display field.

Select separator char

This selection sets the character that will separate the information from each field.



Row Info Field Selection Panel

Select Row Info Fields:

This section allows for selection of the specific fields that will be displayed in the row info and the order of these fields. It has the following buttons:

>> - This button moves the currently hilited field in the left column over to the right column.
<< - This button moves the currently hilited field in the right column over to the left column.
up - This button moves the currently hilited field in the right column up one row.
down - This button moves the currently hilited field in the right column down one row.

Button - Make Changes
This button must be clicked in order for the changes to be made.

Button - Cancel
This button cancels any changes that were made and closes the window.

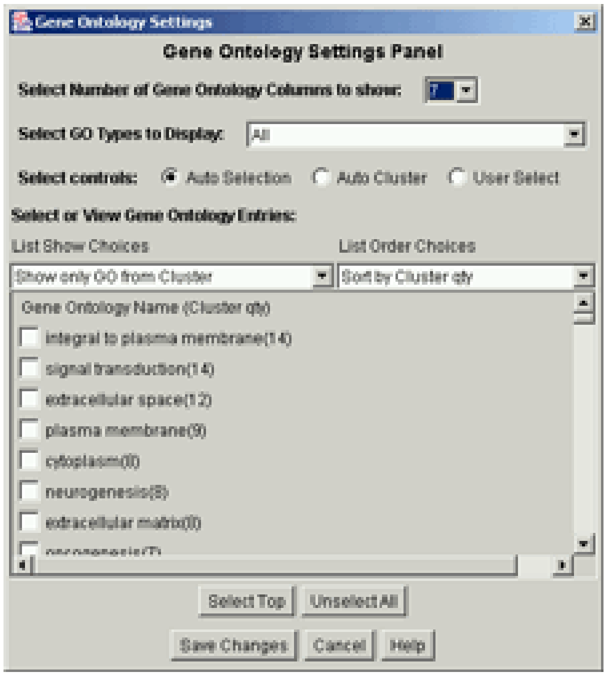
Preferences - Gene Ontology

Description

This interface is used to set the Gene Ontology preferences for this screen. The choices include automatic selection of the most common Gene Ontology categories based upon either the currently selected rows or for all rows. If the selected rows have preference and there aren't enough categories to fill the number of columns, then the most common categories across all rows are used to fill in the categories. Specific Gene Ontology categories can also be specified, this is aided by the List Show Choices and List Order Choices for quickly finding specific Gene Ontologies.

Select Number of Gene Ontology Columns to show

This selection sets the number of columns that will be displayed.



Gene Ontology Settings

Gene Ontology Settings Panel

Select Number of Gene Ontology Columns to show: 7

Select GO Types to Display: All

Select controls: ● Auto Selection ○ Auto Cluster ○ User Select

Select or View Gene Ontology Entries:

List Show Choices: Show only GO from Cluster

List Order Choices: Sort by Cluster qty

Gene Ontology Name (Cluster qty)

□ integral to plasma membrane(14)
□ signal transduction(14)
□ extracellular space(12)
□ plasma membrane(9)
□ cytoplasm(8)
□ neurogenesis(8)
□ extracellular matrix(8)

Gene Ontology Settings Panel

Select GO Types in Display

This selection sets the Gene Ontology Categories to Display with the following options:

Biological Process - Shows only Biological Process
Cellular Component - Shows only Cellular Component
Biological Process & Cellular Component - Shows both Categories
Molecular Function - Shows only Molecular Function
Biological Process & Molecular Function - Shows both Categories
Cellular Component & Molecular Function - Shows both Categories
All - Shows all three Categories

Gene Ontology Preferences

Select control for Gene Ontology Columns

This selection sets the way that Gene Ontology Columns are selected with the following options:

Auto Selection - Chooses most common Gene Ontologies among selected rows (if avail) then fills in remaining, if necessary, from most common among remaining rows.
Auto Set - Chooses most common Gene Ontologies among displayed rows.
User Select - Displays rows that the user has specifically selected. This gets turned off when the user clicks on the Gene Ontology Names.

Select or View Gene Ontology Entries

This section is for viewing or selecting specific Gene Ontologies by checking the checkbox either manually or using the 'Select Top' button under the list.

List Show Choices

This selection sets the Gene Ontologies that will be displayed in the list for selection. It has the following choices:

Show All GO - Displays all of the Gene Ontologies that are associated with the current Dataset.
Show only GO from Set - Displays the Gene Ontologies that are associated with the currently displayed genes.
Show only GO from Selected - Displays only the Gene Ontologies that are associated with the currently selected genes.

List Order Choices

This selection sets the order that the Gene Ontologies are displayed in the list for selection. It has the following choices:

Sort by name - Sorts the Gene Ontologies by their name.
Sort by Set qty - Sorts by the number of times that each Gene Ontology is associated to a gene in the currently displayed genes.
Sort by Selected qty - Sorts by the number of times that each Gene Ontology is associated to a gene in the currently selected genes.

Button - Select Top

This button selects the Gene Ontologies that are at the top of the list, in the quantity specified.

Button - Unselect All

This button unselects all of the currently selected Gene Ontologies.

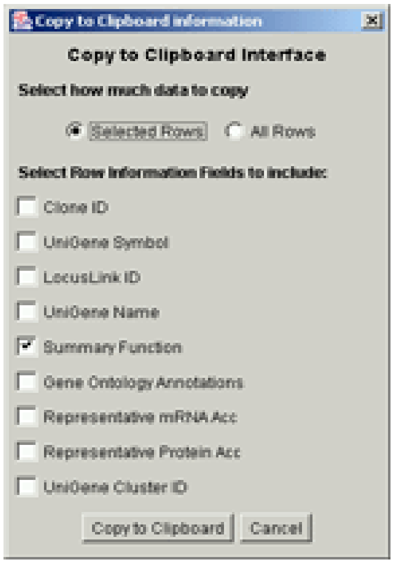
Button - Save Changes

This button sets the currently selected settings and closes the window.

Button - Cancel

This button closes the window without setting any of the changes made.

Interface - Copy to Clipboard



Copy to Clipboard information

Copy to Clipboard Interface

Select how much data to copy

● Selected Rows ○ All Rows

Select Row Information Fields to include:

☐ Clone ID
☐ UniGene Symbol
☐ LocusLink ID
☐ UniGene Name
☑ Summary Function
☐ Gene Ontology Annotations
☐ Representative mRNA Acc
☐ Representative Protein Acc
☐ UniGene Cluster ID

Copy to Clipboard Interface

Description

This Interface allows the user to select which Row Information Fields to include in the Copy to Clipboard and to specify the rows that they want included in the selection.

Select how much data to copy

This section determines which rows will be selected. If ‘All Rows’ are selected, then all of the rows in the left panel are copied to the clipboard. If ‘Selected Rows’ are selected: If rows are selected in the right panel, only those rows will be included. If rows are displayed in the right panel, but no rows are selected, all rows displayed in the right panel will be included. If no rows are selected in the Left Panel, then all rows will be included.

Select Row Information Fields to include

This section allows the user to specify which rows they’d like to include in the Copy to Clipboard. Each Row Information Field that is checked will be included in the copy.

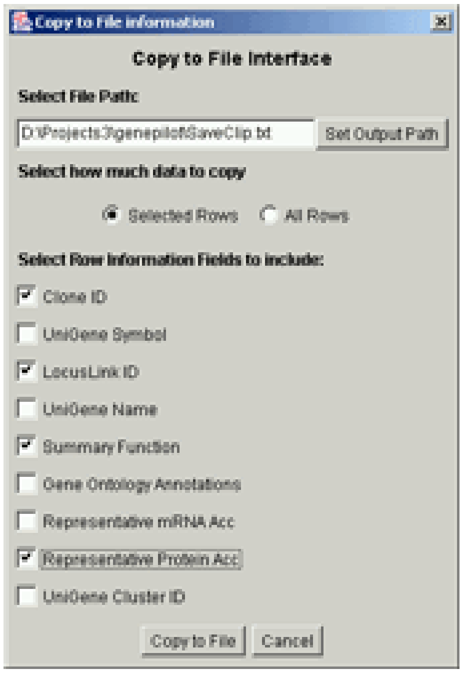
Button - Copy to Clipboard

This Button Initializes the copy and closes the window

Button - Cancel

This Button closes the window without doing the copy.

Interface - Copy to File



Copy to File information

Copy to File Interface

Select File Path:

D:\Projects3\genepilot\SaveClip.txt

Select how much data to copy

● Selected Rows ○ All Rows

Select Row Information Fields to include:

☑ Clone ID
☐ UniGene Symbol
☑ LocusLink ID
☐ UniGene Name
☑ Summary Function
☐ Gene Ontology Annotations
☐ Representative mRNA Acc
☑ Representative Protein Acc
☐ UniGene Cluster ID

Copy to File Interface

Description

This Interface allows the user to select specify a File Path in which to write out information contained in this result. This information will include the rows specified and the Row Information Fields that are chosen.

Select File Path

This selection is used to set the path to the output file. Click on the ‘Set Output Path’ to set the file. Traditional copy and paste methods work in this field.

Select how much data to copy
This section determines which rows will be selected. If 'All Rows' are selected, then all of the rows in the left panel are copied to the file. If 'Selected Rows' are selected: If rows are selected in the right panel, only those rows will be included. If rows are displayed in the right panel, but no rows are selected, all rows displayed in the right panel will be included. If no rows are selected in the Left Panel, then all rows will be included.

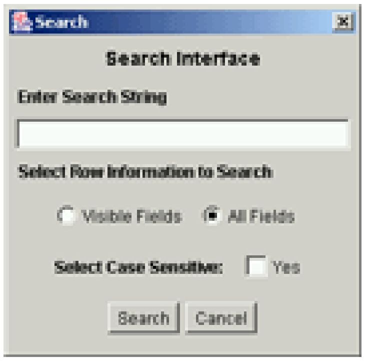
Select Row Information Fields to include
This section allows the user to specify which rows they'd like to include in the Copy to File . Each Row Information Field that is checked will be included in the copy.

Button - Copy to File
This Button Initializes the copy and closes the window

Button - Cancel
This Button closes the window without doing the copy.

Interface - Search

Description
This interface is used to set the search term and the settings for the search. The search can be restricted to just the visible Row Information or to all Row Information. It can also be set to be case sensitive.



Enter Search String

Select Row Information to Search

○ Visible Fields ● All Fields

Select Case Sensitive: □ Yes

Search Interface

Enter Search String
The search string is entered into this field. Traditional copy and paste methods work on this field.

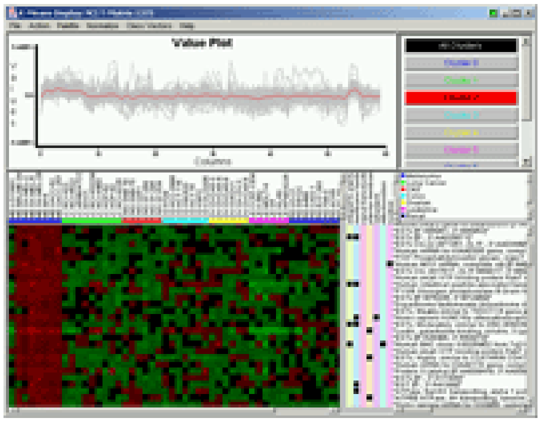
Select Row Information to Search
Selecting 'Visible Fields' restricts the search to just the information that is visible, as set in the Preferences->Row Info Display. Selecting All Fields causes the search to include all Row Info Fields, whether visible or not.

Select Case Sensitive
Checking this box will cause the search to be case sensitive.

Button - Search
This Button activates the search.

Button - Cancel
This button closes the interface without searching.

Chapter 5

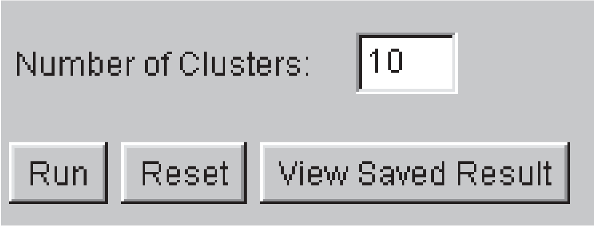


K-Means Clustering

Description

K-Means Clustering is a commonly used unsupervised method for analyzing MicroArray Data. It is a top-down Clustering method, that starts out by randomly distributing the data rows among a preset number of clusters. The average vector of each cluster is then calculated, then all of the rows are then moved to the cluster whose average vector it is closest to. This process is then repeated until either all rows have stopped changing clusters or a few rows are changing back and forth, between clusters, in a cyclic manner. The Vectors for these Clusters is then displayed in the top of the result screen, allowing the user to see the vectors for all of the clusters. Upon selection of a cluster, the vectors for all of it's member rows will be displayed in gray, and the cluster average vector will be displayed in it's assigned color. In the bottom, the heatmap and related information will be displayed, so that the member rows and information can be examined. For more information on K-Means Clustering, read more about it in 'Cluster Analysis for Large Scale Gene Expression Studies' (Reference 5 - A. Sturn).

Run Settings


Number of Clusters: 10

Description

Run settings provide a means to effect the number of clusters that are used.

Number of Clusters

This selection determines the number of clusters to be used. .

Button – Run

This button runs the K-Means Clustering analytic program

Button – Reset

This button resets the values to the last saved values.

Button – View Saved Result

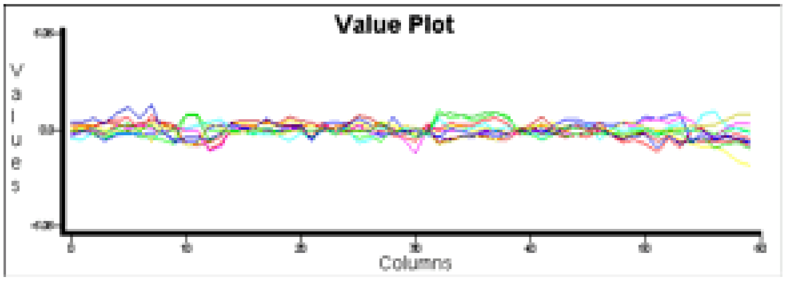
This button is only present when a previous result has been saved. Clicking on this button will launch the saved result.

Results - Parts – Graph and Button Panel

Description

The Graph panel contains a graph of either the average vectors of all of the clusters or of all of the cluster rows along with the cluster average vector, depending on whether the 'All Clusters' button is selected or a specific cluster button is selected. Along the x-axis the column number is represented, which corresponds with the column order as displayed in the heatmap in the lower panel. Along the y-axis, the value is represented.

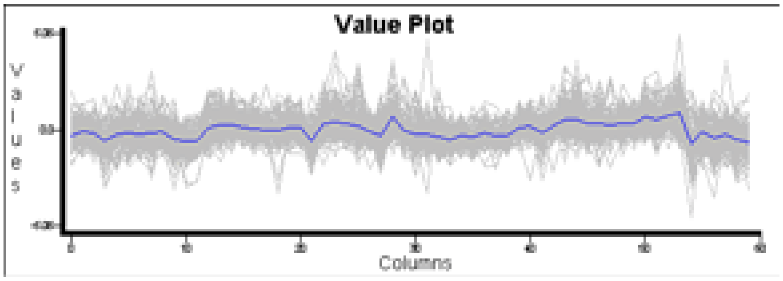
Graph – All Clusters



K-Means Graph - All Clusters displayed

When 'All Clusters' is selected, the graph will contain different colored lines representing the average vector of each cluster. The color of each line corresponds to the color on the cluster buttons on the right.

Graph – Cluster n

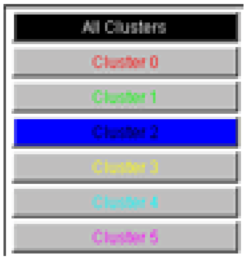


K-Means Graph: Cluster view

When a cluster is selected, in the button panel, to the right of the graph, the graph will contain a set of gray lines, representing the vectors of each row in that cluster. On top of all those lines, will be a line representing the average vector of that cluster, in the color that represents that cluster.

Button Panel

To the right of the Graph is the Button Panel. This panel will contain a button at the top, named ‘All Clusters’, then buttons below it with the label of ‘Cluster n’, where n is 1 to the number of clusters.

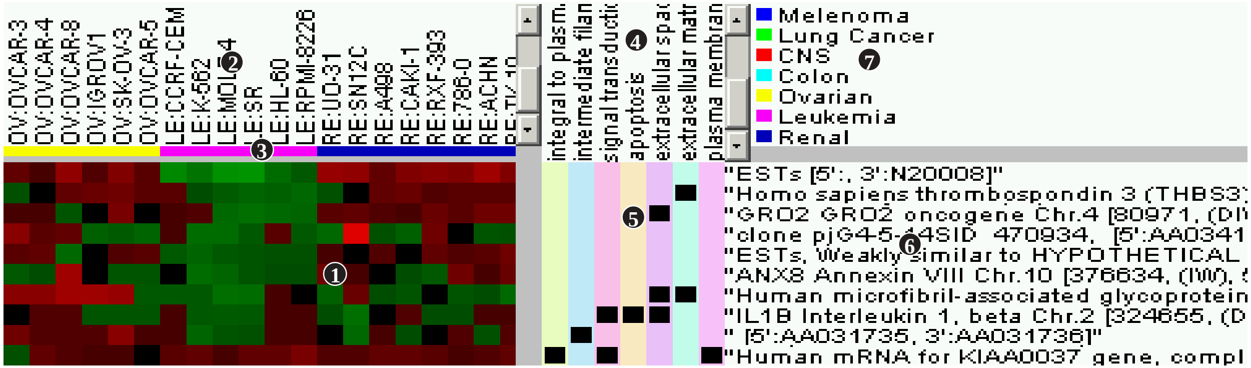


K-Means Button Panel: Cluster 2 Currently selected.

Actions:

1. Clicking on the ‘All Clusters’ button, in the Button Panel will display the average vector for each clusters in the graph, and will display the heatmap representing the average vector, in the heatmap panel, below.
2. Clicking on a cluster button, in the Button Panel will display the vectors for all of the rows in a cluster, along with the average vector for the cluster, in the graph. It will also result in the heatmap for the cluster rows to be displayed in the heatmap panel, below.

Results - Parts – Heatmap Panel



"ESTs [5':, 3':N20008]"
"Homo sapiens thrombospondin 3 (THBS3)
"GRO2 GRO2 oncogene Chr.4 [80971, (DI
"clone pjG4-5-44SID 470934, [5':AA0341
"ESTs, Weakly similar to HYPOTHETICAL
"ANX8 Annexin VIII Chr.10 [376634, (IW),
"Human microfibril-associated glycoprotein
"IL1B Interleukin 1, beta Chr.2 [324655, (D
" [5':AA031735, 3':AA031736]"
"Human mRNA for KIAA0037 gene, compl

Dataset Display Heatmap: 1-Heatmap, 2-Column Information, 3-Class Indication (when avail), 4-Gene Ontology Names (when avail), 5-Gene Ontology Indication (when avail), 6-Row Information, 7-Class Information (when avail)

Description

This panel contains information about the current rows, either in the current cluster or about the average heatmap for each cluster. The central part is the Heatmap, which contains colors from the Palette to indicate values of each row (or cluster). When Classification Vector(s) are available a Class Indication row which, indicates the classes of each column, above the Average Heatmap. At the top is the Column Info, in the form of text rotated counter-clockwise by 90 degrees. To the right of the Heatmap, is the Gene Ontology indication when Gene Ontology information has been imported into GenePilot for this Dataset. To the right is the Row Information, which contains the description of the row. In the top right corner, the legend for the Class Indication will define the colors for each class, when Classification Vector(s) are available. When 'All Clusters' is selected, this Panel will contain information about the average vectors of each cluster. When a cluster is selected, then this panel will contain information about each row in that cluster, sorted in the order of the closest correlated for (to the cluster average vector) to the furthest.

Heatmap – Bottom Left

The heatmap contains rectangles color-coded in the Palette colors to represent the value of each cel (juncture of 1 row and 1 column). By default, these colors are Red (highest value) and Green (lowest value). When No-Normalize is selected, the value range is defined by the min and max values in the entire Dataset (Rows remaining after filtering), with zero being the medium value. When Normalize is selected, then the value range for each row is defined by the min and max values in that row, with the mean value being the medium value.

Class Indication – Above Average Heatmap

If there is Classification Vector(s), there will be Class Indicators above the Heatmap. This is a row of rectangles, which are color-coded to indicate the class of each column as defined by the Class Information block in the Top Right corner.

Column Names – Upper Left
The Column Information contains the column names, rotated counter-clockwise by 90 degrees, above each column in the Heatmap.

Gene Ontology Names - Middle Right (If Gene Ontology Information)
The Gene Ontology Names is a variable-column display that displays the Gene Ontology Names that are currently chosen either automatically (Most common in dataset or most common in selected rows) or by the User. Using File->Preferences->Gene Ontology Display, these different choices can be set. If specific Gene Ontologies are selected by the user, the settings can be easily changed back to automatic selection by clicking on this (Gene Ontology Names) display.

Gene Ontology Indication - Bottom Middle
The Gene Ontology Indication indicates the Gene Ontologies that are associated with each row. The Gene Ontologies are chosen either automatically (Most common in dataset or most common in selected rows) or by the User. Using File->Preferences->Gene Ontology Display, these different choices can be set.

Row Information – Bottom Right
Row information contains the row name and row information to the right of each row in the heatmap.

Class Information – Top Right (If Class Vector)
If there is Classification Vector(s), there will be Class Information which defines the color coding for the Class Indication row.

Actions:

1. Clicking and dragging on the bitmap will select those rows.
2. Holding down the shift key and selecting a row will extend the row from the previous selection down to the selected row. This can be used to select rows in a group that is larger than those seen on the screen.
3. When user selected gene ontologies are displayed, the Gene Ontology displays can be reverted back to automatic mode by clicking on the Gene Ontology Names.

Results - Menu Choices

Bitmap from All **Action->Save All to Bitmap**
This selection creates a gif bitmap containing all of the information in the Lower Panel.

Bitmap from Graph **Action->Graph to Bitmap**
This selection saves the current Graph to a gif bitmap.

K-Means Clustering Menu Choices

Bitmap from Selected **Action->Save Sel to Bitmap**
This selection creates a gif bitmap containing all of the information in the Lower Panel. If 'All Clusters' are selected, the bitmap will contain the Average Vectors of the clusters. If a cluster is selected and rows are selected, then only the selected rows will be included in the bitmap. If a cluster is selected and no rows are selected, the entire contents of the lower panel will be included in the bitmap.

Copy to Clip **Action->Copy to Clip**
This selection launches an interface for copying all or selected rows to the clipboard. This interface also allows for selecting the Row Information Fields which the user is interested in.

Copy to File **Action->Copy to File**
This selection launches an interface for copying all or selected rows to a file. This interface also allows for selecting the Row Information Fields which the user is interested in.

Create Vector **Action->Create Vector**
This selection launches the Create Vector dialog box along with the mean vector of the currently selected rows.

Gene Ontology Display **File->Preferences->Gene Ontology Display**
This selection launches the preferences window for Gene Ontologies. Within this preference window the following can be set; set number of columns, set automatic preferences for selection of Gene Ontologies, select specific Gene Ontologies to display.

Generate Gene Ontology Statistics **Action->Generate GO Statistics**
This selection launches an html page that contains the statistics for the Gene Ontologies for the rows in the following order of precedence: If in Cluster; 1. Selected Rows, 2. All Cluster Rows. If in All Clusters; 1. All Rows in Selected Clusters, 2. All Rows in all Clusters.

Launch Info Page **Action->Launch Info Page**
This selection launches an information page containing a matrix of the genes vs. Gene Ontologies (if available) and a listing of all of the Row Information Fields along with links to websites which have further information.

Make Dataset **Action->Make Sub-Dataset**
This selection launches and interface that let's the user define criteria for creating a new Dataset from rows of the current Dataset. For K-Means Clustering, one or more clusters can be selected, with their rows to be used as the rows for the new Dataset, or to be excluded from the new Dataset. **Note** Rows that have been filtered out will be excluded from any new Dataset created this way.

Quit **File->Quit**
This selection closes the Result Window.

K-Means Clustering Menu Choices

Save Results **File->Save Results**

This selection is used to save the current result. This will save the results of an Analytic run so that the results can be viewed immediately. Once an analytic result is saved, for a Dataset, a button will be added to the Dataset Information screen to quickly view the saved result.

Search **Action->Search**

This selection launches the search interface, which allows the user to search the rows for a String. See 'Search Interface' for more information on the Search Interface. For K-Means Clustering, the search will search the currently selected cluster first, it will then continue the search through the remaining clusters, then start over at the beginning cluster.

Search Again **Action->Search Again**

This selection searches for the next case of the search string using the settings from the last search.

Select Classification Vector (If Avail) **Classes-> 'Class Vector Name'**

If there is one or more Classification Vectors, then one of those vectors can be selected for class indication of columns in either the Column Dendigram (if column clustering was selected) or the Class Indicator, in the Right Panel.

Select Heatmap Palette – Red/Green **Palette->Red-Green**

This selection selects the traditional colors of Red and Green for the heatmap. Red is positive or Correlated and Green is negative or Anti-Correlated.

Select Heatmap Palette – Yellow/Blue **Palete-> Yellow -Blue (Default)**

This selection selects the traditional colors of Yellow and Blue for the heatmap. Yellow is positive or Correlated and Blue is negative or Anti-Correlated.

Select Heatmap Palette – Gray Scale **Palette->Gray Scale**

This selection selects shades of gray for the heatmap. Light Gray is positive or Correlated and Dark Gray is negative or Anti-Correlated.

Show Palette **Palette->Show Palette**

This selection Launches the Palette Window, which shows the color palette used by the heatmap.

Show Rows Normalized **Normalize->Norm**

This selection changes the heatmap to show each row normalized. That means that the color corresponding to the highest value in the palette will be applied to the highest value in the row, the color corresponding to the lowest value in the palette will be applied to the lowest value in the row, and all other values will be scaled accordingly.

Show Rows Un-Normalized **Normalize->NoNorm**

This selection displays the traditional heatmap display where color selections from the palette are determine from the highest and lowest value among the data (not just in a row).

Url Targets for Row Fields Interface **File->Preferences->Field Info Urls**

This selection launches the preferences window for Field Info Target Url.

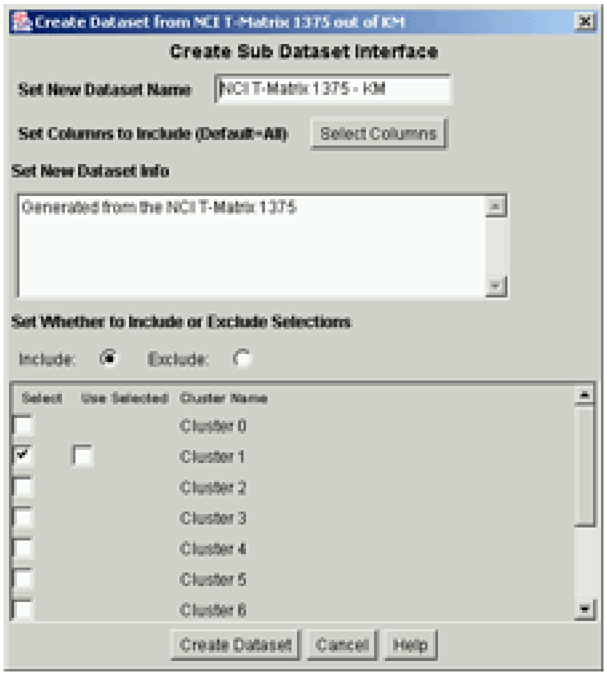
Windows - Create Dataset

Description

This Interface is used to create a new Dataset by using one or more clusters for defining either the rows in the new dataset, or the rows to be excluded in the new Dataset. In the current cluster, the selected rows can be used to further refine the selection

Set New Dataset Name:

This text field is the name that the new Dataset will receive. It is seeded with the result Dataset Name along with '- KM' to indicate that it was created using the results from a K-Means Clustering result screen. That name can bet set to anything, as long as the name is not already in the system.



Create Dataset from NCI T-Matrix 1375 out of KM

Create Sub Dataset Interface

Set New Dataset Name: NCI T-Matrix 1375 - KM

Set Columns to Include (Default=All)

Set New Dataset Info

Generated from the NCI T-Matrix 1375

Set Whether to Include or Exclude Selections

Include: ● Exclude: ○

Select	Use Selected	Cluster Name
☐		Cluster 0
☑	☐	Cluster 1
☐		Cluster 2
☐		Cluster 3
☐		Cluster 4
☐		Cluster 5
☐		Cluster 6

Create Sub Dataset Interface

Select Columns to Include (Default=All)

The button (Select Columns) launches a window that allows for selection of the columns to be included in the new Dataset. If no column selection is made then all columns will be used.

Set New Dataset Info:

This text field is the Dataset information field. It is seeded from the result Dataset Information. Additional information should be added, here to provide a pedigree for the new Dataset.

Set Whether to Include or Exclude Selections

When Include is selected, only the rows designated will be used in the new Dataset. When Exclude is selected, the rows designated will be subtracted from the rows that made it through the filtering process (if any), to create a new Dataset.

Select Check Boxes

These check boxes give the option to select each cluster to be used for defining the rows to be included or excluded.

Use Selected Check Box

This check box is only available for the current cluster. When checked, only the selected rows will be used. It is only shown when rows are selected in the heatmap.

Button – Select All

This button checks all of the Select Check Boxes, making it more convenient to select most of the clusters.

Button – Clear All

This button clears all of the Select Check Boxes, making it more convenient to turn off selection of most or all of the clusters.

Button – Create Dataset

This button creates the Dataset, once all of the settings are made.

Button – Cancel

This button closes this interface without creating the new Dataset.

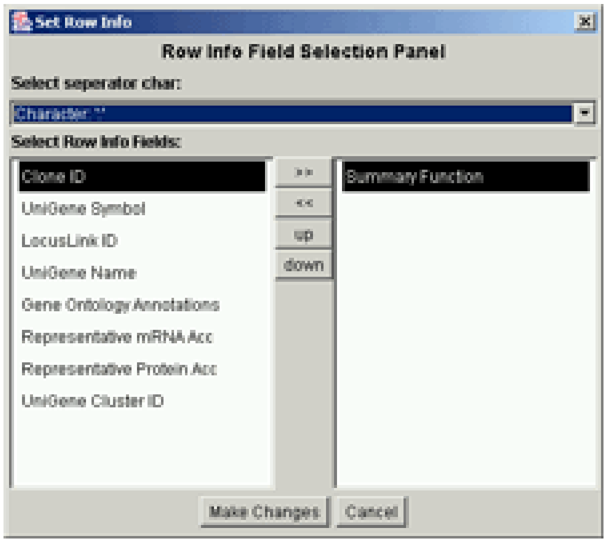
Preferences - Row Info Display

Description

The Row Info Display Preferences Window is used to specify the Row Information fields that will be displayed to the right of the Heatmap. In addition to selecting the fields, the field order can be specified along with the character seperating the contents of each field. Fields are selected for inclusion by moving them from the left column (Available) over to the right column (Included). The order of display is set by their row order, with the first row being the first display field.

Select separator char

This selection sets the character that will separate the information from each field.



Select seperator char: Character: ' '

Row Info Field Selection Panel

Select Row Info Fields:

This section allows for selection of the specific fields that will be displayed in the row info and the order of these fields. It has the following buttons:

>> - This button moves the currently hilited field in the left column over to the right column.
<< - This button moves the currently hilited field in the right column over to the left column.
up - This button moves the currently hilited field in the right column up one row.
down - This button moves the currently hilited field in the right column down one row.

Button - Make Changes

This button must be clicked in order for the changes to be made.

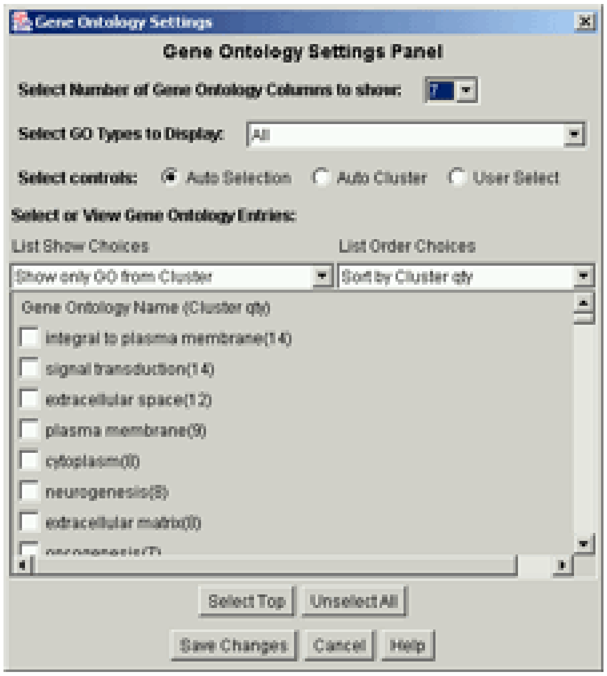
Button - Cancel

This button cancels any changes that were made and closes the window.

Preferences - Gene Ontology

Description

This interface is used to set the Gene Ontology preferences for this screen. The choices include automatic selection of the most common Gene Ontology categories based upon either the currently selected rows or for all rows in the current cluster. If the selected rows have preference and there aren't enough categories to fill the number of columns, then the most common categories across all rows in the cluster are used to fill in the categories. Specific Gene Ontology categories can also be specified, this is aided by the List Show Choices and List Order Choices for quickly finding specific Gene Ontologies.


Gene Ontology Settings

Gene Ontology Settings Panel

Select Number of Gene Ontology Columns to show: 7

Select GO Types to Display: All

Select controls: ● Auto Selection ○ Auto Cluster ○ User Select

Select or View Gene Ontology Entries:

List Show Choices: Show only GO from Cluster

List Order Choices: Sort by Cluster qty

Gene Ontology Name (Cluster qty)

- □ integral to plasma membrane(14)
- □ signal transduction(14)
- □ extracellular space(12)
- □ plasma membrane(9)
- □ cytoplasm(8)
- □ neurogenesis(8)
- □ extracellular matrix(8)
- □ oncogenesis(7)

Gene Ontology Settings Panel

Select Number of Gene Ontology Columns to show

This selection sets the number of columns that will be displayed.

Select GO Types in Display

This selection sets the Gene Ontology Categories to Display with the following options:

Biological Process - Shows only Biological Process
Cellular Component - Shows only Cellular Component
Biological Process & Cellular Component - Shows both Categories
Molecular Function - Shows only Molecular Function
Biological Process & Molecular Function - Shows both Categories
Cellular Component & Molecular Function - Shows both Categories
All - Shows all three Categories

Select control for Gene Ontology Columns

This selection sets the way that Gene Ontology Columns are selected with the following options:

Auto Selection - Chooses most common Gene Ontologies among selected rows (if avail) then fills in remaining, if necessary, from most common among remaining rows in the cluster.

Auto Set - Chooses most common Gene Ontologies among the rows of the current cluster.

User Select - Displays rows that the user has specifically selected. This gets turned off when the user clicks on the Gene Ontology Names.

Select or View Gene Ontology Entries

This section is for viewing or selecting specific Gene Ontologies by checking the checkbox either manually or using the 'Select Top' button under the list.

List Show Choices

This selection sets the Gene Ontologies that will be displayed in the list for selection. It has the following choices:

Show All GO - Displays all of the Gene Ontologies that are associated with the current Dataset.

Show only GO from Cluster- Displays the Gene Ontologies that are associated with genes from the current cluster..

Show only GO from Selected - Displays only the Gene Ontologies that are associated with the currently selected genes.

List Order Choices

This selection sets the order that the Gene Ontologies are displayed in the list for selection. It has the following choices:

Sort by name - Sorts the Gene Ontologies by their name.

Sort by Cluster qty - Sorts by the number of times that each Gene Ontology is associated to a gene in the currently cluster.

Sort by Selected qty - Sorts by the number of times that each Gene Ontology is associated to a gene in the currently selected genes.

Button - Select Top

This button selects the Gene Ontologies that are at the top of the list, in the quantity specified.

Button - Unselect All

This button unselects all of the currently selected Gene Ontologies.

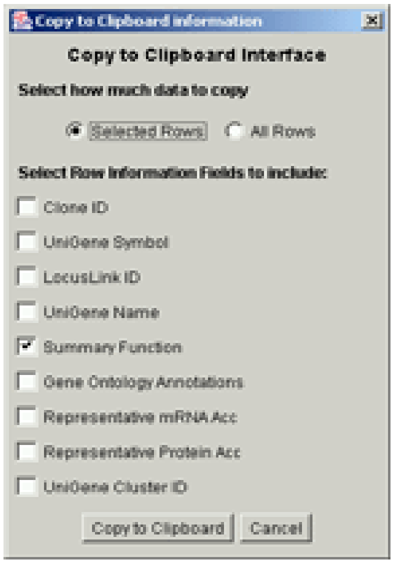
Button - Save Changes

This button sets the currently selected settings and closes the window.

Button - Cancel

This button closes the window without setting any of the changes made.

Interface - Copy to Clipboard



Copy to Clipboard Interface

Select how much data to copy

● Selected Rows ○ All Rows

Select Row Information Fields to include:

☐ Clone ID
☐ UniGene Symbol
☐ LocusLink ID
☐ UniGene Name
☑ Summary Function
☐ Gene Ontology Annotations
☐ Representative mRNA Acc
☐ Representative Protein Acc
☐ UniGene Cluster ID

Copy to Clipboard Interface

Description

This Interface allows the user to select which Row Information Fields to include in the Copy to Clipboard and to specify the rows that they want included in the selection.

Select how much data to copy

This section determines which rows will be selected. When ‘All Rows’ is selected, then all of the rows in the current cluster are copied to the clipboard, or all clusters (along with their rows) are copied if no cluster is selected. When ‘Selected Rows’ is selected: If rows are selected in the Heatmap, only those rows will be included. If a cluster is selected but no rows are selected, all rows in that cluster will be included. If no cluster is selected then all clusters (along with their rows) will be copied.

Select Row Information Fields to include

This section allows the user to specify which rows they’d like to include in the Copy to Clipboard. Each Row Information Field that is checked will be included in the copy.

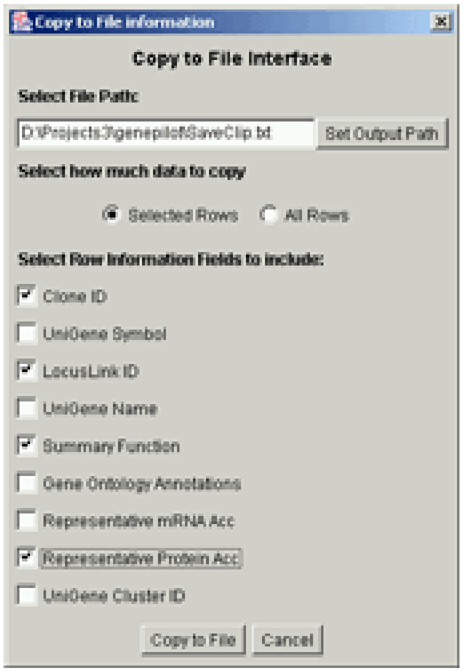
Button - Copy to Clipboard

This Button Initializes the copy and closes the window

Button - Cancel

This Button closes the window without doing the copy.

Interface - Copy to File



Copy to File Interface

Select File Path:

D:\Projects3\genepilot\SaveClip.txt

Select how much data to copy

● Selected Rows ○ All Rows

Select Row Information Fields to include:

☑ Clone ID
☐ UniGene Symbol
☑ LocusLink ID
☐ UniGene Name
☑ Summary Function
☐ Gene Ontology Annotations
☐ Representative mRNA Acc
☑ Representative Protein Acc
☐ UniGene Cluster ID

Copy to File Interface

Description

This Interface allows the user to select specify a File Path in which to write out information contained in this result. This information will include the rows specified and the Row Information Fields that are chosen.

Select File Path

This selection is used to set the path to the output file. Click on the ‘Set Output Path’ to set the file. Traditional copy and paste methods work in this field.

Select how much data to copy

This section determines which rows will be selected. When ‘All Rows’ is selected, then all of the rows in the current cluster are copied to the file, or all clusters (along with their rows) are copied if no cluster is selected. When ‘Selected Rows’ is selected: If rows are selected in the Heatmap, only those rows will be included. If a cluster is selected but no rows are selected, all rows in that cluster will be included. If no cluster is selected then all clusters (along with their rows) will be copied.

Select Row Information Fields to include

This section allows the user to specify which rows they’d like to include in the Copy to File . Each Row Information Field that is checked will be included in the copy.

Button - Copy to File

This Button Initializes the copy and closes the window

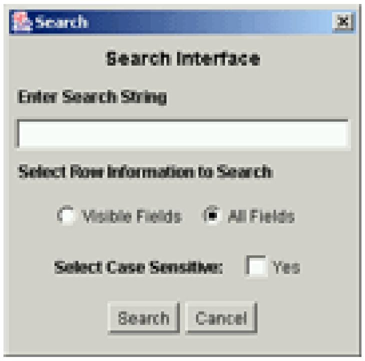
Button - Cancel

This Button closes the window without doing the copy.

Interface - Search

Description

This interface is used to set the search term and the settings for the search. The search can be restricted to just the visible Row Information or to all Row Information. It can also be set to be case sensitive. The search function remembers the last row that it searched in a cluster. When a new cluster is selected, that information is reset.



Search Interface

Enter Search String

Select Row Information to Search

○ Visible Fields ● All Fields

Select Case Sensitive: □ Yes

Search Interface

Enter Search String

The search string is entered into this field. Traditional copy and paste methods work on this field.

Select Row Information to Search

Selecting ‘Visible Fields’ restricts the search to just the information that is visible, as set in the Preferences->Row Info Display. Selecting All Fields causes the search to include all Row Info Fields, whether visible or not.

Select Case Sensitive

Checking this box will cause the search to be case sensitive.

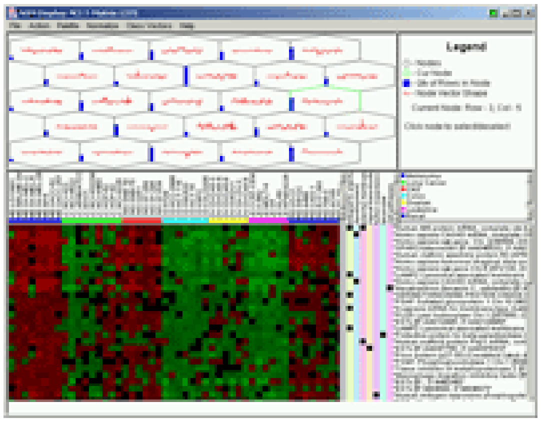
Button - Search

This Button activates the search.

Button - Cancel

This button closes the interface without searching.

Chapter 6

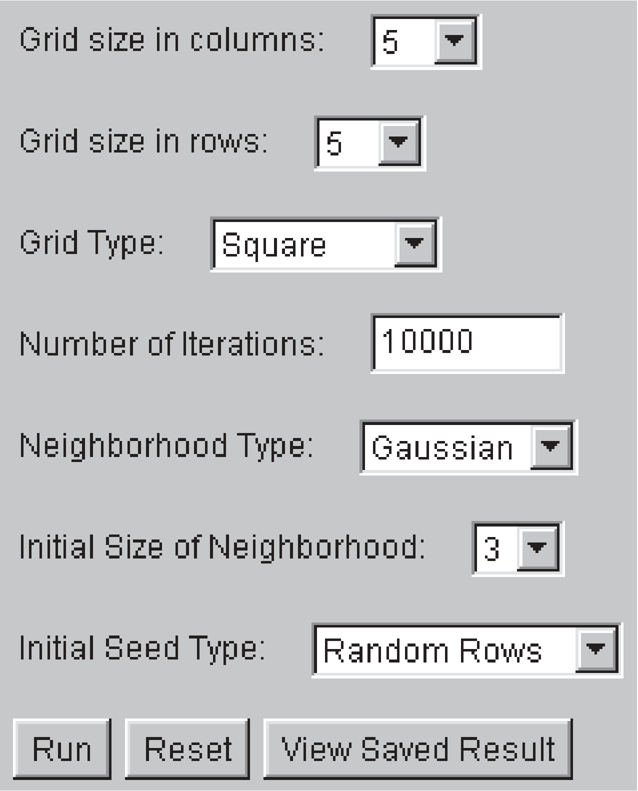


SOM

Description

SOM is a commonly used unsupervised method for analyzing MicroArray Data. It is a top-down Clustering method, with a grid of nodes (square or hexagonal) which influence the vector shape of the adjacent nodes which are in their ‘neighborhood’. The grid is initially ‘seeded’ with either random values or random rows. Then, successive rows are randomly selected and placed into the node to which their vector is closest. Upon the placement of the row vector into the node, the node vector is recalculated based upon the relationship of it’s vector to the row vector. Then, the adjacent node vectors are also recalculated, out to the extent of the current ‘neighborhood’ size. These ‘neighborhood’ nodes are influenced to a lesser degree than the central node, based upon their distance from the central node and the neighborhood type. A neighborhood is a decreasing (with time) sphere of influence that one node has on it’s neighbors when each new row is placed within. This process is then repeated for the number of iterations that are specified, with the neighborhood decreasing in size throughout the process until the final iterations when only the central node is influenced. In the result screen the nodes are displayed in their grid, in the top pane. Each node displays a small graph in the middle and a bar along the left side that indicates the number of rows that are represented by that Node. The heatmap representing the vectors for these nodes are displayed in the lower panel, until a node is select. Upon selection of a node, the heatmap and related information will be displayed, so that the member rows and information can be examined. For more information on SOM, read more about it in ‘Cluster Analysis for Large Scale Gene Expression Studies’ (Reference 5 - A. Sturn).

Run Settings



Grid size in columns: 5

Grid size in rows: 5

Grid Type: Square

Number of Iterations: 10000

Neighborhood Type: Gaussian

Initial Size of Neighborhood: 3

Initial Seed Type: Random Rows

SOM Run Settings Menu

Description

The Run Settings provide a means to alter the way that SOM analyzes the Data. It is recommended that the default settings are used, at first, then to try out the other settings to view other results. Because SOM is randomly seeded, initially, each run (even with the same settings) will result in unique results. It may therefore be useful to try several runs on a new Dataset to see what type of results can be yielded.

Grid size in columns

This selection determines the number node columns that will be used.

Grid size in rows

This selection determines the number of node rows that will be used.

Grid Type

This selection determines the type of grid that will be used:

1. Square – each node is in the shape of a square and has 8 adjacent nodes.
2. Hexagon – each node is in the shape of a hexagon and has 6 adjacent nodes.

Number of Iterations

This selection determines the number of times that a row vector is randomly selected, then fit into the grid. The number selected should be at least 10 times the number of rows.

Neighborhood Type

This determines the distribution of strength of the influence that a placement in a node has on it's neighbors:

1. Gausian – Linear weakening of influence with distance, from full strength on updated node, down to zero strength at current boundary of neighborhood.
2. Bubble – No weakening of influence within neighborhood.

Initial Size of Neighborhood

This determines the size of the neighborhood, at the start. With each cycle, the Neighborhood becomes marginally smaller, until the final placed nodes have no influence on the nodes around them.

Initial Seed Type

This determines how the vectors within the nodes are initially seeded:

1. Random Rows – Each node is seeded with a random row from the Dataset.
2. Random Values – After calculating the Mean Value and Standard Deviation, random values are calculated and placed into each position of the vector for each node.

Button – Run

This button runs the K-Means Clustering analytic program

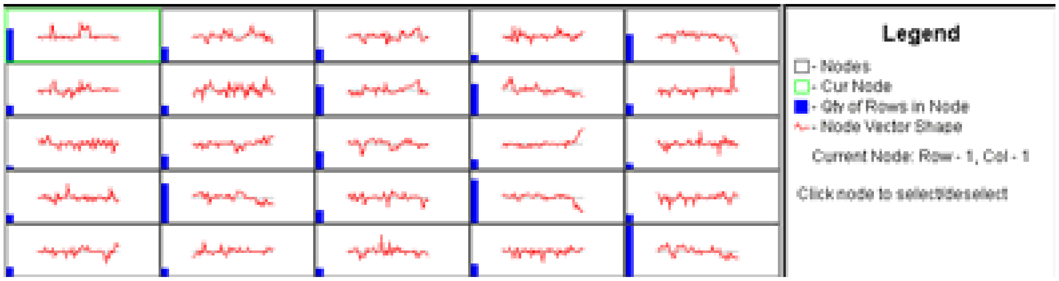
Button – Reset

This button resets the values to the last saved values.

Button – View Saved Result

This button is only present when a previous result has been saved. Clicking on this button will launch the saved result.

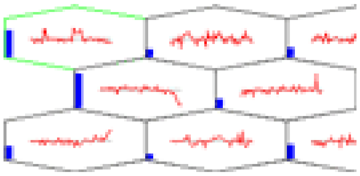
Results - Parts – Node Panel



SOM Node Panel: Square Grid (5 Columns X 5 Rows)

Description

The Node Panel contains the grid of nodes. Each node contains a small graph of the node vector, in the middle, and a bar along the left that indicates the number of member rows. In reference to a node, the first column is 0 and the first row is 0. To the right of the Nodes is the Legend, which gives information on the nodes and indicates which node is currently selected. The currently selected node will have it's borders colored in green.



SOM Hexagonal Nodes

Nodes – Left

The nodes are either square or Hexagonal, depending on the Grid Type selected. Within each node is a small graph in red, representing the node Vector. On the left side of each node is a blue bar, representing the number of rows associated with this node (this node having the vector which was closest to each row vector).

Legend – Right

The upper portion of the Legend serves as a legend for the node panel. In the lower part is text indicating to ‘Click node to select/deselect’, if no node is currently selected, or the coordinates of the node (0 based), if a node is selected.

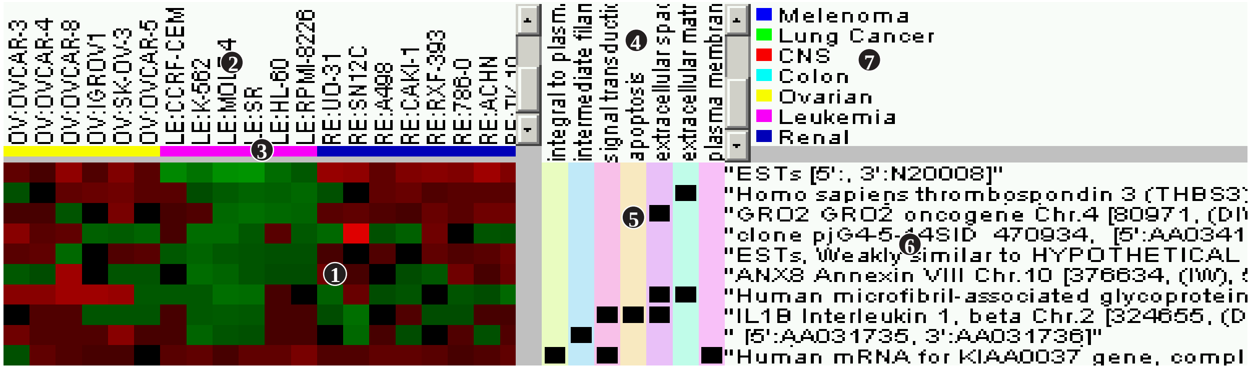
Actions:

1. Clicking on an unselected node will select that node as the current node.
2. Clicking on a selected node will unselect it.

Results - Parts – Heatmap Panel

Description

This panel contains the specific information, for the rows (or nodes). The central part is the Heatmap, which contains colors from the Palette to indicate values of each row (or cluster). To the top, of each column is the Column Info, in the form of text rotated counter-clockwise by 90 degrees. A row of rectangles which are color-coded to indicate the class of each column, will be in between the Heatmap and the Column Info, when Classification Vector(s) are available. To the right of the Heatmap is the Gene Ontology (when available) indications. To the far right is the Row Information, which contains the description of the row. In the top middle is the Gene Ontology descriptions (when available). In the top right corner, the legend for the Class Indication will define the colors for each class, when Classification Vector(s) are available. When no node is selected, this Panel will contain information about the node vector of each node. When a node is selected, then this panel will contain information about each member row for that node, sorted in the order of the closest correlated for (to the node vector) to the furthest.



"ESTs [5':, 3':N20008]"
"Homo sapiens thrombospondin 3 (THBS3)
"GRO2 GRO2 oncogene Chr.4 [80971, (DI
"clone pjG4-5-14SID 470934, [5':AA0341
"ESTs, Weakly similar to HYPOTHETICAL
"ANX8 Annexin VIII Chr.10 [376634, (IW), 5
"Human microfibril-associated glycoprotein
"IL1B Interleukin 1, beta Chr.2 [324655, (D
" [5':AA031735, 3':AA031736]"
"Human mRNA for KIAA0037 gene, compl

Dataset Display Heatmap: 1-Heatmap, 2-Column Information, 3-Class Indication (when avail), 4-Gene Ontology Names (when avail), 5-Gene Ontology Indication (when avail), 6-Row Information, 7-Class Information (when avail)

Heatmap – Bottom Left
The heatmap contains rectangles color-coded in the Palette colors to represent the value of each cel (juncture of 1 row and 1 column). By default, these colors are Red (highest value) and Green (lowest value). When No-Normalize is selected, the value range is defined by the min and max values in the entire Dataset (Rows remaining after filtering), with zero being the medium value. When Normalize is selected, then the value range for each row is defined by the min and max values in that row, with the mean value being the medium value.

Class Indication – Above Average Heatmap
If there is Classification Vector(s), there will be Class Indicators above the Heatmap. This is a row of rectangles, which are color-coded to indicate the class of each column as defined by the Class Information block in the Top Right corner.

Column Names – Upper Left
The Column Information contains the column names, rotated counter-clockwise by 90 degrees, above each column in the Heatmap.

Gene Ontology Names - Middle Right (If Gene Ontology Information)
The Gene Ontology Names is a variable-column display that displays the Gene Ontology Names that are currently chosen either automatically (Most common in dataset or most common in selected rows) or by the User. Using File->Preferences->Gene Ontology Display, these different choices can be set. If specific Gene Ontologies are selected by the user, the settings can be easily changed back to automatic selection by clicking on this (Gene Ontology Names) display.

Gene Ontology Indication - Bottom Middle
The Gene Ontology Indication indicates the Gene Ontologies that are associated with each row. The Gene Ontologies are chosen either automatically (Most common in dataset or most common in selected rows) or by the User. Using File->Preferences->Gene Ontology Display, these different choices can be set.
Row Information – Bottom Right
Row information contains the row name and row information to the right of each row in the heatmap

Row Information – Bottom Right
Row information contains the row name and row information to the right of each row in the heatmap.

Class Information – Top Right (If Class Vector)
If there is Classification Vector(s), there will be Class Information which defines the color coding for the Class Indication row.

Actions:
1. Clicking and dragging on the bitmap will select those rows.

2. Holding down the shift key and selecting a row will extend the row from the previous selection down to the selected row. This can be used to select rows in a group that is larger than those seen on the screen.
3. When user selected gene ontologies are displayed, the Gene Ontology displays can be reverted back to automatic mode by clicking on the Gene Ontology Names.

Results - Menu Choices

Bitmap from All **Action->Save All to Bitmap**
This selection creates a gif bitmap containing all of the information in the Lower Panel.

Bitmap from Grid **Action->Graph to Bitmap**
This selection saves the current node grid to a gif bitmap.

Bitmap from Selected **Action->Save Sel to Bitmap**
This selection creates a gif bitmap containing all of the information in the Lower Panel. If no nodes are selected, the bitmap will contain the node vector heatmap. If a node is selected and rows are selected, then only the selected rows will be included in the bitmap. If a node is selected and no rows are selected, the entire contents of the lower panel will be included in the bitmap.

Copy to Clip **Action->Copy to Clip**
This selection launches an interface for copying all or selected rows to the clipboard. This interface also allows for selecting the Row Information Fields which the user is interested in.

Copy to File **Action->Copy to File**
This selection launches an interface for copying all or selected rows to a file. This interface also allows for selecting the Row Information Fields which the user is interested in.

Create Vector **Action->Create Vector**
This selection launches the Create Vector dialog box along with the mean vector of the currently selected rows.

Gene Ontology Display **File->Preferences->Gene Ontology Display**
This selection launches the preferences window for Gene Ontologies. Within this preference window the following can be set; set number of columns, set automatic preferences for selection of Gene Ontologies, select specific Gene Ontologies to display.

Generate Gene Ontology Statistics **Action->Generate GO Statistics**
This selection launches an html page that contains the statistics for the Gene Ontologies for the rows in the following order of precedence: If in Node; 1. Selected Rows, 2. All NodeRows. If in All Node; 1. All Rows in Selected Nodes, 2. All Rows in all Nodes.

Launch Info Page **Action->Launch Info Page**
This selection launches an information page containing a matrix of the genes vs. Gene Ontologies (if available) and a listing of all of the Row Information Fields along with links to websites which have further information.

Copy to Clip – Selected **Action->Copy Sel to Clip**
This selection copies the currently selected Rows to the system Clipboard, along with row information and column headings. If no rows are selected, but a node is selected, then all of the rows in the current cluster are copied to the clipboard. If no nodes are selected, then the node Vector will be copied to the clipboard..

Copy to Clip – All **Action->Copy All to Clip**
This selection copies all Rows to the System Clipboard, sorted by Nodes.

Make Dataset **Action->Make Sub-Dataset**
This selection launches and interface that let's the user define criteria for creating a new Dataset from rows of the current Dataset. For SOM, one or more nodes can be selected, with their rows to be used as the rows for the new Dataset, or to be excluded from the new Dataset. **Note** Rows that have been filtered out will be excluded from any new Dataset created this way.

Quit **File->Quit**
This selection closes the Result Window.

Save Results **File->Save Results**
This selection is used to save the current result. This will save the results of an Analytic run so that the results can be viewed immediately. Once an analytic result is saved, for a Dataset, a button will be added to the Dataset Information screen to quickly view the saved result.

Search **Action->Search**
This selection launches the search interface, which allows the user to search the rows for a String with added options of searching visible Row Info fields or all Row Info Fields and case sensitive or not.. See 'Search Interface' for more information on the Search Interface. For SOM, the search will search the currently selected node first, it will then continue the search through the remaining nodes, then start over at the beginning cluster.

Search Again **Action->Search Again**
This selection searches for the next case of the search string using the settings from the previous search.

Select Classification Vector (If Avail) **Classes-> 'Class Vector Name'**
If there is one or more Classification Vectors, then one of those vectors can be selected for class indication of columns in either the Column Dendigram (if column clustering was selected) or the Class Indicator, in the Right Panel.

Select Heatmap Palette – Red/Green **Palette->Red-Green**
This selection selects the traditional colors of Red and Green for the heatmap. Red is positive or Correlated and Green is negative or Anti-Correlated.

Select Heatmap Palette – Yellow/Blue **Palete-> Yellow -Blue (Default)**
This selection selects the traditional colors of Yellow and Blue for the heatmap. Yellow is positive or Correlated and Blue is negative or Anti-Correlated.

Select Heatmap Palette – Gray Scale **Palette->Gray Scale**
This selection selects shades of gray for the heatmap. Light Gray is positive or Correlated and Dark Gray is negative or Anti-Correlated.

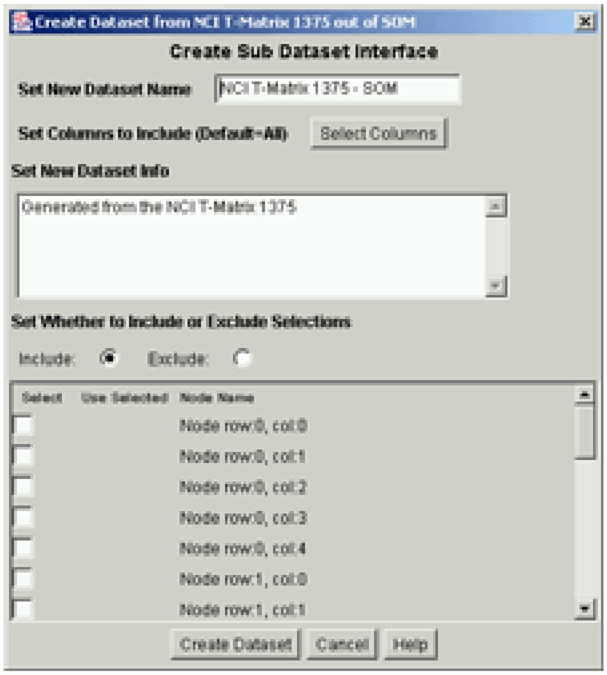
Show Palette **Palette->Show Palette**
This selection Launches the Palette Window, which shows the color palette used by the heatmap.

Show Rows Normalized **Normalize->Norm**
This selection changes the heatmap to show each row normalized. That means that the color corresponding to the highest value in the palette will be applied to the highest value in the row, the color corresponding to the lowest value in the palette will be applied to the lowest value in the row, and all other values will be scaled accordingly.

Show Rows Un-Normalized **Normalize->NoNorm**
This selection displays the traditional heatmap display where color selections from the palette are determine from the highest and lowest value among the data (not just in a row).

Url Targets for Row Fields Interface **File->Preferences->Field Info Urls**
This selection launches the preferences window for Field Info Target Url.

Windows - Create Dataset



Create Dataset from NCI T-Matrix 1375 out of SOM

Create Sub Dataset Interface

Set New Dataset Name: NCI T-Matrix 1375 - SOM

Set Columns to Include (Default=All)

Set New Dataset Info: Generated from the NCI T-Matrix 1375

Set Whether to Include or Exclude Selections

Include: ● Exclude: ○

Select	Use Selected	Node Name
□		Node row:0, col:0
□		Node row:0, col:1
□		Node row:0, col:2
□		Node row:0, col:3
□		Node row:0, col:4
□		Node row:1, col:0
□		Node row:1, col:1

Description

This Interface is used to create a new Dataset by using one or more nodes for defining either the rows in the new dataset, or the rows to be excluded in the new Dataset. In the current node, the selected rows can be used to further refine the selection

Set New Dataset Name:

This text field is the name that the new Dataset will receive. It is seeded with the result Dataset Name along with '- SOM' to indicate that it was created using the results from a SOM result screen. That name can bet set to anything, as long as the name is not already in the system.

Select Columns to Include (Default=All)

The button (Select Columns) launches a window that allows for selection of the columns to be included in the new Dataset. If no column selection is made then all columns will be used.

Set New Dataset Info:

This text field is the Dataset information field. It is seeded from the result Dataset Information. Additional information should be added, here to provide a pedigree for the new Dataset.

Set Whether to Include or Exclude Selections

When Include is selected, only the rows designated will be used in the new Dataset. When Exclude is selected, the rows designated will be subtracted from the rows that made it through the filtering process (if any), to create a new Dataset.

Select Check Boxes

These check boxes give the option to select each node to be used for defining the rows to be included or excluded.

Use Selected Check Box

This check box is only available for the current node. When checked, only the selected rows will be used. It is only shown when rows are selected in the heatmap.

Button – Select All

This button checks all of the Select Check Boxes, making it more convenient to select most of the nodes.

Button – Clear All
This button clears all of the Select Check Boxes, making it more convenient to turn off selection of most or all of the nodes.

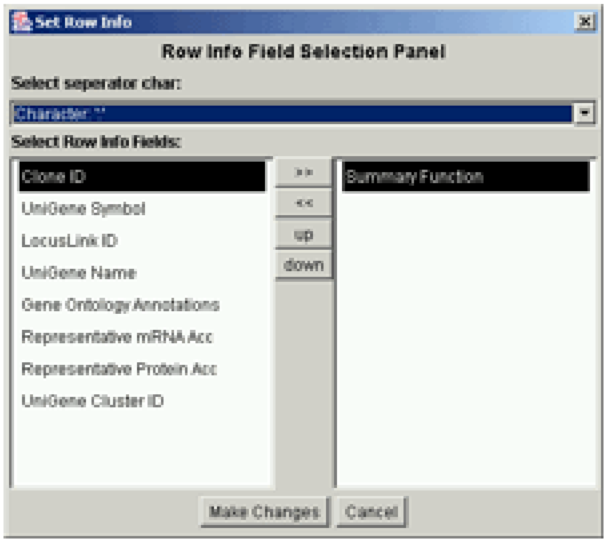
Button – Create Dataset
This button creates the Dataset, once all of the settings are made.

Button – Cancel
This button closes this interface without creating the new Dataset.

Preferences - Row Info Display

Description
The Row Info Display Preferences Window is used to specify the Row Information fields that will be displayed to the right of the Heatmap. In addition to selecting the fields, the field order can be specified along with the character seperating the contents of each field. Fields are selected for inclusion by moving them from the left column (Available) over to the right column (Included). The order of display is set by their row order, with the first row being the first display field.

Select separator char
This selection sets the character that will separate the information from each field.



Set Row Info

Row Info Field Selection Panel

Select seperator char: Character ":"

Select Row Info Fields:

Row Info Field Selection Panel

Select Row Info Fields:
This section allows for selection of the specific fields that will be displayed in the row info and the order of these fields. It has the following buttons:

>> - This button moves the currently hilited field in the left column over to the right column.
<< - This button moves the currently hilited field in the right column over to the left column.
up - This button moves the currently hilited field in the right column up one row.
down - This button moves the currently hilited field in the right column down one row.

Button - Make Changes
This button must be clicked in order for the changes to be made.

Button - Cancel
This button cancels any changes that were made and closes the window.

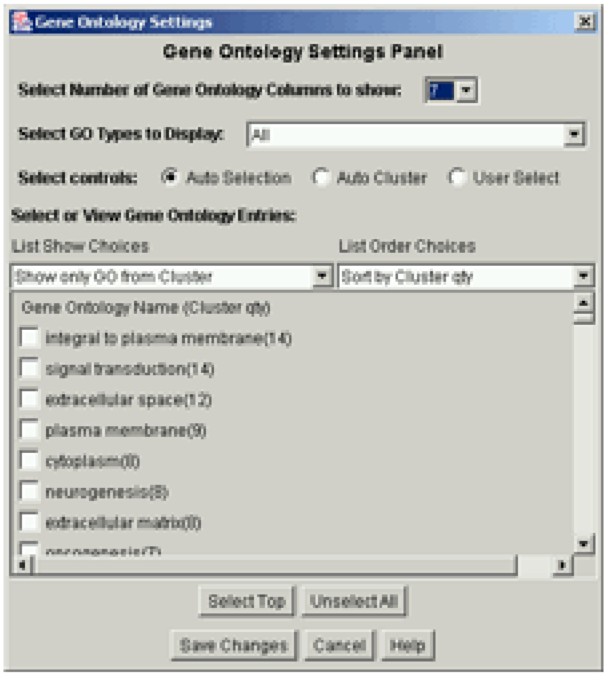
Preferences - Gene Ontology

Description

This interface is used to set the Gene Ontology preferences for this screen. The choices include automatic selection of the most common Gene Ontology categories based upon either the currently selected rows or for all rows in the current node. If the selected rows have preference and there aren't enough categories to fill the number of columns, then the most common categories across all rows in the node are used to fill in the categories. Specific Gene Ontology categories can also be specified, this is aided by the List Show Choices and List Order Choices for quickly finding specific Gene Ontologies.

Select Number of Gene Ontology Columns to show

This selection sets the number of columns that will be displayed.



Select Number of Gene Ontology Columns to show: 7

Select GO Types to Display: All

Select controls: ● Auto Selection ○ Auto Cluster ○ User Select

Select or View Gene Ontology Entries:

List Show Choices: Show only GO from Cluster

List Order Choices: Sort by Cluster qty

Gene Ontology Name (Cluster qty)

□ integral to plasma membrane(14)
□ signal transduction(14)
□ extracellular space(12)
□ plasma membrane(9)
□ cytoplasm(8)
□ neurogenesis(8)
□ extracellular matrix(8)
□ oncogenesis(7)

Gene Ontology Settings Panel

Select GO Types in Display

This selection sets the Gene Ontology Categories to Display with the following options:

Biological Process - Shows only Biological Process
Cellular Component - Shows only Cellular Component
Biological Process & Cellular Component - Shows both Categories
Molecular Function - Shows only Molecular Function
Biological Process & Molecular Function - Shows both Categories
Cellular Component & Molecular Function - Shows both Categories
All - Shows all three Categories

Select control for Gene Ontology Columns

This selection sets the way that Gene Ontology Columns are selected with the following options:

Auto Selection - Chooses most common Gene Ontologies among selected rows (if avail) then fills in remaining, if necessary, from most common among remaining rows in the node.
Auto Node- Chooses most common Gene Ontologies among the rows in the current node.
User Select - Displays rows that the user has specifically selected. This gets turned off when the user clicks on the Gene Ontology Names.

Select or View Gene Ontology Entries

This section is for viewing or selecting specific Gene Ontologies by checking the checkbox either manually or using the 'Select Top' button under the list.

List Show Choices
This selection sets the Gene Ontologies that will be displayed in the list for selection. It has the following choices:

Show All GO - Displays all of the Gene Ontologies that are associated with the current Dataset.
Show only GO from Node- Displays the Gene Ontologies that are associated with genes from the current node..
Show only GO from Selected - Displays only the Gene Ontologies that are associated with the currently selected genes.

List Order Choices
This selection sets the order that the Gene Ontologies are displayed in the list for selection. It has the following choices:

Sort by name - Sorts the Gene Ontologies by their name.
Sort by Node qty - Sorts by the number of times that each Gene Ontology is associated to a gene in the currently node.
Sort by Selected qty - Sorts by the number of times that each Gene Ontology is associated to a gene in the currently selected genes.

Button - Select Top
This button selects the Gene Ontologies that are at the top of the list, in the quantity specified.

Button - Unselect All
This button unselects all of the currently selected Gene Ontologies.

Button - Save Changes
This button sets the currently selected settings and closes the window.

Button - Cancel
This button closes the window without setting any of the changes made.

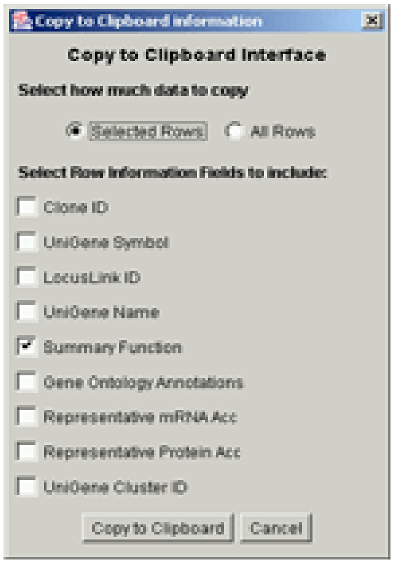
Interface - Copy to Clipboard

Description

This Interface allows the user to select which Row Information Fields to include in the Copy to Clipboard and to specify the rows that they want included in the selection.

Select how much data to copy

This section determines which rows will be selected. When ‘All Rows’ is selected, then all of the rows in the current node are copied to the clipboard, or all nodes (along with their rows) are copied if no node is selected. When ‘Selected Rows’ is selected: If rows are selected in the Heatmap, only those rows will be included. If a node is selected but no rows are selected, all rows in that node will be included. If no node is selected then all nodes (along with their rows) will be copied.



Select how much data to copy

● Selected Rows ○ All Rows

Select Row Information Fields to include:

- □ Clone ID
- □ UniGene Symbol
- □ LocusLink ID
- □ UniGene Name
- ☑ Summary Function
- □ Gene Ontology Annotations
- □ Representative mRNA Acc
- □ Representative Protein Acc
- □ UniGene Cluster ID

Copy to Clipboard Interface

Select Row Information Fields to include

This section allows the user to specify which rows they’d like to include in the Copy to Clipboard. Each Row Information Field that is checked will be included in the copy.

Button - Copy to Clipboard

This Button Initializes the copy and closes the window

Button - Cancel

This Button closes the window without doing the copy.

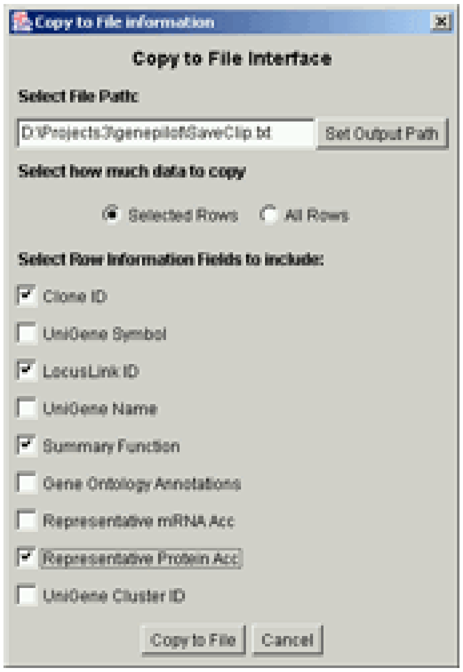
Interface - Copy to File

Description

This Interface allows the user to select specify a File Path in which to write out information contained in this result. This information will include the rows specified and the Row Information Fields that are chosen.

Select File Path

This selection is used to set the path to the output file. Click on the ‘Set Output Path’ to set the file. Traditional copy and paste methods work in this field.



Select File Path:

D:\Projects3\genepilot\SaveClip.txt

Select how much data to copy

● Selected Rows ○ All Rows

Select Row Information Fields to include:

- ☑ Clone ID
- □ UniGene Symbol
- ☑ LocusLink ID
- □ UniGene Name
- ☑ Summary Function
- □ Gene Ontology Annotations
- □ Representative mRNA Acc
- ☑ Representative Protein Acc
- □ UniGene Cluster ID

Copy to File Interface

Select how much data to copy
This section determines which rows will be selected. When 'All Rows' is selected, then all of the rows in the current node are copied to the file, or all nodes (along with their rows) are copied if no node is selected. When 'Selected Rows' is selected: If rows are selected in the Heatmap, only those rows will be included. If a node is selected but no rows are selected, all rows in that node will be included. If no node is selected then all nodes (along with their rows) will be copied.

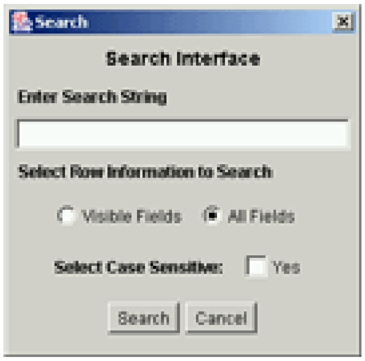
Select Row Information Fields to include
This section allows the user to specify which rows they'd like to include in the Copy to File . Each Row Information Field that is checked will be included in the copy.

Button - Copy to File
This Button Initializes the copy and closes the window

Button - Cancel
This Button closes the window without doing the copy.

Interface - Search

Description
This interface is used to set the search term and the settings for the search. The search can be restricted to just the visible Row Information or to all Row Information. It can also be set to be case sensitive. The search function remembers the last row that it searched in a cluster. When a new cluster is selected, that information is reset.



Search

Search Interface

Enter Search String

Select Row Information to Search

○ Visible Fields ● All Fields

Select Case Sensitive: □ Yes

Search Cancel

Search Interface

Enter Search String
The search string is entered into this field. Traditional copy and paste methods work on this field.

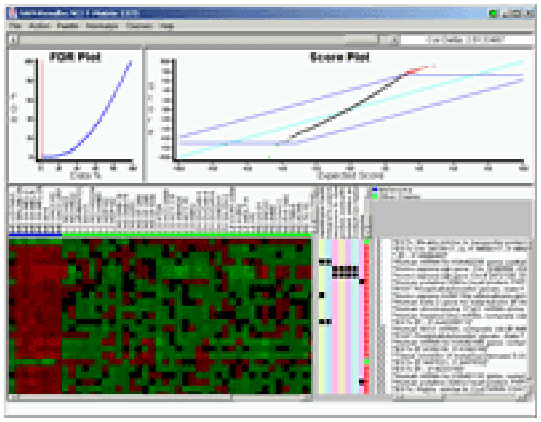
Select Row Information to Search
Selecting 'Visible Fields' restricts the search to just the information that is visible, as set in the Preferences->Row Info Display. Selecting All Fields causes the search to include all Row Info Fields, whether visible or not.

Select Case Sensitive
Checking this box will cause the search to be case sensitive.

Button - Search
This Button activates the search.

Button - Cancel
This button closes the interface without searching.

Chapter 7

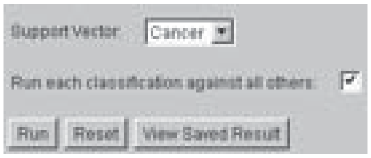


SAM

Description

SAM is a supervised method for finding significant rows in a MicroArray. It requires Vector information in the form of either a Classification Vector or a Shape Vector. Using the information contained in one of these vectors, SAM ranks rows according to their significance to the Vector. In the case of Shape Vectors, rows whose vector has a shape similar to the Shape Vector (or it's mirror image) will have greater significance, resulting in a lower FDR than those rows will less resemblance to the Shape Vector. In the case of Classification Vectors, Rows which have similar values in columns of the same class and (these similar values are) distinct from columns of other classes, have greater significance, resulting in a lower FDR than those rows with less cohesion within a class. An option that is available for Classification Vectors is to add additional runs where each class is compared against all other classes. This helps find rows where values in one class are similar, and the values of this class are distinct from the rest of the values. The result screen is made up of three panels. The FDR Graph Panel plots the FDR against the percentage of rows. The Score Graph Panel plots the score against the expected score. The Heatmap Panel displays the Heatmap and other related information for the significant rows, as selected by the Delta Slider on the top. For more information on SOM, read more about it in 'Significance Analysis of microarrays applied to the ionizing radiation respons (Reference 1 - V. Tusher, et al.).

Run Settings



Support Vector: Cancer

Run each classification against all others: ☑

SAM Run Settings Panel

Description
The Run Settings provide a means to alter the way that SAM Analyzes the Data. The primary action to make, is to select the Support Vector that will be used by SAM. It is recommended to Run each classification against all others, if a classification vector is selected. This will take longer to run, but will provide a wealth of information, well beyond that given by a single run, when more than two classes are present.

Support Vector
This selects the vector to be used by SAM.

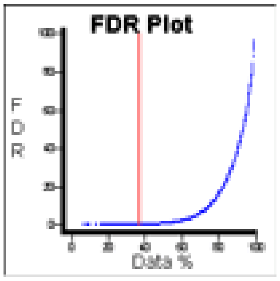
Run each classification against all others
If the selected vector is a Classification vector and this selection is checked, SAM runs more than one time. In the first run, the Classification Vector is used. Then, for each class, as vector is generated where the current class retains it's classification and all other classes are grouped together. This finds significant rows which are specific to a single class.

Button – Run
This button runs the K-Means Clustering analytic program

Button – Reset
This button resets the values to the last saved values.

Button – View Saved Result
This button is only present when a previous result has been saved. Clicking on this button will launch the saved result.

Results - Parts – FDR Graph



SAM FDR Graph

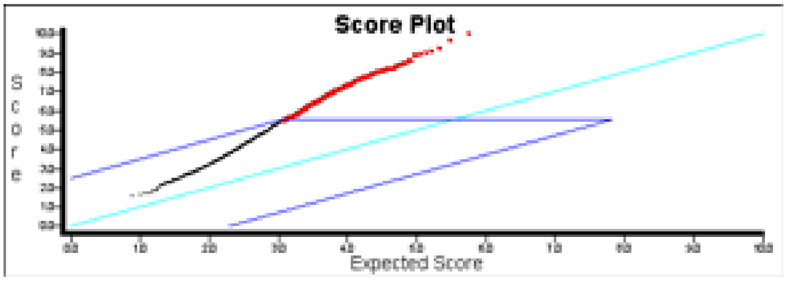
Description

The FDR graph plots the FDR, in the y-axis, against the percentage of the Dataset, in the x-axis. A vertical red line indicates the percentage of rows and FDR for the current Delta selection. The shape of the graph can quickly indicate the quality and distribution of data in a Dataset or Run. A line that starts out closer to horizontal, then curves up indicates a large number of rows that correlate well with the vector. A line that starts out with a steep upward curve, on the other hand indicates few (if any) rows that correlate with the vector.

Actions:

1. Clicking along the line will change the Delta value to the closest delta value to the point selected.

Results - Parts – Score Graph



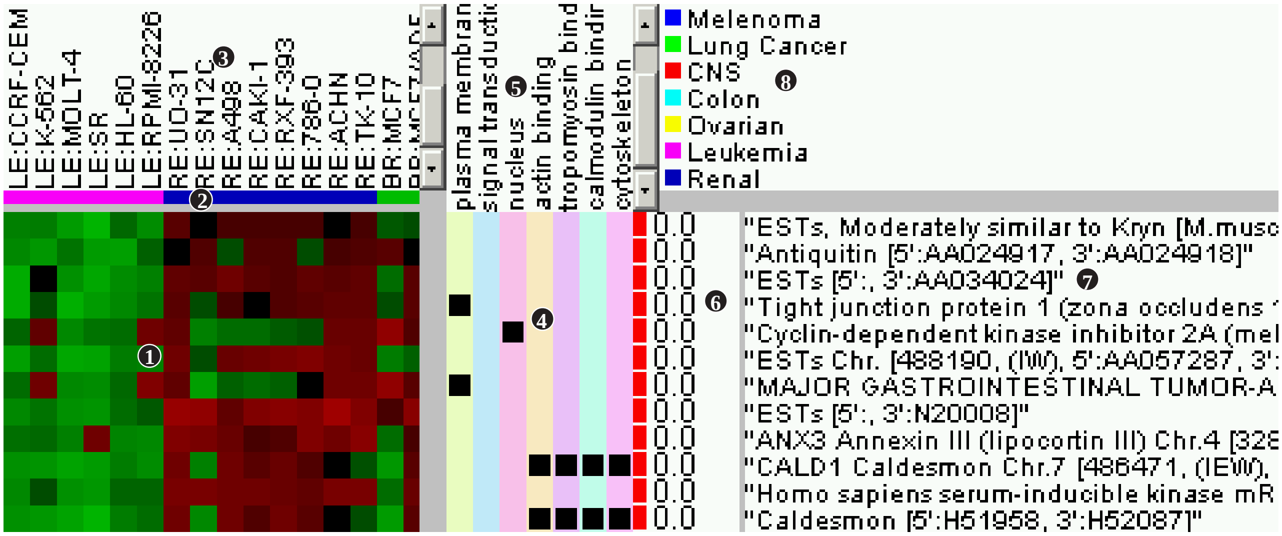
SAM Score Graph: Score vs Expected Score. Current Delta is plotted in Blue

Description

The Score Graph plots the score, in the y-axis, against the expected score (as calculated through regression), in the x-axis. A red parallelogram indicates the boundaries of the current Delta Selection. Points above the boundaries are red, to indicate correlated significance. Points below the boundaries are green, to indicate anti-correlated significance.

Actions: None

Results - Parts – Heatmap Panel


LE:CCRF-CEM
LE:K-562
LE:MOLT-4
LE:SR
LE:HL-60
LE:RPMI-8226
RE:UO-31
RE:SN12C
RE:A498
RE:CAKI-1
RE:RXF-393
RE:786-0
RE:ACHN
RE:TK-10
BR:MCF7
BR:MCF7/ADF
plasma membran
signal transductic
nucleus
actin binding
tropomyosin bind
calmodulin bindir
cytoskeleton
Melenoma
Lung Cancer
CNS
Colon
Ovarian
Leukemia
Renal
0.0 "ESTs, Moderately similar to Kryn [M.musc
0.0 "Antiquitin [5':AA024917, 3':AA024918]"
0.0 "ESTs [5':, 3':AA034024]"
0.0 "Tight junction protein 1 (zona occludens 1
0.0 "Cyclin-dependent kinase inhibitor 2A (mel
0.0 "ESTs Chr. [488190, (IW), 5':AA057287, 3':
0.0 "MAJOR GASTROINTESTINAL TUMOR-A
0.0 "ESTs [5':, 3':N20008]"
0.0 "ANX3 Annexin III (lipocortin III) Chr.4 [328
0.0 "CALD1 Caldesmon Chr.7 [486471, (IEW),
0.0 "Homo sapiens serum-inducible kinase mR
0.0 "Caldesmon [5':H51958, 3':H52087]"
1
2
3
4
5
6
7
8

SAM Heatmap Panel: 1-Heatmap, 2-Class Indicator, 3-Column Information, 4-Gene Ontology Indicator, 5-Gene Ontology Names, 6-Correlation Indicator and FDR, 7-Row Information, 8-Class Information

Description

This panel contains the specific information, for the rows (or clusters). The central part is the Heatmap, which contains colors from the Palette to indicate values of each row (or cluster). To the top, of each column is the Column Info, in the form of text rotated counter-clockwise by 90 degrees. A row of rectangles which are color-coded to indicate the class of each column, will be in between the Heatmap and the Column Info, when Classification Vector(s) are available. To the right of the Heatmap is a column indicating Correlation and the FDR. To the far right of the Heatmap, is the Row Information, which contains the description of the row. In the top right corner, the legend for the Class Indication will define the colors for each class, when Classification Vector(s) are available. When the Delta is set to it's largest value, there will be few rows displayed. As the Delta value is lowered, more rows will be added, as the significance requirements are relaxed.

Heatmap – Bottom Left

The heatmap contains rectangles color-coded in the Palette colors to represent the value of each cel (juncture of 1 row and 1 column). By default, these colors are Red (highest value) and Green (lowest value). When No-Normalize is selected, the value range is defined by the min and max values in the entire Dataset (Rows remaining after filtering), with zero being the medium value. When Normalize is selected, then the value range for each row is defined by the min and max values in that row, with the mean value being the medium value.

Class Indication – Above Average

If there is Classification Vector(s), there will be Class Indicators above the Heatmap. This is a row of rectangles, which are color-coded to indicate the class of each column as defined by the Class Information block in the Top Right corner.

Column Names – Upper Left

The Column Information contains the column names, rotated counter-clockwise by 90 degrees, above each column in the Heatmap.

Gene Ontology Names - Middle Right (If Gene Ontology Information)

The Gene Ontology Names is a variable-column display that displays the Gene Ontology Names that are currently chosen either automatically (Most common in dataset or most common in selected rows) or by the User. Using File->Preferences->Gene Ontology Display, these different choices can be set. If specific Gene Ontologies are selected by the user, the settings can be easily changed back to automatic selection by clicking on this (Gene Ontology Names) display.

Gene Ontology Indication - Bottom Middle

The Gene Ontology Indication indicates the Gene Ontologies that are associated with each row. The Gene Ontologies are chosen either automatically (Most common in dataset or most common in selected rows) or by the User. Using File->Preferences->Gene Ontology Display, these different choices can be set.

Correlation and FDR Column – Bottom Middle

This column indicates whether a row is correlated (red) or anti-correlated(green) with the input vector. It also indicates the FDR value (percentage) for this row.

Row Information – Bottom Right

Row information contains the row name and row information to the right of each row in the heatmap.

Class Information – Top Right (If Class Vector)

If there is Classification Vector(s), there will be Class Information which defines the color coding for the Class Indication row.

Actions:

1. Clicking and dragging on the bitmap will select those rows.
2. Holding down the shift key and selecting a row will extend the row from the previous selection down to the selected row. This can be used to select rows in a group that is larger than those seen on the screen.

Results - Menu Choices

Bitmap from All **Action->Save All to Bitmap**
This selection creates a gif bitmap containing all of the information in the Lower Panel.

Bitmap from Graphs **Action->Graph to Bitmap**
This selection saves the two graphs (FDR and Score Plot) to a gif bitmap.

Bitmap from Selected **Action->Save Sel to Bitmap**
This selection creates a gif bitmap containing all of the information in the Lower Panel. If rows are selected, then only the selected rows will be included in the bitmap. If no rows are selected then all of the rows in the current Delta are saved to the bitmap.

Copy to Clip **Action->Copy to Clip**
This selection launches an interface for copying all or selected rows to the clipboard. This interface also allows for selecting the Row Information Fields which the user is interested in.

Copy to File **Action->Copy to File**
This selection launches an interface for copying all or selected rows to a file. This interface also allows for selecting the Row Information Fields which the user is interested in.

Create Vector **Action->Create Vector**
This selection launches the Create Vector dialog box along with the mean vector of the currently selected rows.

Gene Ontology Display **File->Preferences->Gene Ontology Display**
This selection launches the preferences window for Gene Ontologies. Within this preference window the following can be set; set number of columns, set automatic preferences for selection of Gene Ontologies, select specific Gene Ontologies to display.

Generate Gene Ontology Statistics **Action->Generate GO Statistics**
This selection launches an html page that contains the statistics for the Gene Ontologies for the rows in the following order of precedence: 1. Selectd rows in Right Panel. 2. Rows of currently selected cluster in left panel. 3. All rows in left panel.

Display All Data **Display->All Data**
This choice selects all Data to be displayed which includes Correlated Data and Anti-Correlated Data.

Display Correlated Data **Display->Correlated Data**
This Choice selects only Correlated Data to be displayed. Anti-Correlated data will be excluded from the display.

Display Anti-Correlated Data **Display->Ani-Corr Data**
This Choice selects only Anti-Correlated Data to be displayed. Correlated data will be excluded from the display.

Launch Info Page **Action->Launch Info Page**
This selection launches an information page containing a matrix of the genes vs. Gene Ontologies (if available) and a listing of all of the Row Information Fields along with links to websites which have further information.

Make Dataset **Action->Make Sub-Dataset**
This selection launches and interface that let's the user define criteria for creating a new Dataset from rows of the current Dataset. For SAM, the FDR or selected rows can be used to determine rows to be included or excluded from the new Dataset. If a classification vector was used and multiple runs was selected, then one or more runs can be used to determine the rows to be included or excluded. Selected rows in the current run can also be used.

Quit **File->Quit**
This selection closes the Result Window.

Save Results **File->Save Results**
This selection is used to save the current result. This will save the results of an Analytic run so that the results can be viewed immediately. Once an analytic result is saved, for a Dataset, a button will be added to the Dataset Information screen to quickly view the saved result.

Search **Action->Search**
This selection launches the search interface, which allows the user to search the rows for a String. See 'Search Interface' for more information on the Search Interface. For K-Means Clustering, the search will search the currently selected cluster first, it will then continue the search through the remaining clusters, then start over at the beginning cluster.

Search Again **Action->Search Again**
This selection searches for the next case of the search string.

Select Classification Vector (If Avail) **Classes-> 'Class Vector Name'**
If the vector is a classification Vector, then it will be in this menu. If multiple runs were selected, then each class will also be included in the menu, so that each run can be viewed. If the Vector was a Shape Vector, but there are classification Vector(s) available, then they will be selectable for being used to indicate classes for the columns.

Select Heatmap Palette – Red/Green **Palette->Red-Green**
This selection selects the traditional colors of Red and Green for the heatmap. Red is positive or Correlated and Green is negative or Anti-Correlated.

Select Heatmap Palette – Yellow/Blue **Palete-> Yellow -Blue (Default)**
This selection selects the traditional colors of Yellow and Blue for the heatmap. Yellow is positive or Correlated and Blue is negative or Anti-Correlated.

Select Heatmap Palette – Gray Scale **Palette->Gray Scale**
This selection selects shades of gray for the heatmap. Light Gray is positive or Correlated and Dark Gray is negative or Anti-Correlated.

Show Palette **Palette->Show Palette**
This selection Launches the Palette Window, which shows the color palette used by the heatmap.

Show Rows Normalized **Normalize->Norm**
This selection changes the heatmap to show each row normalized. That means that the color corresponding to the highest value in the palette will be applied to the highest value in the row, the color corresponding to the lowest value in the palette will be applied to the lowest value in the row, and all other values will be scaled accordingly.

Show Rows Un-Normalized **Normalize->NoNorm**
This selection displays the traditional heatmap display where color selections from the palette are determine from the highest and lowest value among the data (not just in a row).

Url Targets for Row Fields Interface **File->Preferences->Field Info Urls**
This selection launches the preferences window for Field Info Target Url.

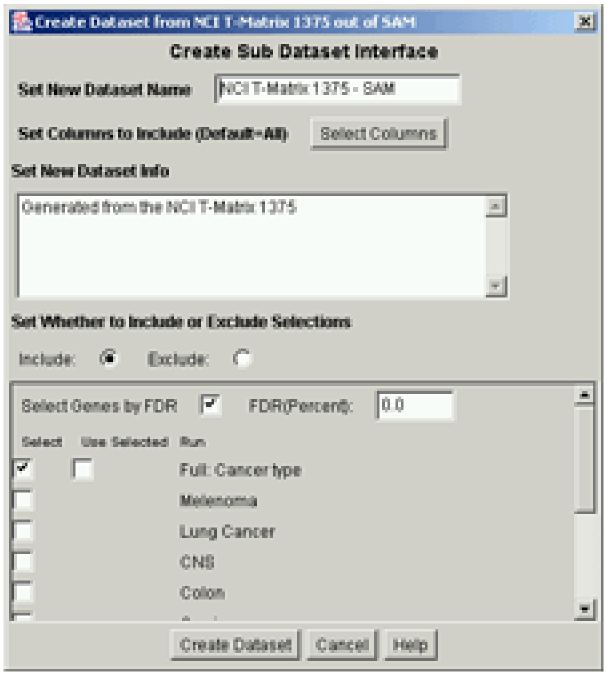
Windows - Create Dataset

Description

This Interface is used to create a new Dataset by using one or more nodes for defining either the rows in the new dataset, or the rows to be excluded in the new Dataset. In the current node, the selected rows can be used to further refine the selection

Select New Dataset Name:

This text field is the name that the new Dataset will receive. It is seeded with the result Dataset Name along with '- SAM' to indicate that it was created using the results from a SAM result screen. That name can bet set to anything, as long as the name is not already in the system.



Create Dataset from NCI T-Matrix 1375 out of SAM

Create Sub Dataset Interface

Set New Dataset Name: NCI T-Matrix 1375 - SAM

Set Columns to Include (Default=All) [Select Columns]

Set New Dataset Info: Generated from the NCI T-Matrix 1375

Set Whether to Include or Exclude Selections

Include: ● Exclude: ○

Select Genes by FDR ☑ FDR(Percent): 0.0

Select	Use Selected	Run
☑	□	Full: Cancer type
□		Melenoma
□		Lung Cancer
□		CNS
□		Colon

Create Sub Dataset Interface

Select Columns to Include (Default=All)

The button (Select Columns) launches a window that allows for selection of the columns to be included in the new Dataset. If no column selection is made then all columns will be used.

Select New Dataset Info:

This text field is the Dataset information field. It is seeded from the result Dataset Information. Additional information should be added, here to provide a pedigree for the new Dataset.

Select Whether to Include or Exclude Selection

When Include is selected, only the rows designated will be used in the new Dataset. When Exclude is selected, the rows designated will be subtracted from the rows that made it through the filtering process (if any), to create a new Dataset.

Select Genes by FDR Checkbox and FDR value

When checked, the row FDR's will be used to select rows from all of the runs selected (no selection necessary if only one run). The FDR value sets the FDR to be used to determine rows selected.

Select Check Boxes

These check boxes will only be available if the was more than one run (i.e. The vector was a classification vector and the option to 'Run each classification against all others' was selected) These check boxes give the option to select each run to be used for defining the rows to be included or excluded.

Use Selected Check Box

This check box is only available for the current run. When checked, only the selected rows will be used. It is only shown when rows are selected in the heatmap.

Button – Select All

This button checks all of the Select Check Boxes, making it more convenient to select most of the runs.

Button – Clear All

This button clears all of the Select Check Boxes, making it more convenient to turn off selection of most or all of the runs.

Button – Create Dataset

This button creates the Dataset, once all of the settings are made.

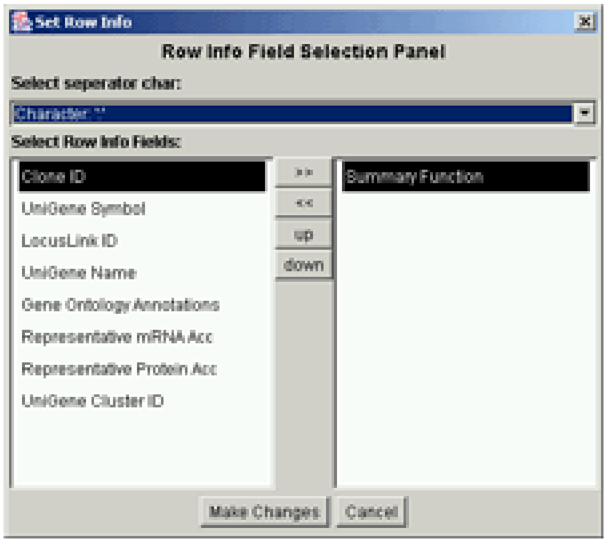
Button – Cancel

This button closes this interface without creating the new Dataset.

Preferences - Row Info Display

Description

The Row Info Display Preferences Window is used to specify the Row Information fields that will be displayed to the right of the Heatmap. In addition to selecting the fields, the field order can be specified along with the character seperating the contents of each field. Fields are selected for inclusion by moving them from the left column (Available) over to the right column (Included). The order of display is set by their row order, with the first row being the first display field.



Set Row Info

Row Info Field Selection Panel

Select seperator char: Character: ' '

Row Info Field Selection Panel

Select separator char

This selection sets the character that will separate the information from each field.

Select Row Info Fields:

This section allows for selection of the specific fields that will be displayed in the row info and the order of these fields. It has the following buttons:

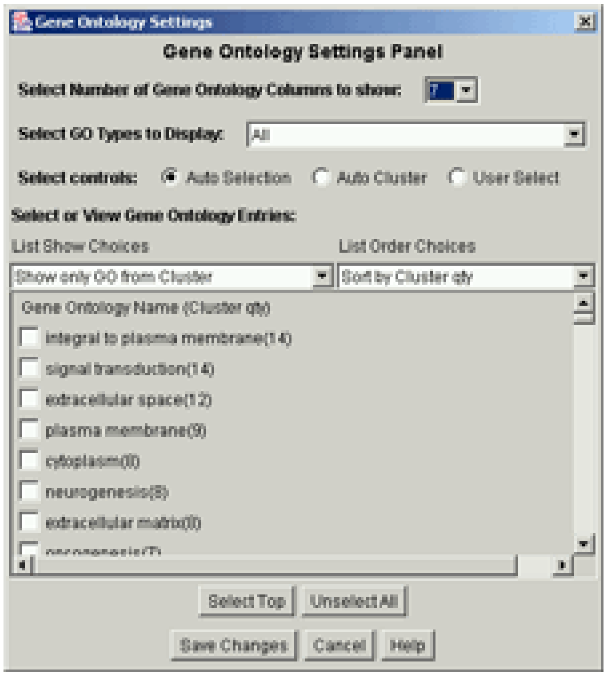
>> - This button moves the currently hilited field in the left column over to the right column.
<< - This button moves the currently hilited field in the right column over to the left column.
up - This button moves the currently hilited field in the right column up one row.
down - This button moves the currently hilited field in the right column down one row.

Button - Make Changes
This button must be clicked in order for the changes to be made.

Button - Cancel
This button cancels any changes that were made and closes the window.

Preferences - Gene Ontology

Description
This interface is used to set the Gene Ontology preferences for this screen. The choices include automatic selection of the most common Gene Ontology categories based upon either the currently selected rows or for all rows in the current delta range. If the selected rows have preference and there aren't enough categories to fill the number of columns, then the most common categories across all rows in the delta range are used to fill in the categories. Specific Gene Ontology categories can also be specified, this is aided by the List Show Choices and List Order Choices for quickly finding specific Gene Ontologies.



Select Number of Gene Ontology Columns to show: 7

Select GO Types to Display: All

Select controls: ● Auto Selection ○ Auto Cluster ○ User Select

Select or View Gene Ontology Entries:

List Show Choices: Show only GO from Cluster

List Order Choices: Sort by Cluster qty

Gene Ontology Name (Cluster qty)

□ integral to plasma membrane(14)
□ signal transduction(14)
□ extracellular space(12)
□ plasma membrane(9)
□ cytoplasm(8)
□ neurogenesis(8)
□ extracellular matrix(8)
□ oncogenesis(7)

Gene Ontology Settings Panel

Select Number of Gene Ontology Columns to show
This selection sets the number of columns that will be displayed.

Select GO Types in Display
This selection sets the Gene Ontology Categories to Display with the following options:

Biological Process - Shows only Biological Process
Cellular Component - Shows only Cellular Component
Biological Process & Cellular Component - Shows both Categories
Molecular Function - Shows only Molecular Function
Biological Process & Molecular Function - Shows both Categories
Cellular Component & Molecular Function - Shows both Categories
All - Shows all three Categories

Select control for Gene Ontology Columns

This selection sets the way that Gene Ontology Columns are selected with the following options:

Auto Selection - Chooses most common Gene Ontologies among selected rows (if avail) then fills in remaining, if necessary, from most common among remaining rows in the current delta range.

Auto Delta Range - Chooses most common Gene Ontologies among the rows displayed for the current delta range.

User Select - Displays rows that the user has specifically selected. This gets turned off when the user clicks on the Gene Ontology Names.

Select or View Gene Ontology Entries

This section is for viewing or selecting specific Gene Ontologies by checking the checkbox either manually or using the 'Select Top' button under the list.

List Show Choices

This selection sets the Gene Ontologies that will be displayed in the list for selection. It has the following choices:

Show All GO - Displays all of the Gene Ontologies that are associated with the current Dataset.

Show only GO from Delta Range- Displays the Gene Ontologies that are associated with genes from the current delta range.

Show only GO from Selected - Displays only the Gene Ontologies that are associated with the currently selected genes.

List Order Choices

This selection sets the order that the Gene Ontologies are displayed in the list for selection. It has the following choices:

Sort by name - Sorts the Gene Ontologies by their name.

Sort by Delta Range qty - Sorts by the number of times that each Gene Ontology is associated to a gene in the currently delta range.

Sort by Selected qty - Sorts by the number of times that each Gene Ontology is associated to a gene in the currently selected genes.

Button - Select Top

This button selects the Gene Ontologies that are at the top of the list, in the quantity specified.

Button - Unselect All

This button unselects all of the currently selected Gene Ontologies.

Button - Save Changes

This button sets the currently selected settings and closes the window.

Button - Cancel

This button closes the window without setting any of the changes made.

Interface - Copy to Clipboard

Description

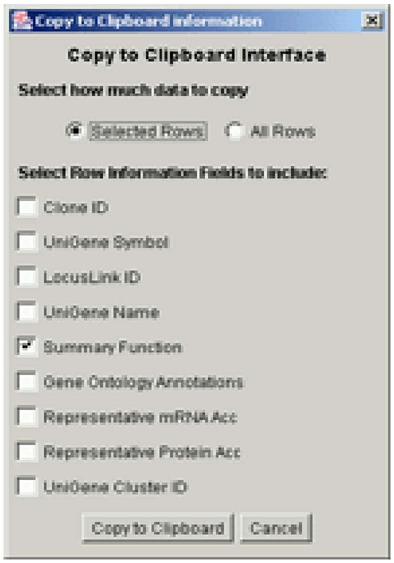
This Interface allows the user to select which Row Information Fields to include in the Copy to Clipboard and to specify the rows that they want included in the selection.

Select how much data to copy

This section determines which rows will be selected. When ‘All Rows’ is selected, then all of the rows in the current delta range are copied to the clipboard. When ‘Selected Rows’ is selected then only the selected rows will be included unless there are no selected rows, in which case all the rows in the current delta will be included.

Select Row Information Fields to include

This section allows the user to specify which rows they’d like to include in the Copy to Clipboard. Each Row Information Field that is checked will be included in the copy.


Copy to Clipboard information

Copy to Clipboard Interface

Select how much data to copy

● Selected Rows ○ All Rows

Select Row Information Fields to include:

- □ Clone ID
- □ UniGene Symbol
- □ LocusLink ID
- □ UniGene Name
- ☑ Summary Function
- □ Gene Ontology Annotations
- □ Representative mRNA Acc
- □ Representative Protein Acc
- □ UniGene Cluster ID

Copy to Clipboard Interface

Button - Copy to Clipboard

This Button Initializes the copy and closes the window

Button - Cancel

This Button closes the window without doing the copy.

Interface - Copy to File

Description

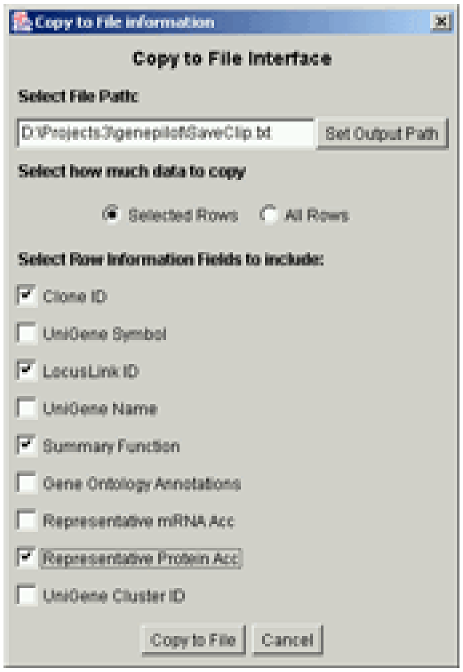
This Interface allows the user to select specify a File Path in which to write out information contained in this result. This information will include the rows specified and the Row Information Fields that are chosen.

Select File Path

This selection is used to set the path to the output file. Click on the ‘Set Output Path’ to set the file. Traditional copy and paste methods work in this field.

Select how much data to copy

This section determines which rows will be selected. When ‘All Rows’ is selected, then all of the rows in the current delta range are copied to the file. When ‘Selected Rows’ is selected then only the selected rows will be included unless there are no selected rows, in which case all the rows in the current delta will be included.


Copy to File information

Copy to File Interface

Select File Path:

D:\Projects3\genepilot\SaveClip.txt

Select how much data to copy

● Selected Rows ○ All Rows

Select Row Information Fields to include:

- ☑ Clone ID
- □ UniGene Symbol
- ☑ LocusLink ID
- □ UniGene Name
- ☑ Summary Function
- □ Gene Ontology Annotations
- □ Representative mRNA Acc
- ☑ Representative Protein Acc
- □ UniGene Cluster ID

Copy to File Interface

Select Row Information Fields to include

This section allows the user to specify which rows they'd like to include in the Copy to File . Each Row Information Field that is checked will be included in the copy.

Button - Copy to File

This Button Initializes the copy and closes the window

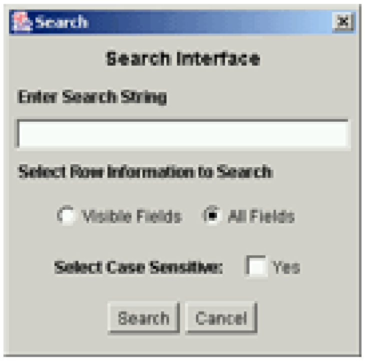
Button - Cancel

This Button closes the window without doing the copy.

Interface - Search

Description

This interface is used to set the search term and the settings for the search. The search can be restricted to just the visible Row Information or to all Row Information. It can also be set to be case sensitive. The search function remembers the last row that it searched in a set. When a new set is selected, that information is reset.



○ Visible Fields ● All Fields

Select Case Sensitive: □ Yes

Search Interface

Enter Search String

The search string is entered into this field. Traditional copy and paste methods work on this field.

Select Row Information to Search

Selecting 'Visible Fields' restricts the search to just the information that is visible, as set in the Preferences->Row Info Display. Selecting All Fields causes the search to include all Row Info Fields, whether visible or not.

Select Case Sensitive

Checking this box will cause the search to be case sensitive.

Button - Search

This Button activates the search.

Button - Cancel

This button closes the interface without searching.

References

1. SAM Reference:
Virginia Tusher, Robert Tibshirani and Gilbert Chu (2000),
'Significance analysis of microarrays applied to the ionizing radiation response'
PNAS 2001 98: 5116-5121, (Apr 24).

2. Good Paper on various clustering methods (1999)
Tibshirani, R., Hastie, T. Eisen, M., Ross, D. , Botstein, D. and Brown, P.
Clustering methods for the analysis of DNA microarray data
Tech. report Oct. 1999.

3. Good paper on Hierarchical Clustering.
Eisen MB, Spellman PT, Brown PO and Botstein D. (1998)
Cluster Analysis and Display of Genome-Wide Expression Patterns.
Proc Natl Acad Sci U S A 95, 14863-8.

4. Very good cluster Analysis Paper
Sturn A, Quackenbush J, Trajanoski Z. Genesis (2002)
Cluster analysis of microarray data.
Bioinformatics. 2002 Jan;18(1):207-8.

5. Best paper on Clustering that I had seen, up to that date
Alexander Sturn - Master Thesis (2001), Institute for Biomedical Engineering, Graz University of Technology, Graz, Austria.
Cluster Analysis for Large Scale Gene Expression Studies

Index

A

Adjusted Data
  Show 25
Archive Dataset 13
Average Heatmap
  Hierarchical Clustering 39

B

Bitmap from All
  Dataset Display 23
  Hierarchical Clustering 41
  K-Means Clustering 53
  SAM 81
  SOM 67
Bitmap from Graph
  K-Means Clustering 53
Bitmap from Graphs
  SAM 81
Bitmap from Grid
  SOM 67
Bitmap from Selected
  Dataset Display 23
  Hierarchical Clustering 41
  K-Means Clustering 54
  SAM 81
  SOM 67

C

Chromosome Location
  Creating Dataset Templates 7
Class Indicator
  Hierarchical Clustering 40
  K-Means Clustering 52
  SAM 80
  SOM 66
Class Information
  Hierarchical Clustering 40
  K-Means Clustering 53
  SAM 80
  SOM 66
Classification Vector
  Creating Dataset Templates 7
Clone ID
  Creating Dataset Templates 7
Col Similarity Metric 36
Column Dendigram
  Hierarchical Clustering 40
Column Info
  Hierarchical Clustering 40
Column Names
  K-Means Clustering 53
  SAM 80
  SOM 66
Copy to Clip
  Dataset Display 23
  Hierarchical Clustering 41
  K-Means Clustering 54
  SAM 81
  SOM 67, 68
Copy to Clipboard Interface
  Dataset Display 29
  Hierarchical Clustering 47
  K-Means Clustering 60
  SAM 88
  SOM 74
Copy to File
  Dataset Display 24
  Hierarchical Clustering 41
  K-Means Clustering 54

SAM 81
SOM 67
Copy to File Interface
Dataset Display 29
Hierarchical Clustering 47
K-Means Clustering 60
SAM 88
SOM 74
Correlation and FDR Column
SAM 80
Create Dataset
Hierarchical Clustering 43
K-Means Clustering 56
SAM 84
SOM 70
Create Vector
Dataset Display 24
Hierarchical Clustering 41
K-Means Clustering 54
SAM 81
SOM 67
Creating Dataset Template
Replicate Column Vector 6
Row Type 6
Creating Dataset Templates 6
Select Data Chip Type 6

D

Data Adjustment
Data Adjustment Options 15
Filtering Options 15
Linear Calibration 15
Mean Centering 15
Median Centering 15
View 13
Data Adjustment Settings 15
Dataset
Archive Dataset 13
Close 13
Export Dataset 13
Filtering Options 15
Graph Panel 16
Import 13
Information Panel 16
Load Dataset Archive 13
Make Template 13
Missing Data Imputation 16
New 13
Open 13
Results Panel 16
Save 14
Select Current 14
Vew Info 15
View Heatmap 14
Dataset Display **22**
Copy to Clipboard 23
Copy to Clipboard Interface 29
Copy to File 24
Copy to File Interface 29
Create Vector 24
Field Info Url Targets Interface 26
Gene Ontology Display Settings 24
Gene Ontology Statistics 24
Heatmap Panel 22
Help 24
Launch Info Page 24
Make Bitmap 23
Normalize Rows 25
Palette 31
Quit 24
Row Info Display 26
Row Information Page 32
Row Information Window 31, 33
Search 24, 30
Search Again 25
Select Heatmap 25
Url Lookup Editor 33
View Adjusted Data 25
View Filtered out Data 26
View Original Data 25
Dataset Info 16

E

Enzymatic Function
Creating Dataset Templates 7
Export
Export Dataset 13
Export Dataset 13
Exporting
Archive Dataset 13

F

FDR Graph
SAM 78
Field Info Url Targets Interface
Dataset Display 26
GenePilot 14

Hierarchical Clustering 43
K-Means Clustering 56
SAM 83
SOM 69
Filter
View 13
Filtered Data
Show 26
Filtering Options 15
Min-Max 15
Negative Values 15
Observations 15
Percent Present 15
Standard Deviation 15

G

GenBank Accession
Creating Dataset Templates 7
Gene Aliases
Creating Dataset Templates 7
Gene Ontology
Dataset Display 23
Hierarchical Clustering 40
K-Means Clustering 53, 66
SAM 80
Gene Ontology Anotations
Creating Dataset Templates 7
Gene Ontology Display
Dataset Display 24
Hierarchical Clustering 41
K-Means Clustering 54
SAM 81
SOM 67
Gene Ontology Settings
K-Means Clustering 58
SAM 86
SOM **72**
Gene Ontology Statistics
Dataset Display 24
Hierarchical Clustering 41
K-Means Clustering 54
SAM 81
SOM 68
GenePilot
Field Info Url Targets Interface 14
Grid size
SOM 63
Grid Type
SOM 63

H

Heatmap
Hierarchical Clustering 39
K-Means Clustering 52
SAM 79
SOM 66
Heatmap Panel
K-Means Clustering 52
SAM 79
SOM 65
Help
Dataset Display 24
Hierarchical Clustering **35**
Bitmap 41
Copy to Clipboard 41
Copy to Clipboard Interface 47
Copy to File 41
Copy to File Interface 47
Create Dataset 43
Create Vector 41
Field Info Url Targets Interface 43
Gene Ontology 40
Gene Ontology Display Settings 41
Gene Ontology Statistics 41
Launch Info Page 41
Left Panel 37
Memory Model 37
Palette 31
Quit 42
Results - Menu Choices 41
Right Panel 39
Row Info Display 44
Row Information Page 32
Row Information Window 31, 33
Run 14
Run Settings 36
Save Results 42
Search 42, 48
Search Again 42
Select Palette 42
Show Palette 43

I

Import
Load Dataset Archive 13
Importing
Additional Row Info 17
Info Page

Dataset Display 24
Hierarchical Clustering 41
K-Means Clustering 54
SAM 82
SOM 68
Installation **2**
All Other Platforms Instructions 4
Generic Unix Instructions 3
HP-UX Instructions 3
Linux Instructions 3
Mac OS Classic (8.1 or greater) Instructions 2
Mac OS X Instructions 2
Solaris Instructions 3
Windows Instructions 2

K

K-Means Clustering **49**
Bitmap 53
Copy to Clipboard 54
Copy to Clipboard Interface 60
Copy to File 54
Copy to File Interface 60
Create Dataset 56
Create Vector 54
Field Info Url Targets Interface 56
Gene Ontology 53, 66
Gene Ontology Display Settings 54
Gene Ontology Settings 58
Gene Ontology Statistics 54
Graph 50
Graph and Button Panel 50
Heatmap Panel 52
Launch Info Page 54
Normalize Rows 55
Palette 31
Quit 54
Row Info Display 57
Row Information Page 32
Row Information Window 31, 33
Run 14
Run Settings 50
Save Results 55
Search 55, 61
Select Palette 55
Show Palette 55

L

Linear Calibration 15
Linkage Measure 36
Load Dataset Archive 13
LocusLink ID
Creating Dataset Templates 7

M

Main Window **12**, **21**
Description 12, 21
Menu Choices 13
Make Dataset
Dataset Display 24
Hierarchical Clustering 42
K-Means Clustering 54
SAM 82
SOM 68
Make Template
Dataset 13
Mean Centering 15
Median Centering 15
Missing Data Imputation 16
Nearest Neighbor 16

N

Nearest Neighbor 16
Neighborhood Type
SOM 63
New Dataset
Make Template 13
Node Panel
SOM 64
Normalize Row
Hierarchical Clustering 43
Normalize Rows
Dataset Display 25
K-Means Clustering 55
SAM 83
SOM 69
Number of Clusters
K-Means Clustering 50

O

Original Data
Show 25

P

Pairs Vector
  Creating Dataset Templates 7
Palette 31
Palette Show
  Dataset Display 25
Pre-Process Settings
  View 13
Preparing Data 8
  Classification Vector 10
  Column Names 9
  Data Header 10
  Data Rows 11
  Dataset Name 8
  Information 9
  Lead Col Count 9
  Model Number 9
  Pair Data Vector 10
  Replicates Vector 9
  Shape Vector 10
  Type of Data 8
  Vectors 10

Q

Quit 13

R

Replicate Column Combine Panel 18
  Column and Row Statistics Display 19
  Mean Value 19
  Median Value 19
  Set Description 19
  Set Name 19
  Weighted Mean Value 19
Replicate Column Vector
  Add to Template 6
Representative mRNA Acc
  Creating Dataset Templates 7
Representative Protein Acc
  Creating Dataset Templates 7
Results
  Launching Row Info Display 14
Row Info Display
  Dataset Display 24, 26
  Hierarchical Clustering 44
  K-Means Clustering 57
  Launching settings interface 14
  SAM 85
  SOM 71
Row Info Import 17
Row Information
  Hierarchical Clustering 40
  K-Means Clustering 53, 66
  SAM 80
  SOM 66
Row Information Page 32
Row Information Window 31, 33
Row Similarity Metric 36
Row Type
  Creating Dataset Templates 6

S

SAM **76**
  Bitmap 81
  Copy to Clipboard 81
  Copy to Clipboard Interface 88
  Copy to File 81
  Copy to File Interface 88
  Create Dataset 84
  Create Vector 81
  Display All Data 81
  Display Anti-Correlated Data Only 82
  Display Correlated Data Only 81
  FDR Graph 78
  Field Info Url Targets Interface 83
  Gene Ontology 80
  Gene Ontology Display Settings 81
  Gene Ontology Settings 86
  Gene Ontology Statistics 81
  Heatmap Panel 79
  Launch Info Page 82
  Make Dataset 82
  Normalize Rows 83
  Palette 31
  Quit 82
  Row Info Display 85
  Row Information Page 32
  Row Information Window 31, 33
  Run 14
  Run Settings 77
  Save Results 82
  Score Graph 78
  Search 82, 89
  Select Palette 83
Save Results

SAM 82
Score Graph
SAM 78
Search
Dataset Display 30
Hierarchical Clustering 42, 48
K-Means Clustering 55, 61
SAM 82, 89
SOM 68, 75
Search Again
K-Means Clustering 55
Select Palette
SAM 83
SOM 69
Self Organizing Map
See SOM **62**
Self Organizing Maps
Run 14
Settings
Save 14
Shape Vector
Creating Dataset Templates 7
Show Palette
Hierarchical Clustering 43
K-Means Clustering 55
SOM 69
Significance of MicroArrays
See SAM **76**
SOM **62**
Bitmap 67
Copy to Clip 68
Copy to Clipboard 67
Copy to Clipboard Interface 74
Copy to File 67
Copy to File Interface 74
Create Dataset 70
Create Vector 67
Field Info Url Targets Interface 69
Gene Ontology Display Settings 67
Gene Ontology Settings **72**
Gene Ontology Statistics 68
Grid size 63
Grid Type 63
Heatmap Panel 65
Initial Seed Type 64
Initial Size of Neighborhood 63
Launch Info Page 68
Make Dataset 68
Menu Choices 67
Neighborhood Type 63
Node Panel 64
Normalize Rows 69
Number of Iterations 63
Palette 31
Quit 68
Row Info Display 71
Row Information Page 32
Row Information Window 31, 33
Run 14
Run Settings 63
Save Results 68
Search 68, 75
Select Palette 69
Show Palette 69
Summary Function
Creating Dataset Templates 7
Support Vector
SAM 77

T

Tutorials
Sample Data 5
Your Data 5

U

UniGene Cluster ID
Creating Dataset Templates 7
UniGene Name
Creating Dataset Templates 7
UniGene Symbol
Creating Dataset Templates 7
Url Lookup Editor
Dataset Display 33

V

Vectors
Classification Vector 10
Pair Data 10
Preparing Data 10
Replicates Vector 9
Shape Vector 10