

# **GenePattern Java Client Guide**

**Software Copyright**

The Broad Institute  
SOFTWARE COPYRIGHT NOTICE AGREEMENT

This software and its documentation are copyright 2006 by the Broad Institute/Massachusetts Institute of Technology.  
All rights are reserved.

This software is supplied without any warranty or guaranteed support whatsoever. Neither the Broad Institute nor MIT  
can be responsible for its use, misuse, or functionality.

# Table of Contents

Getting Started .....	5
Installing GenePattern .....	5
Starting the GenePattern Server .....	5
Starting the GenePattern Java Client .....	5
Exiting from GenePattern .....	6
Getting Help .....	7
Working with Projects.....	7
Projects Pane .....	7
Opening a Project Directory .....	8
Refreshing a Project Directory.....	8
Closing a Project Directory .....	9
Working with Project Files .....	9
Working with Analysis Results .....	9
Results Pane .....	9
Saving Analysis Results .....	10
Deleting Analysis Results .....	10
Working with Analysis Results Files .....	11
Running Analysis and Visualization Modules.....	12
About Analysis and Visualization Modules .....	12
What are Analysis and Visualization Modules? .....	12
Adding Modules to GenePattern.....	12
How GenePattern Runs Analyses .....	13
GenePattern Jobs.....	13
Running an Analysis or Visualization Module.....	14
Checking Job Status .....	15
Viewing Analysis Results .....	16
Rerunning Analyses .....	16
Working with Pipelines .....	16
About Pipelines .....	17
What Are Pipelines? .....	17
How GenePattern Runs Pipelines .....	17
Running a Pipeline .....	17
Viewing a Pipeline .....	18
Creating a Pipeline .....	20
Editing a Pipeline.....	21
Editing the Pipeline Description .....	21
Adding Tasks.....	22
Removing Tasks.....	24
Reordering Tasks .....	24
Exporting and Importing Pipelines.....	24

Working with Suites.....	25
Viewing or Editing a Suite .....	25
Creating a Suite.....	26
Using Filters .....	27
Exporting and Importing Suites .....	27
Creating Input Files .....	28
Tutorial: Creating a gct File .....	28
Understanding Version Numbers .....	31
Menu Bar .....	31
File .....	31
Projects .....	31
Results .....	32
Analysis.....	32
Visualization .....	32
Pipelines.....	32
Suites .....	32
History .....	33
Help.....	33

# Getting Started

GenePattern provides access to a broad array of computational methods used to analyze genomic data. Its extendable architecture makes it easy for computational biologists to add analysis and visualization modules, which ensures that GenePattern users have access to new computational methods on a regular basis.

GenePattern has three components:

- The server, which is the GenePattern engine.
- The Java Client, which is the primary GenePattern interface for most users. This guide describes the Java Client and how to use it.
- The Web Client, which is primarily used for administrative operations, such as adding and deleting analysis modules. The [GenePattern Web Client Guide](#) describes the Web Client and how to use it.

This section describes how to get started with the GenePattern Java Client. It contains the following topics:

- [Installing GenePattern](#)
- [Starting the GenePattern Server](#)
- [Starting the GenePattern Java Client](#)
- [Exiting from GenePattern](#)
- [Getting Help](#)

## Installing GenePattern

When you install GenePattern, you typically install all three components: the server, Java Client, and Web Client. GenePattern can be run standalone, or separated into its client and server components to take advantage of a more powerful compute server. A single installer works for either local or distributed configurations.

If you have not yet installed GenePattern, you can do so by following the [installation instructions](#). If you are new to GenePattern, you may find it helpful to work through the [GenePattern Tutorial](#) before reading the more detailed information in this guide.

## Starting the GenePattern Server

You must start the GenePattern server before you can start either the Java Client or the Web Client. The server takes several seconds to start; clients cannot connect to it while it is starting.

To start the GenePattern server, double-click the Start GenePattern Server icon, shown below. By default, installing GenePattern places this icon on your desktop.



StartGenePatt  
ernServer

On Windows, while the server is starting, a console window displays messages. The server is ready when the console displays the following message:

```
*****  
GenePattern server version 2.0.0 build 2.0.0.358 built Thu, Dec 15 2005 is ready  
*****
```

On Mac OS X, while the server is starting, the server icon bounces in the Dock. The server is ready when the icon stops bouncing. If you want to read the messages, see the Console.

## Starting the GenePattern Java Client

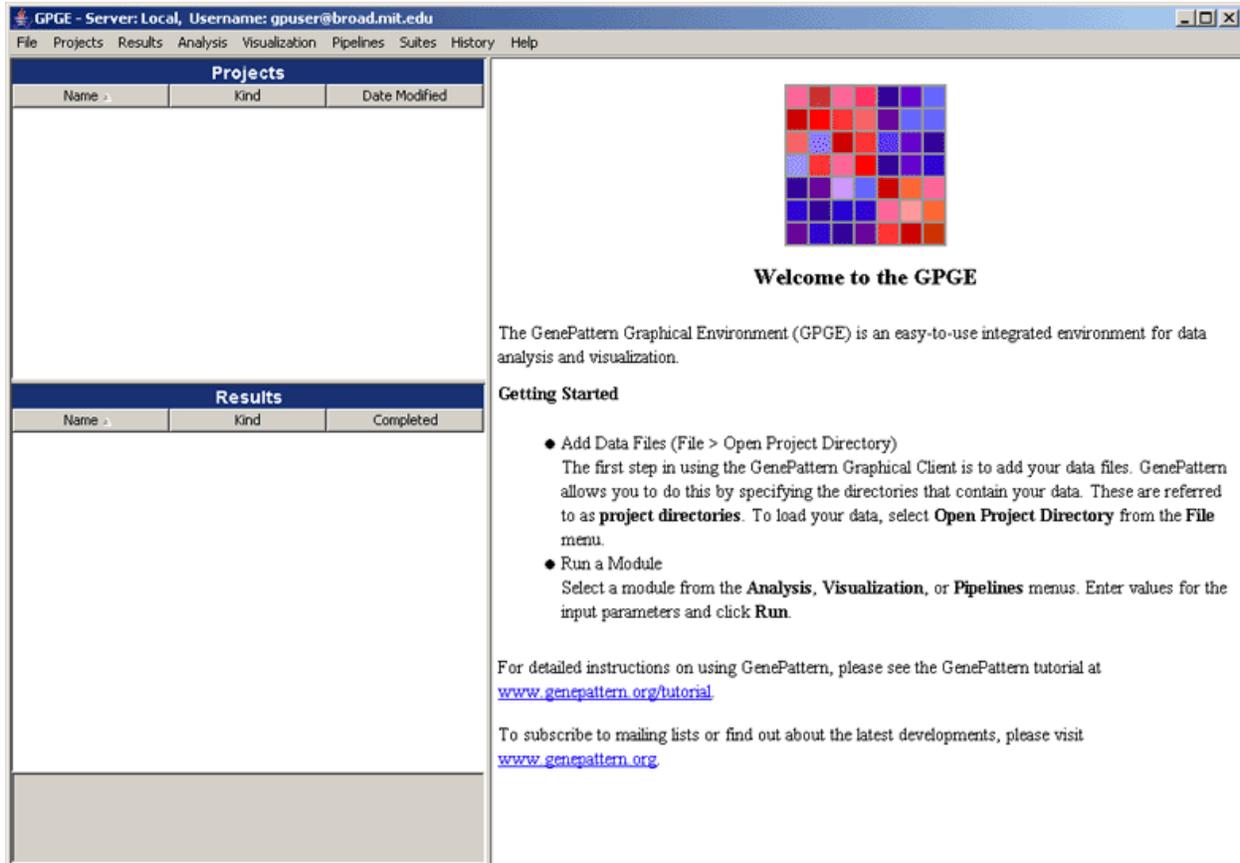
You must [start the GenePattern server](#) before you can start the Java Client.

To start the GenePattern Java Client, double-click the GenePattern Java Client icon, shown below. By default, installing GenePattern places this icon on your desktop.



GenePattern  
Client

As the client starts, it connects to the GenePattern server, retrieves the list of installed analysis modules, visualization modules, pipelines, and suites, and builds its menus based on that list. When it is finished, the Java Client displays the following welcome message and you can begin working:



The title bar shows you the GenePattern server and user name that you are using. The Java Client, which remains connected to the server throughout the session, runs all analyses on this server. Typically, your GenePattern user name is your e-mail address.

## Exiting from GenePattern

To exit from GenePattern:

1. Exit from the Java and Web clients.
  - To exit from the GenePattern Java Client, select *File>Quit*.
  - To exit from the GenePattern Web Client, if it is running, click the *sign out* link in the top right corner of the page.
2. Shutdown the server by double-clicking the Stop GenePattern Server icon, shown below. Alternative: if you are using Windows, you can close the console window.



StopGenePatt  
ernServer

On Windows, when you shutdown GenePattern, the GenePattern console window closes. On Mac OS X, the GenePattern server icon disappears from the Dock. On Linux machines, the GenePattern server exits silently.

If a GenePattern client remains running after you shutdown the server, the client displays error messages as it attempts to contact the server.

## Getting Help

The [GenePattern web site](#) provides an overview of GenePattern and its analysis modules, as well as links to the GenePattern software and documentation. The documentation is your primary source for help with GenePattern:

- The [GenePattern Tutorial](#) provides a brief tour of common operations.
- This guide, the [GenePattern Java Client Guide](#), describes the Java Client and how to use it.
- The [GenePattern Web Client Guide](#) describes the Web Client and how to use it.
- The [GenePattern Programmer's Guide](#) describes how to access GenePattern from various programming environments and provides instructions for writing GenePattern modules.
- The [Modules](#) page lists all GenePattern modules, with links to their documentation. Each analysis, visualization, and pipeline module has its own documentation.
- The [File Formats](#) page describes the GenePattern file formats.

We welcome your feedback. If you have suggestions, comments, or questions, contact the GenePattern help desk ([gp-help@broad.mit.edu](mailto:gp-help@broad.mit.edu)).

## Working with Projects

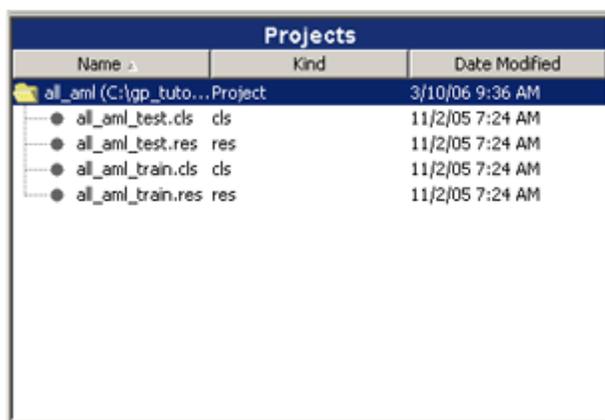
In GenePattern, the directories that contain your data are called **project directories**. Although you do not have to open your project directories in GenePattern, doing so makes it easier to find and analyze the data in those directories.

This section describes how to work with project directories in GenePattern:

- [Projects Pane](#)
- [Opening a Project Directory](#)
- [Refreshing a Project Directory](#)
- [Closing a Project Directory](#)
- [Working with Project Files](#)

### Projects Pane

The Projects pane lists the project directories that you have opened in GenePattern:



Name	Kind	Date Modified
all_aml (C:\gp_tuto...Project	Project	3/10/06 9:36 AM
● all_aml_test.cls	cls	11/2/05 7:24 AM
● all_aml_test.res	res	11/2/05 7:24 AM
● all_aml_train.cls	cls	11/2/05 7:24 AM
● all_aml_train.res	res	11/2/05 7:24 AM

The three columns in the Projects pane describe the project directories and files:

- **Name:** shows the name of the directory or file. Each directory name is followed by the directory path in parentheses.
- **Kind:** describes the content of the directory or file. For directories, Kind is Project. For files, Kind is usually the file extension. If the file extension is ODF, Kind is one of the following:
  - Gene List: the file contains lists of genes or markers

- PredictionResults: the file contains the results of a prediction analysis
- SOMCluster: the file contains the results of a SOMClustering analysis
- ComparativeMarkerSelection: the file contains the results of a ComparativeMarkerSelection analysis
- **Date Modified:** shows the date and time the directory or file was last modified.

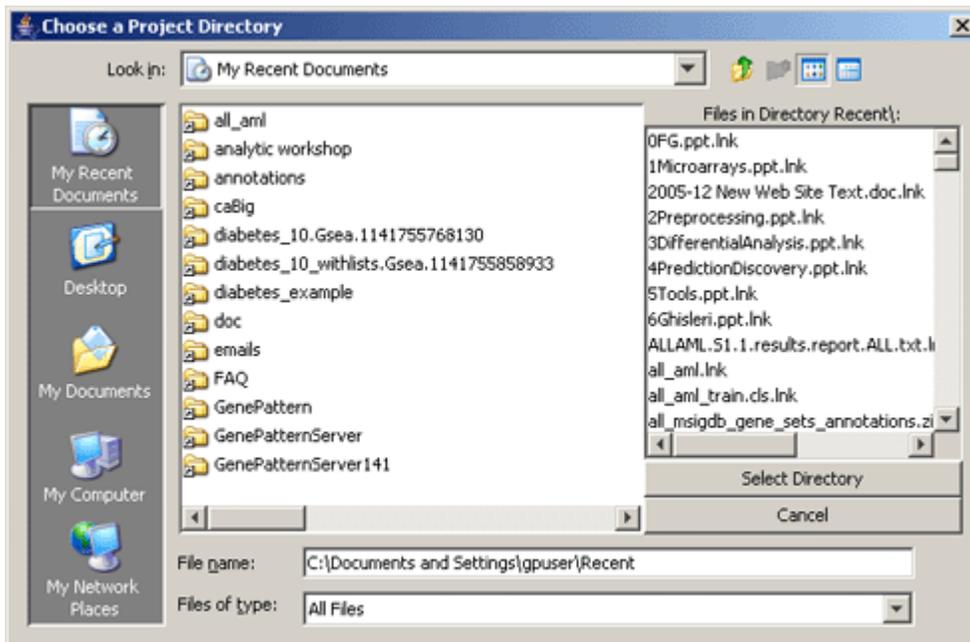
In the Projects pane, you can:

- Click a column label to sort the directories and files based on the information in that column.
- Click-and-drag the column boundary to change the width of the column.
- Double-click a directory name to hide and show the files in that directory.
- Click a file name to display its size and statistics relevant to the file type; file statistics are displayed below the Results pane.
- Double-click a file name to open the file in the default application (if any) for this file type. For information about associating an application with a file type, see your operating system documentation.

## Opening a Project Directory

To open a project directory:

1. Select *File>Open Project Directory*. GenePattern displays the Choose a Project Directory window:



2. In the Choose a Project Directory window, navigate to the directory that contains your data files and click *Select Directory*.

GenePattern adds the directory to the Projects pane. When you next restart the GenePattern Java Client, GenePattern automatically adds this directory to the Projects pane.

## Refreshing a Project Directory

GenePattern periodically reads the contents of your project directories, updating the Projects pane to reflect any changes that have been made. If you change the files in a project directory, you can update the Projects pane immediately by refreshing the directory.

To refresh a project directory:

1. Click the project directory in the Projects pane.
2. Select *Projects>Refresh*. GenePattern reads the contents of the directory and updates the Projects pane.

Alternative: right-click (Mac users Option-click) the project directory that you want to refresh and select *Refresh* from the context menu.

## Closing a Project Directory

To close a project directory, removing it from the Projects pane:

1. Click the project directory.
2. Select *Projects>Close Project*. GenePattern removes the directory from the Projects pane; it is not added when you next restart the GenePattern Java Client.

Alternative: right-click (Mac users Option-click) the project directory that you want to remove and select *Close Project* from the context menu.

## Working with Project Files

By adding a project directory to GenePattern, you make it easier to work with the files in that directory. To work with a file in the Projects pane:

1. Select the file.
2. Select one of the following actions from the Projects menu:
  - *Send To>parameter\_name*: When you run an analysis (see [Running Analysis and Visualization Modules](#)), you must enter values for each parameter. *Send To* makes this file the input value for the selected parameter.
  - *Open With>Text Viewer*: Display the content of this file in the Text Viewer. You can use the Text Viewer to copy text to the clipboard, but not to modify the file.
  - *Open With>Default Application*: Display this file in the default application (if any) associated with this file type. For information about associating an application with a file type, see your operating system documentation.
  - *Show File Location*: Display this file and its parent directory in a file browser.
  - *Modules>module\_name*: When you run an analysis, you must enter values for each parameter. *Modules* runs the selected analysis and makes this file the input value for the first parameter that accepts this file type.

*Modules* provides a quick way to determine which modules accept this file as input. However, if a module is not listed, it may be because it has not yet been annotated, not because it cannot be used with this file. You can access all modules from the Analysis and Visualization menus.

Alternative: right-click (Mac users Option-click) a file and select the desired action from the context menu.

## Working with Analysis Results

When you run an analysis (see [Running Analysis and Visualization Modules](#)), the analysis results are stored on the GenePattern server and displayed in the Results pane of the application window.

This section describes how to view and manage analysis results:

- [Results Pane](#)
- [Saving Analysis Results](#)
- [Deleting Analysis Results](#)
- [Working with Analysis Results Files](#)

## Results Pane

The Results pane displays the analyses that you have run. More precisely, it displays the analysis results files stored for you on the GenePattern server:

Results		
Name	Kind	Completed
ComparativeMarkerSelection (233)	Job	3/10/06 1:57 PM
all_aml_train.comp.marker.odf	Comparative ...	
gp_task_execution_log.txt	txt	
ExtractComparativeMarkerResults (...)	Job	3/10/06 2:46 PM
Golub,Slonim.1999.Nature.all.aml.pi...	Job	Processing

The three columns in the Results pane describe your analysis results:

- **Name:** shows each analysis you have run and its generated files. Each analysis is identified by its name followed by its job number in parentheses. The generated files include the analysis results files and one or more of the following log files:
  - `gp_task_execution_log.txt`: contains the parameter values used to run the analysis, which is useful for reproducing analysis results.
  - `*.stdout`: contains "standard output" messages; that is, comments generated as the analysis module runs.
  - `*.stderr`: contains "standard error" messages; that is, information about errors (if any) that occurred during the analysis.
- **Kind:** describes the analysis or file content. For analyses, Kind is Job. For files, Kind is the file extension, unless the file extension is ODF (as described in [Projects Pane](#)).
- **Completed:** shows the date and time the analysis finished, or Processing if the job is still running.

You can manipulate the columns and files in the Results pane, just as you can the Projects pane (see [Projects Pane](#)).

## Saving Analysis Results

When you run an analysis, the analysis results are stored on the GenePattern server and displayed in the Results pane. Server storage is temporary and results files are deleted after they have been on the server for a certain length of time (by default, one week). To save a results file, you must copy it from the server to a more permanent location.

To save an analysis results file:

1. In the Results pane, click the file that you want to save.
2. Select *Results>Save To*. GenePattern displays the *Save To* submenu, which lists your project directories.
3. Select a project directory, or select *Other* to choose a directory that you have not added to GenePattern. GenePattern copies the file to the selected directory. If you copy the file to a project directory, GenePattern refreshes the Projects pane to display the new file.

Alternative: right-click (Mac users Option-click) the file to save and select *Save To* from the context menu.

## Deleting Analysis Results

As you run more analyses, the Results pane contains an increasing number of jobs. If you no longer need the results of an analysis, you can delete that job. Deleting a job removes it from the Results pane and deletes its results files from the GenePattern server.

You can delete analysis jobs in many ways:

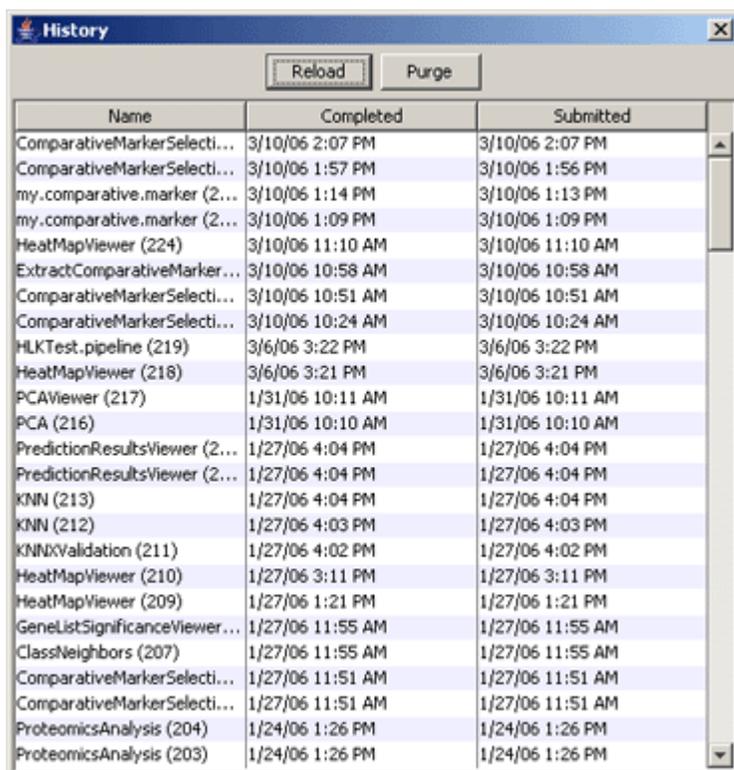
- To delete all analyses and their results files, select *Results>Delete All Jobs*.
- To delete a single results file, click the analysis results file and select *Results>Delete File*.

Alternative: right-click (Mac users Option-click) the file and select *Delete File*.

- To delete an analysis and all its results files, click the analysis and select *Results>Delete Job*.

Alternative: right-click (Mac users Option-click) the analysis and select *Delete Job*.

- To delete one or more analyses, use the History window:
  1. Select *History>View All*. GenePattern displays the History window, as shown below.
  2. For each analysis that you want to delete, select the analysis and click *Purge*.
  3. When you are finished deleting analyses, close the History window by clicking the Close (X) icon in the upper right corner.



Name	Completed	Submitted
ComparativeMarkerSelecti...	3/10/06 2:07 PM	3/10/06 2:07 PM
ComparativeMarkerSelecti...	3/10/06 1:57 PM	3/10/06 1:56 PM
my.comparative.marker (2...	3/10/06 1:14 PM	3/10/06 1:13 PM
my.comparative.marker (2...	3/10/06 1:09 PM	3/10/06 1:09 PM
HeatMapView (224)	3/10/06 11:10 AM	3/10/06 11:10 AM
ExtractComparativeMarker...	3/10/06 10:58 AM	3/10/06 10:58 AM
ComparativeMarkerSelecti...	3/10/06 10:51 AM	3/10/06 10:51 AM
ComparativeMarkerSelecti...	3/10/06 10:24 AM	3/10/06 10:24 AM
HLKTest.pipeline (219)	3/6/06 3:22 PM	3/6/06 3:22 PM
HeatMapView (218)	3/6/06 3:21 PM	3/6/06 3:21 PM
PCAVIEW (217)	1/31/06 10:11 AM	1/31/06 10:11 AM
PCA (216)	1/31/06 10:10 AM	1/31/06 10:10 AM
PredictionResultsViewer (2...	1/27/06 4:04 PM	1/27/06 4:04 PM
PredictionResultsViewer (2...	1/27/06 4:04 PM	1/27/06 4:04 PM
KNN (213)	1/27/06 4:04 PM	1/27/06 4:04 PM
KNN (212)	1/27/06 4:03 PM	1/27/06 4:03 PM
KNNValidation (211)	1/27/06 4:02 PM	1/27/06 4:02 PM
HeatMapView (210)	1/27/06 3:11 PM	1/27/06 3:11 PM
HeatMapView (209)	1/27/06 1:21 PM	1/27/06 1:21 PM
GeneListSignificanceViewer...	1/27/06 11:55 AM	1/27/06 11:55 AM
ClassNeighbors (207)	1/27/06 11:55 AM	1/27/06 11:55 AM
ComparativeMarkerSelecti...	1/27/06 11:51 AM	1/27/06 11:51 AM
ComparativeMarkerSelecti...	1/27/06 11:51 AM	1/27/06 11:51 AM
ProteomicsAnalysis (204)	1/24/06 1:26 PM	1/24/06 1:26 PM
ProteomicsAnalysis (203)	1/24/06 1:26 PM	1/24/06 1:26 PM

## Working with Analysis Results Files

GenePattern makes it easy to work with the files in the Results pane:

1. Select the file.
2. Select one of the following actions from the Results menu:
  - *Send To>parameter\_name*: When you run an analysis (see [Running Analysis and Visualization Modules](#)), you must enter values for each parameter. *Send To* makes this file the input value for the selected parameter.
  - *Save To>directory*: Copies this file from the GenePattern server (temporary storage) to the selected directory. See [Saving Analysis Results](#).
  - *Delete File*: Deletes this file from the GenePattern server. See [Deleting Analysis Results](#).
  - *Open With>Text Viewer*: Displays the content of this file in the Text Viewer. You can use the Text Viewer to copy text to the clipboard, but not to modify the file.
  - *Open With>Default Application*: Displays this file in the default application (if any) associated with this file type. For information about associating an application with a file type, see your operating system documentation.
  - *Modules>module\_name*: When you run an analysis, you must enter values for each parameter. *Modules* runs the selected analysis and makes this file the input value for the first parameter that accepts this file type.

*Modules* provides a quick way to determine which modules accept this file as input. However, if a module is not listed, it may be because it has not yet been annotated, not because it cannot be used with this file. You can access all modules from the Analysis and Visualization menus.

- *Create Pipeline*: Create a pipeline that contains the modules used to create this file. See [Working with Pipelines](#).

Alternative: right-click (Mac users Option-click) a file and select the desired action from the context menu.

## Running Analysis and Visualization Modules

GenePattern provides access to a broad array of computational methods used to analyze genomic data. This section describes how to run the analysis modules that implement the computational methods and the visualization modules that display the analysis results graphically:

- [About Analysis and Visualization Modules](#)
- [Running an Analysis or Visualization Module](#)
- [Checking Job Status](#)
- [Viewing Analysis Results](#)
- [Rerunning Analyses](#)

## About Analysis and Visualization Modules

This section provides background information about analysis and visualization modules:

- [What are Analysis and Visualization Modules?](#)
- [Adding Modules to GenePattern](#)
- [How GenePattern Runs Analyses](#)
- [GenePattern Jobs](#)

## What are Analysis and Visualization Modules?

Analysis and visualization modules are at the heart of GenePattern:

- Analysis modules provide computational methods and tools for gene expression analysis, proteomics data analysis, SNP analysis, and data preprocessing and conversion.
- Visualization modules display your data and analysis results graphically. GenePattern includes two types of visualization modules:
  - Viewers display your data graphically and allow you to manipulate that view interactively. By convention, these modules have "Viewer" in the name.
  - Image creators create static graphics (for example, gif and pdf files) for display in other applications. By convention, these modules have "Image" in the name.

Each module includes its own documentation, which is supplied by the module developer. To display the documentation for a module:

1. Select the module from the Analysis or Visualization menu. The parameters for the selected module appear in the main window.
2. Click *Help* to display the documentation for that module.

For a list of the modules in the repository maintained by the Broad Institute, with links to their documentation, see the [Modules](#) page of the GenePattern web site.

## Adding Modules to GenePattern

The Analysis and Visualization menus list all of the modules available to you; that is, all of the modules on your GenePattern server. You can add modules to your server from the Java Client or the Web Client:

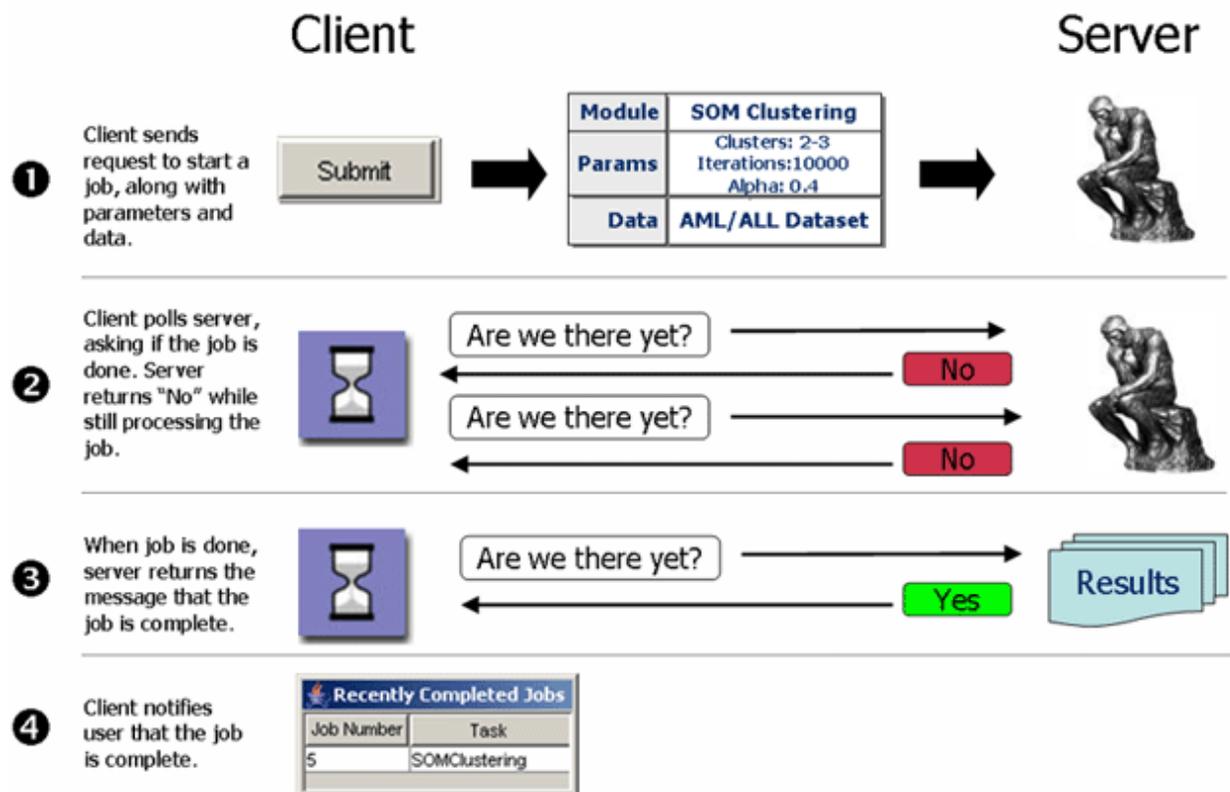
- From the Java Client: import a module from a zip file (*File>Import Module*).
- From the Web Client: import a module from a zip file, install modules from the module repository maintained by the Broad Institute, or create and deploy your own modules. For more information about the Web Client, see the [GenePattern Web Client Guide](#). Typically, you use the Web Client to add modules to GenePattern.

Module names are color coded to indicate their source. By default, your modules are displayed in pink, modules from the Broad Institute in black, and modules from other sources in blue. Click *Help>Module Color Key* to display a legend for your current colors.

## How GenePattern Runs Analyses

GenePattern has a client/server architecture designed for flexibility and ease of use. The server holds all analysis modules, visualization modules, and pipelines. As shown in the following figure, you use the GenePattern client to run analyses on the GenePattern server:

1. From the client, you select the module that you want to use, specify the necessary parameters, and run the analysis. The client sends your request to server, where the analysis will be run with the specified parameters.
2. After sending the request, the client periodically polls the server asking whether the analysis has finished running. While the server works on the analysis, it returns “No” to the client.
3. When the server has finished running the analysis, it returns “Yes” when polled by the client.
4. The client notifies you that the analysis has finished. You can then view the analysis results, which are in temporary storage on the server (see [Saving Analysis Results](#)).



## GenePattern Jobs

When you run an analysis module, visualization module, or pipeline, the client starts a job on the GenePattern server. Viewers are the one exception: viewers run on the client, not the server.

Every job run on the GenePattern server has an owner and is persistent:

- The owner of the job is the person logged into the GenePattern client that submitted the job to the server. The owner is identified by the GenePattern user name (usually an e-mail address) that was used to log into the client. For example: you log into the Java Client using the user name john\_doe@abc.com and run an analysis; the Java Client submits a job to the server; on the server, that job is owned by john\_doe@abc.com.
- Every job is persistent, which means:
  - Jobs run independently of the client. Once you start a job on the server, the job continues to run, even if you exit from the client that started the job. The next time you start the client, the client shows you the status of any jobs that were running when you exited from the client.

- Jobs are restarted if necessary. If the GenePattern server is shut down or interrupted while executing a job, the next time you start the server, the server restarts the interrupted job.

## Running an Analysis or Visualization Module

To run an analysis or visualization module:

1. Select a module from the Analysis or Visualization menu. The parameters for the selected module appear in the main window, as shown below.

The Analysis and Visualization menus list all of the modules on your GenePattern server. If you think your server has been updated since you started the Java Client, click *File>Refresh>Modules* to update the menus.

2. Enter values for each parameter, as described below.
3. If necessary, click *Reset* to restore the parameters to their default values or *Help* to display documentation for the module.
4. Click *Run* to start the job.

When you select a module from the Analysis or Visualization menu, the parameters for the selected module appear in the main window:

**ComparativeMarkerSelection** 2

Computes significance values for features using several metrics, including FDR(BH), Q Value, maxT P-Value, FWER, Feature-Specific P-Value, and Bonferroni.

Show Parameter Descriptions

**input filename:**  Browse...  
The input file - .res, .gct, Dataset

**cls filename:**  Browse...  
The class file - .cls

confounding variable cls filename:  Browse...  
The class file containing the confounding variable.

**test direction:** 2 Sided  
The test to perform

**test statistic:** SNR  
The statistic to use

min std:   
The minimum standard deviation if test statistic is T-Test (min std)

**number of permutations:** 1000  
The number of permutations to perform

**complete:** no  
Whether to perform all possible permutations

**balanced:** no  
Whether to perform balanced permutations

**random seed:** 779948241  
The seed of the random number generator used to produce permutations

Run Reset Help

Generate Code

The parameter forms for all modules are similar:

- The form lists each parameter followed by an entry box for the parameter value.
- Below each parameter is a brief description of that parameter. Use the *Show Parameter Descriptions* check box in the upper left corner of the form to hide and show the descriptions.
- Required parameters are listed in **bold**. Other parameters are optional.

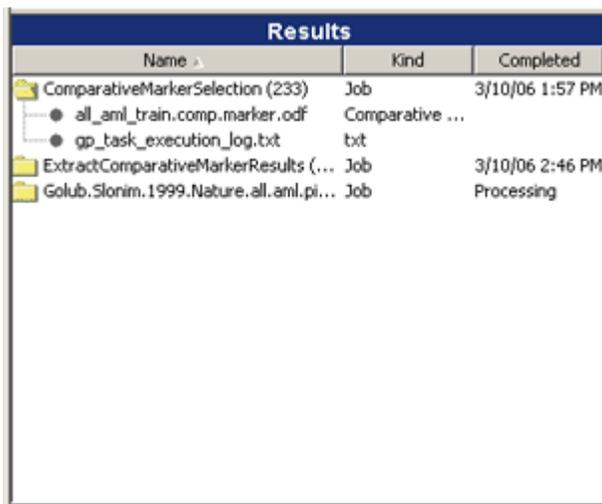
- When a parameter requires a file name, you can enter it in many ways:
  - Enter a directory name to create a batch process that runs the same analysis on every valid file in that directory. If the directory contains a file in the wrong format for this analysis, GenePattern displays a message informing you of that and moves on to the next file. You can enter a directory name for, at most, one file parameter.
  - Click-and-drag a file from the Projects or Results pane to the file parameter box. Selecting a file from the Results pane has the benefit of chaining the previously executed module to the current module, which can be useful for creating pipelines (see [Working with Pipelines](#)).
  - Right-click a file in the Projects or Results pane and select *Send To>parameter-name*.  
Alternative: select the file and select either *Projects>Send To>parameter-name* or *Results>Send To>parameter-name*.
  - Click-and-drag a file from a file browser or a web browser to the file parameter box.
  - Click *Browse* (next to the file parameter box) to display the file selection window. Navigate to the desired file, select it, and click *Open*.
  - Enter the full path name of a file in the file parameter box. This method is not recommended because file names can too easily be mistyped.

The *Generate Code* field at the bottom of the form displays the code (Java, MATLAB, or R) used to run the module with the parameters that you have entered. This can be useful for programmers writing batch procedures or new modules.

## Checking Job Status

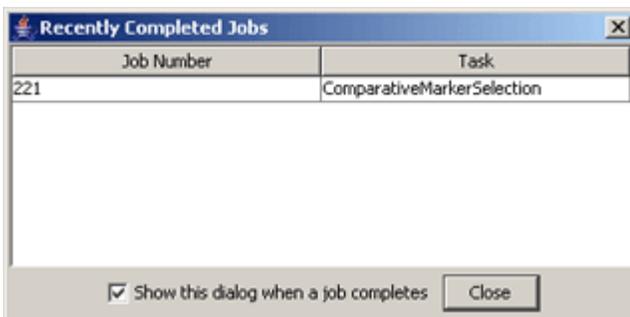
After starting an analysis job on the GenePattern server, you can check its status in two ways:

- In the Results pane, locate the job and check its Completed column. If the job is still processing, the column says processing; if it has finished, the column shows the date and time it finished.



Results		
Name	Kind	Completed
ComparativeMarkerSelection (233)	Job	3/10/06 1:57 PM
all_aml_train.comp.marker.odf	Comparative ...	
gp_task_execution_log.txt	txt	
ExtractComparativeMarkerResults (...)	Job	3/10/06 2:46 PM
Golub.Slonim.1999.Nature.all.aml.pi...	Job	Processing

- By default, when a job finishes, GenePattern adds it to the Recently Completed Jobs window. If the window is not open, GenePattern opens it. The window accumulates finished jobs until you close it, which can be useful if you have several jobs running at once or you want a log of recent jobs.



Recently Completed Jobs	
Job Number	Task
221	ComparativeMarkerSelection

Show this dialog when a job completes Close

Use *File>Alert On Job Completion* to control whether GenePattern updates the Recently Completed Jobs window when a job completes. If the menu item has a check mark, GenePattern updates the window, opening it if necessary; otherwise, it does not. Select the menu item to toggle the check mark.

## Viewing Analysis Results

When you run an analysis or visualization module, the analysis results are stored on the GenePattern server and displayed in the Results pane of the application window, as described in [Working with Analysis Results](#).

You can view an analysis results file in many ways:

- **Viewer:** select the viewer from the Visualization menu and run it as described in [Running an Analysis or Visualization Module](#). Viewers, which allow you to view your data graphically and to manipulate that view interactively, are generally the best way to view analysis results. By convention, viewer names end in "Viewer;" for example, HeatMapView.
- **Image creator:** select the image creator from the Visualization menu and run it as described in [Running an Analysis or Visualization Module](#). Image creators, which create static graphics for view in other applications, are often useful for documenting or sharing analysis results. By convention, image creator names end in "Image;" for example, HeatMapImage.
- **Text Viewer:** right-click the file and select *Open With>Text Viewer* to display the file in a text viewer. This is generally the best way to view log files.
- **Default Application:** right-click the file and select *Open With>Default Application* to display the file in the default application (if any) associated with this file type. This can be useful when you have assigned default applications to particular file types. For information about associating an application with a file type, see your operating system documentation.

## Rerunning Analyses

To rerun an analysis:

1. Reload the previously run analysis in one of two ways:
  - In the Results pane, click the analysis and select *Results>Reload*.  
Alternative: right-click (Mac users Option-click) the analysis and select *Reload*.
  - Select *History>View All*. When GenePattern displays the History window, select an analysis and click *Reload*.

The Java Client displays the analysis module in the main window, populating the parameter fields with the values used to run the previous analysis.

2. Optionally, modify the parameters.
3. Click *Run* to run the analysis.

Another way to rerun an analysis or series of analyses is to build an analysis pipeline. For more information, see [Working with Pipelines](#).

## Working with Pipelines

Pipelines allow you to combine analysis and visualization modules into a single workflow. Running a pipeline runs the series of tasks defined by that pipeline. Since pipelines are easily exported and imported from GenePattern, you can share them with colleagues who would like to reproduce your analysis results or use them in documenting your research. By providing a way to create and distribute an entire computational analysis methodology in a single executable script, pipelines enable a form of *in silico* reproducible research.

This section describes pipelines and how to use them:

- [About Pipelines](#)
- [Running a Pipeline](#)
- [Viewing a Pipeline](#)
- [Creating a Pipeline](#)
- [Editing a Pipeline](#)
- [Exporting and Importing Pipelines](#)

**Note:** You can delete pipelines from the Web Client, but not from the Java Client. For more information, see the [GenePattern Web Client Guide](#).

## About Pipelines

This section provides background information about pipelines:

- [What Are Pipelines?](#)
- [How GenePattern Runs Pipelines](#)

## What Are Pipelines?

A GenePattern pipeline captures a workflow. Each GenePattern pipeline contains one or more tasks. Each task identifies an analysis or visualization module and its parameter values. When you run the pipeline, GenePattern runs each task defined in the pipeline, one after the other. As each task completes, GenePattern adds its analysis results files to the Results pane. When all the tasks are finished, the pipeline is finished.

GenePattern offers significant flexibility in how you define the parameter values for each task:

- **Static:** you can specify parameter values.
- **Variable:** you can have GenePattern prompt the user for parameter values each time the pipeline is run.
- **Chained:** you can use an output file from one task as the input parameter value for a subsequent task.

This flexibility means that you can create a pipeline to run an analysis methodology against a particular data set; reproduce an analysis methodology to be run against variable data sets; run multiple analyses against a single data file; or run a progressive series of analyses, where the output from one analysis is used as input for the next.

## How GenePattern Runs Pipelines

The Pipeline menu lists all of the pipelines available to you; that is, all of the pipelines on your GenePattern server. You can add pipelines to your server from the Java Client or the Web Client:

- From the Java Client: create your own pipelines or import them from a zip file.
- From the Web Client: create your own pipelines, import them from a zip file, or install pipelines from the module repository maintained by the Broad Institute. For more information about the Web Client, see the [GenePattern Web Client Guide](#).

Pipeline names are color coded to indicate the source of the pipeline. By default, your pipelines are displayed in pink, pipelines from the Broad Institute in black, and pipelines from other sources in blue. Click *Help>Module Color Key* to display a legend for your current colors.

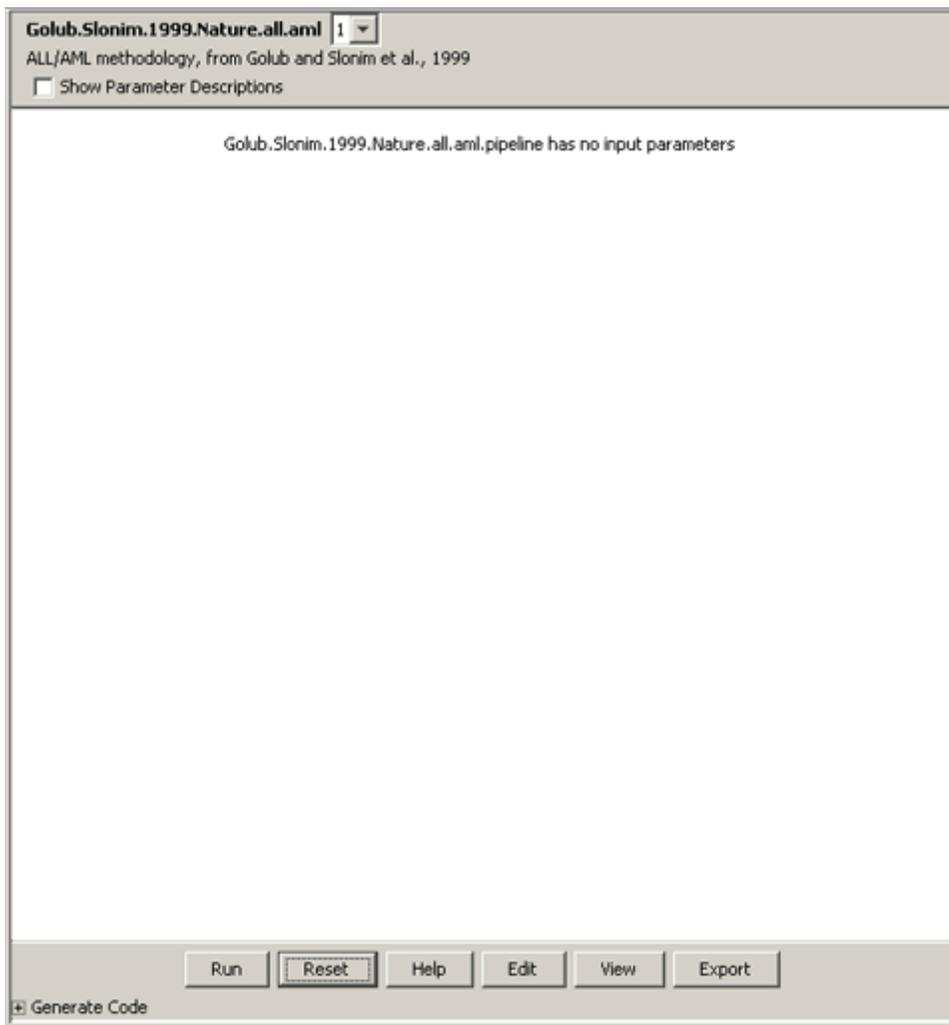
When you run a pipeline, the GenePattern client starts a job on the GenePattern server. All of the tasks in the pipeline are run as part of that one job. For more information about how the client runs jobs on the server, see [About Analysis and Visualization Modules](#).

## Running a Pipeline

To run a pipeline:

1. Select the pipeline from the Pipelines menu. The pipeline and its parameters (if any) appear in the main window, as shown below.

The Pipelines menu lists all of the pipelines on your GenePattern server. If you think your server has been updated since you started the Java Client, click *File>Refresh>Modules* to update the menu.



2. Enter any parameters, if necessary.
3. Click *Run*.

GenePattern adds the pipeline to the Results pane; the Completed column shows a status of processing. As each task completes, GenePattern adds its analysis results files to the Results pane. When the last task completes, GenePattern writes the pipeline execution log files to the Results pane and updates the Completed column to show the date and time that the pipeline finished.

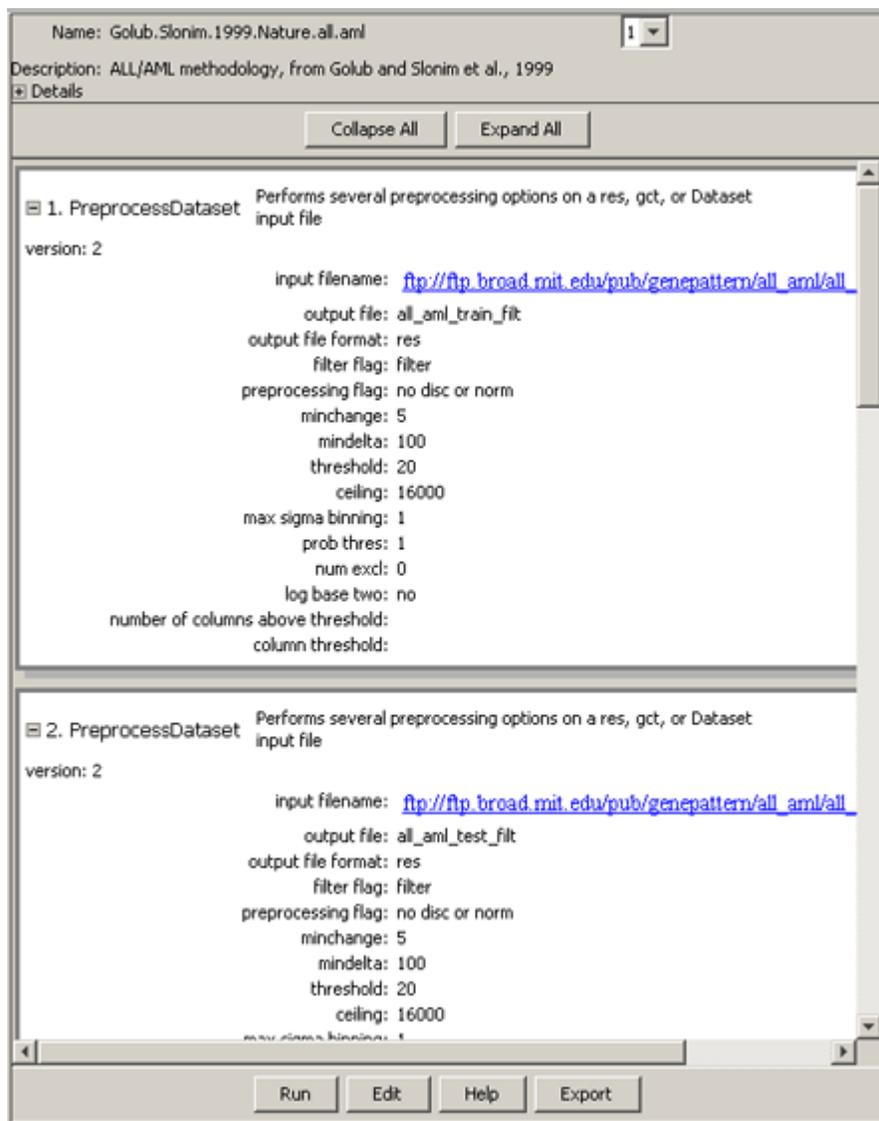
Analysis and visualization modules run from a pipeline work exactly the same as those run directly from the Java Client. If you run a Viewer from a pipeline, it gets displayed just as if you ran it directly from the Java Client. Analysis results files generated by a pipeline are exactly the same as those generated by modules run directly from the Java Client.

## ***Viewing a Pipeline***

Viewing a pipeline shows you its definition, including the tasks it runs and the parameters for each task. You might want to view a pipeline before running it for the first time.

To view a pipeline:

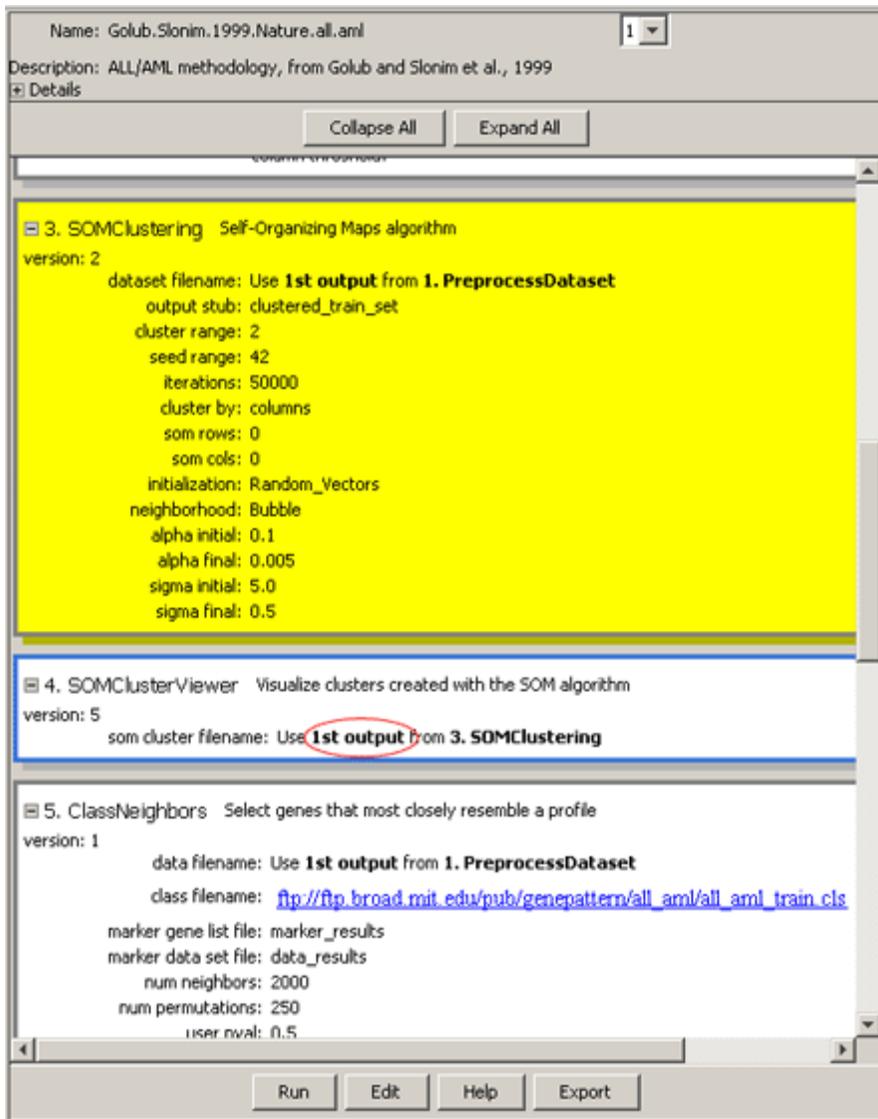
1. Select the pipeline from the Pipelines menu. The pipeline and its parameters (if any) appear in the main window.
2. Click *View*. The pipeline definition appears in the main window, as shown below.



The definition form for each pipeline is similar:

- The first line shows the name of the pipeline and a drop-down list of versions. By default, you are viewing the most recent version of the pipeline. To view a different version, select that version from the drop-down list.
- The second line is a brief description of the pipeline.
- To view author information and documentation for the pipeline, expand the *Details* field by clicking the plus (+) sign.
- Below the *Details* field is a scrolling window that lists each task in the pipeline, including its parameters. You can hide/show parameter details for each task individually (plus/minus icon to the left of each task) or for all tasks at once (*Collapse All/Expand All* buttons at the top of the window).
- Chained parameters allow you to take the output from one task and use it as the input for another task. If you place your cursor on a chained input parameter, GenePattern highlights the task that generates this input file, as shown below. (You only see this highlighting if both tasks are visible in the window.)
- The buttons at the bottom of the window allow you to run, edit, or export the pipeline to a zip file. *Help* displays the documentation for the pipeline.

For more information about the fields on the definition form, see [Editing a Pipeline](#).



## Creating a Pipeline

You can create a pipeline in two ways:

- Use *Pipelines>New* to create an empty pipeline.
- Use *Results>Create Pipeline* to create a pipeline that contains the tasks used to generate the selected analysis results file.

To create an empty pipeline:

1. Select *Pipelines>New*. GenePattern displays an empty pipeline definition form in the main window.
2. Enter a name for the pipeline in the *Name* field.
3. Edit the fields at the top of the form and add tasks to pipeline, as described in [Editing a Pipeline](#).

To create a pipeline based on an analysis results file:

1. Select an analysis results file from the Results pane and then select *Results>Create Pipeline*. GenePattern displays the following window prompting you for a pipeline name.

Alternative: right-click (Mac users Option-click) an analysis results file and select *Create Pipeline*.



2. Enter a name for the pipeline and click *OK*. GenePattern displays a pipeline definition form that contains the name you entered and the tasks used to create the analysis results file.

GenePattern adds tasks to the pipeline based on the following logic: add the module that created the results file and then check its input file; if the input file for the module was the output file of a previous module, add the previous module and then check its input file; continue to walk back through the chain of modules, adding modules to the pipeline, until you reach the initial input file.

3. Edit the fields at the top of the form and add additional tasks to pipeline, as described in [Editing a Pipeline](#).

## Editing a Pipeline

You can edit any pipeline listed in the Pipelines menu, regardless of whether you created it. Editing a pipeline creates a new version of the pipeline; it does not change the existing version of the pipeline. For more information, see [Understanding Version Numbers](#).

To edit a pipeline:

1. Select the pipeline from the Pipelines menu.
2. Click *Edit*. The pipeline definition appears in the main window.
3. Edit the pipeline definition, as described below.
4. Click *Save* to create a new version of the pipeline. GenePattern creates a new version of the pipeline.
5. When the confirmation message appears, click *OK*.

The remaining topics in this section describe how to edit the pipeline definition:

- [Editing the Pipeline Description](#)
- [Adding Tasks](#)
- [Removing Tasks](#)
- [Reordering Tasks](#)

## Editing the Pipeline Description

The fields at the top of the pipeline definition form describe the pipeline. If necessary, expand the *Details* field by clicking the plus (+) sign.

- **Name:** The name of the pipeline. If you change the name of the pipeline, you create a new pipeline. If you leave the name unchanged, you create a new version of the existing pipeline. When naming pipelines, note the following:
  - Pipeline names can include alphanumeric characters, periods (.), and underscores (\_).
  - Avoid special characters in pipeline names, such as: exclamation points (!), at signs (@), pound signs (#), dollar signs (\$), percent signs (%), carets (^), ampersands (&), and asterisks (\*).
  - For cross-platform compatibility, avoid the following names: con, prn, aux, nul, com1, com2, com3, com4, lpt1, lpt2, and lpt3. Machines running Windows cannot accept files with these names, regardless of the file extension.
  - GenePattern does not prevent you from using the same name for multiple pipelines; however, using unique names is strongly recommended.
- **Version:** To the right of the name field is a drop-down list of versions. By default, you are editing the most recent version of the pipeline. To edit a different version, select that version from the drop-down list.

Editing a pipeline creates a new version of the pipeline; it does not change the existing version of the pipeline. For more information, see [Understanding Version Numbers](#).

- **Description:** A brief description of the pipeline. When you select a pipeline from the Pipelines menu, GenePattern displays the pipeline name and this brief description in the main window.

- *Author*: The name and organization of the pipeline's author. By default, this is your user name (typically, your e-mail address).
- *Owner*: The user name of the person responsible for maintaining the pipeline. By default, this is your user name (typically, your e-mail address).
- *Privacy*: Select Private (default) or Public. A private pipeline can only be seen and run by a user logged in using the user name specified in the *Owner* field of this form; a public pipeline can be seen and run by all users.
- *Version comment*: A brief description of this version. This comment is visible on pages used to keep versions up to date, so keep your comments complete and concise.
- *Documentation*: The full path name of the file that contains the pipeline documentation. Documentation is strongly encouraged for public pipelines. Click *Add* to select the (previously created) pipeline documentation file. Click *Delete* to clear this field.
- *LSID*: The Life Science Identifiers (LSIDs) for this pipeline. GenePattern maintains this field; you cannot edit it. For more information, see [Understanding Version Numbers](#).

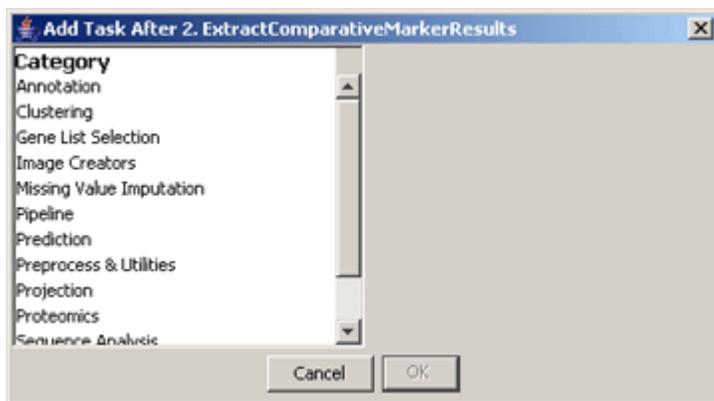
## Adding Tasks

Below the *Details* field is a scrolling window that defines the tasks in the pipeline. You can hide and show parameter details for each task individually (plus/minus icon to the left of each task) or for all tasks at once (*Collapse All/Expand All* buttons at the top of the window).

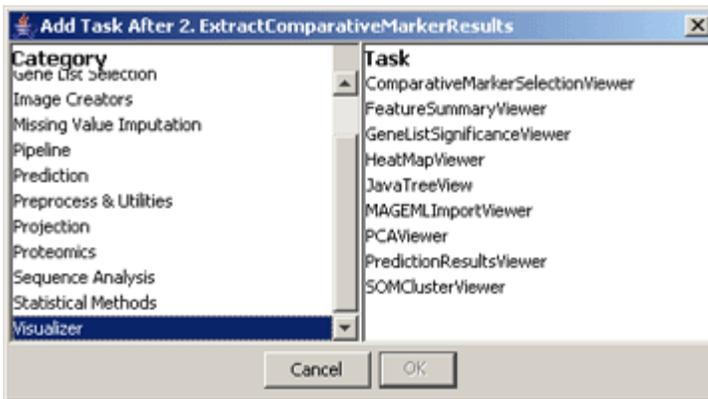
To add a task to the pipeline:

1. Select *Add Task*, *Add Task After*, or *Add Task Before*:
  - *Add Task* appears only for empty pipelines. Use this button to add the first task to the pipeline.
  - *Add Task After* adds a task below the currently selected task. To select a task, click its name in the drop-down list to the left of the *Add Task After* button or click its definition in the pane below the buttons.
  - *Add Task Before* adds a task above the currently selected task.

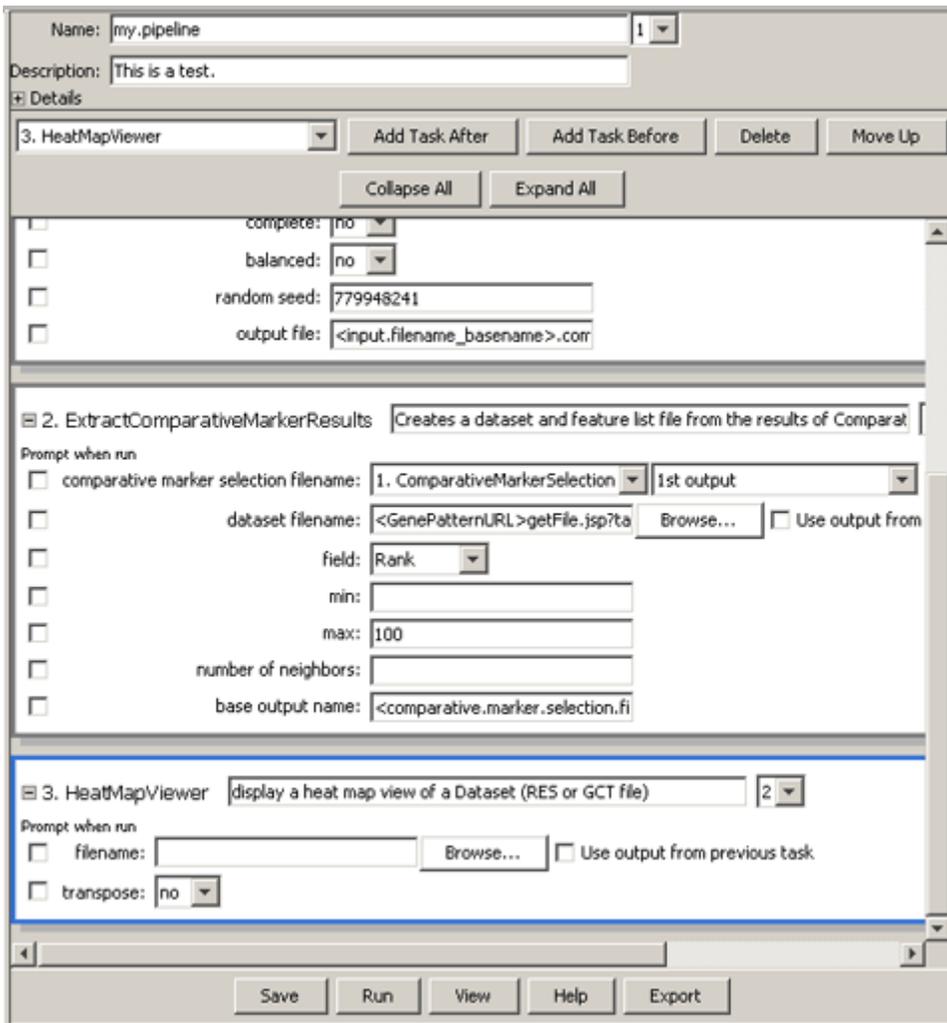
GenePattern displays the Add Task window.



2. Select a category from the Category list. GenePattern populates the Task list with the modules in the selected category. The categories are the same as those on the Analysis menu, with the addition of Visualizer, which populates the Task list with the Visualization modules.



3. Select a task from the Tasks list and click *OK*. GenePattern adds the selected task to the pipeline definition.



4. The first line of the task definition panel shows the module name and version number. By default, you are using the most recent version of the module. To have the pipeline run a different version of the module, select that version from the drop-down list.
5. Enter the parameter values for the module. For most parameters, you enter a value, select a value from a drop-down list, or use the default value supplied by GenePattern.

To have GenePattern prompt the user for a parameter value, select the *Prompt when run* box to the left of the parameter field.

For input file parameters, you can enter a file name in many ways:

- Drag-and-drop a file from the Projects or Results pane to the parameter field.

- Click the *Browse* button to the right of the parameter field to select a file from the File Selection window.
- To have the input file be an output file generated by a previous task, select the *Use output from previous task* box to the right of the parameter field. The parameter field changes to a drop-down list of the previous tasks. Select the task that generates the output file that you want to use. If the selected task generates multiple output files, GenePattern displays a drop-down list of the generated output files. Select the output file that you want to use.

## Removing Tasks

To remove a task from the pipeline:

1. Select the task that you want to delete. To select a task, click its name in the drop-down list to the left of the *Add Task After* button or click its definition in the pane below the buttons.
2. Click the *Delete* button.

## Reordering Tasks

The order of the tasks in the pipeline definition form determines the order in which the tasks are executed when you run the pipeline. This can be important, for example, when you want to use the output from one task as the input for a subsequent task or when you want to preprocess a data file before analyzing it.

To change the order of tasks:

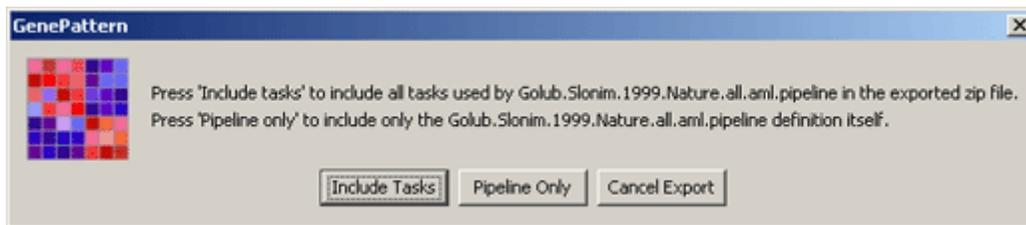
1. Select the task that you want to move. To select a task, click its name in the drop-down list to the left of the *Add Task After* button or click its definition in the pane below the buttons.
2. Click the *Move Up* or *Move Down* button.

## Exporting and Importing Pipelines

Exporting and importing pipelines allows you to easily share them with colleagues who have access to GenePattern.

To export a pipeline:

1. Select the pipeline from the Pipelines menu. The pipeline appears in the main window.
2. Click *Export*. The export confirmation window appears:



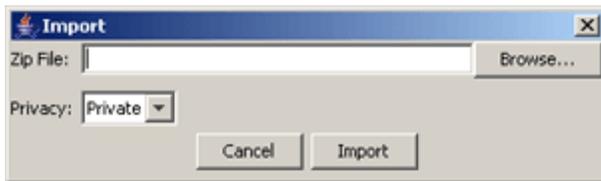
3. Choose whether to export the pipeline or the pipeline and all of its tasks:
  - Click *Pipeline Only* to create a zip that contains a programmatic description of the pipeline and the pipeline documentation.
  - Click *Include Tasks* to create a zip file that contains multiple zip files: the pipeline zip file (as described above) and a zip file for each task. The zip file for each task contains the program files that implement the task and the task documentation.

GenePattern displays the file selection window.

4. Select a name and directory for the zip file and click *Open*.  
GenePattern exports the pipeline to the zip file.

To import a pipeline:

1. Select *File>Import Module*. The Import window appears:



2. Click *Browse*. The File Selection window appears.
3. Navigate to the directory that contains the zip file, select the zip file, and click *Open*. GenePattern updates the Import window to show the selected zip file.
4. Click *Import* to import the pipeline from the zip file.

## Working with Suites

Use suites to group modules into packages of analysis modules, visualization modules, and pipelines; for example, the ClusteringSuite contains the ConsensusClustering, HierarchicalClustering, HierarchicalClustering.MATLAB, NMFClustering, and SOMClustering analysis modules. Suites allow you to work with groups of modules; for example, you can filter the Java Client menus to show only the modules in the ClusteringSuite.

The Suites menu lists all of the suites available to you; that is, all of the suites on your GenePattern server. You can add suites to your server from the Java Client or the Web Client:

- From the Java Client: create your own suites or import them from a zip file.
- From the Web Client: create your own suites, import them from a zip file, or install them from the module repository maintained by the Broad Institute. For more information about the Web Client, see the [GenePattern Web Client Guide](#).

This section describes how to use suites:

- [Viewing or Editing a Suite](#)
- [Creating a Suite](#)
- [Using Filters](#)
- [Exporting and Importing Suites](#)

**Note:** You can delete suites from the Web Client, but not from the Java Client. For more information, see the [GenePattern Web Client Guide](#).

### Viewing or Editing a Suite

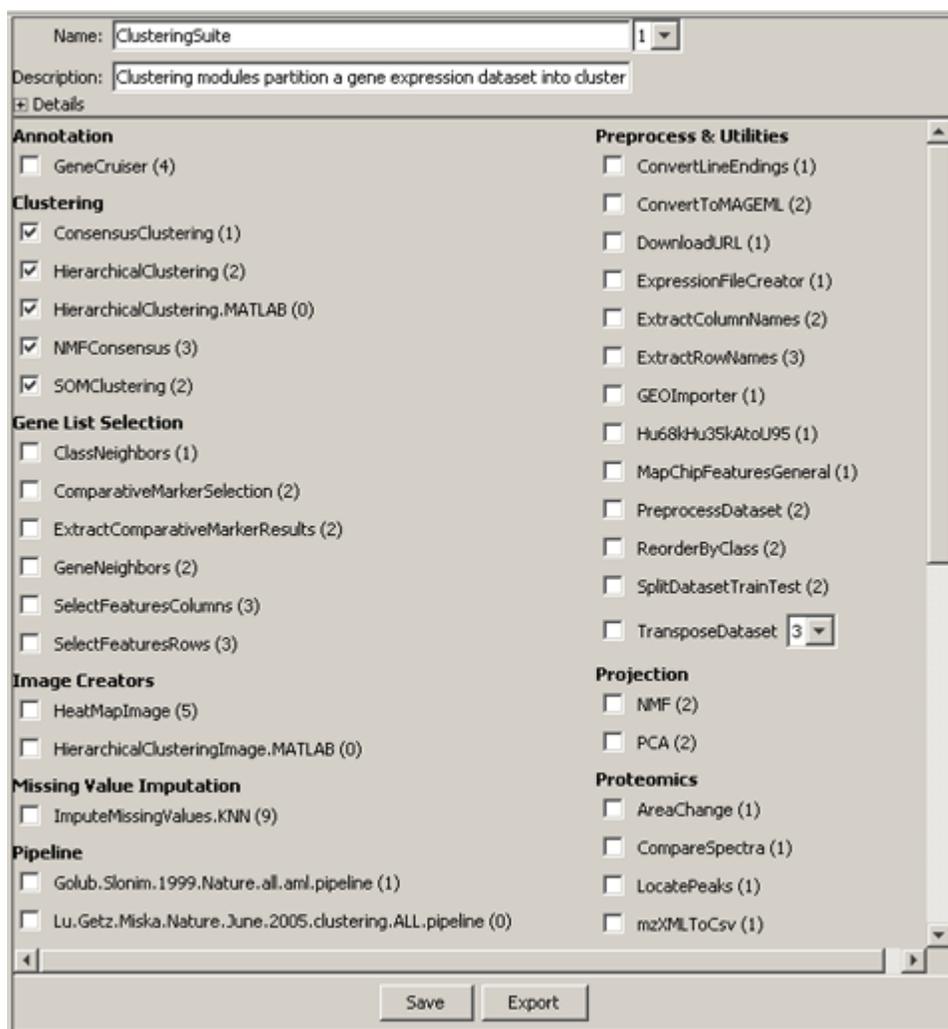
You can view or edit any suite listed in the Suites menu, regardless of whether you created it. Editing a suite creates a new version of the suite; it does not change the existing version. For more information, see [Understanding Version Numbers](#).

To view or edit a suite:

1. Select the suite from the Suites menu. The menu lists all of the suites on your GenePattern server. If you think your server has been updated since you started the Java Client, click *File>Refresh>Modules* to update the menu.

The suite definition appears in the main window, as shown below. The definition form lists all of the analysis modules, visualization modules, and pipelines that are on your GenePattern server. A check mark next to a module or pipeline indicates that it is in this suite.

2. To edit the suite definition:
  1. Edit the suite definition form.
  2. Click *Save*. GenePattern creates a new version of the suite.
  3. When the confirmation message appears, click *OK*.



All suites use a similar definition form:

- The first line shows the name of the suite and a drop-down list of versions. By default, you are viewing the most recent version of the suite. To view a different version, select that version from the drop-down list.
- The second line is a brief description of the suite.
- To view author information and documentation for the suite, expand the *Details* field by clicking the plus (+) sign. The details fields for suites are similar to those for pipelines; for descriptions, see [Editing the Pipeline Description](#).
- Below the *Details* field is a scrolling window that lists all of the analysis modules, visualization modules, and pipelines on your GenePattern server. A check mark next to a module or pipeline indicates that it is in this suite. If you are editing the suite, toggle the check marks by clicking the check boxes.
- The buttons at the bottom of the window allow you to save your changes or export the suite to a zip file.

## Creating a Suite

To create a suite:

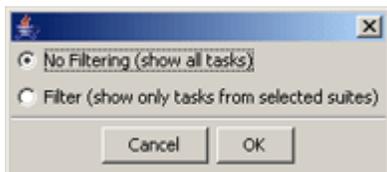
1. Select *Suites>New*. An empty suite definition appears in the main window.
2. Enter a suite name and description in the *Name* and *Description* fields.
3. Edit the suite definition, as described in [Viewing or Editing a Suite](#).
4. Click *Save*. GenePattern creates version 1 of the suite.
5. When the confirmation message appears, click *OK*.

## Using Filters

You can use suites to filter the Java Client menus so that they list only the modules in selected suites.

To apply a filter:

1. Select *Suites>Filter*. The Java Client displays the Filter window:



2. Select the *Filter* radio button. The Java Client updates the window to show the available suites:



3. Click the check box next to each suite that you want to use.
4. Click *OK*.

The Java Client updates the menus to show only the analysis modules, visualization modules, and pipelines in the selected suites.

To remove the filter from the menus:

1. Select *Suites>Filter*. The Java Client displays the Filter window.
2. Select the *No Filtering* radio button.
3. Click *OK*.

The Java Client updates the menus to show all of the analysis modules, visualization modules, and pipelines that are on your GenePattern server.

## Exporting and Importing Suites

Exporting and importing suites allows you to easily share them with colleagues who have access to GenePattern.

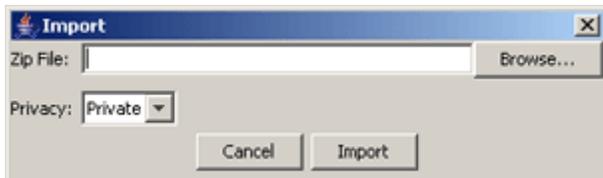
To export a suite:

1. Select the suite from the Suites menu. The suite appears in the main window.
2. Click *Export*. GenePattern displays the file selection window.
3. Select a name and directory for the zip file and then click *Open*.

GenePattern exports the suite to the zip file.

To import a suite:

1. Select *File>Import Suite*. The Import window appears:



2. Click *Browse*. The File Selection window appears.
3. Navigate to the directory that contains the zip file, select the zip file, and click *Open*. GenePattern updates the Import window to show the selected zip file.
4. Click *Import* to import the suite from the zip file.

## Creating Input Files

When you run an analysis module, visualization module, or pipeline, GenePattern displays the parameters for the selected modules. Often, one or more of these parameters are input files, which must have a particular format; for example, you might need to supply a gct or res file. For descriptions of the different GenePattern file formats, see [File Formats](#).

Although different GenePattern modules require different file formats, all of the files are tab-delimited or space-delimited text files. Most of your gene expression data is already in tab-delimited text files or in spreadsheet and database programs, which have export features that allow you to export the data into tab-delimited text files. Therefore, creating input files for GenePattern is relatively easy:

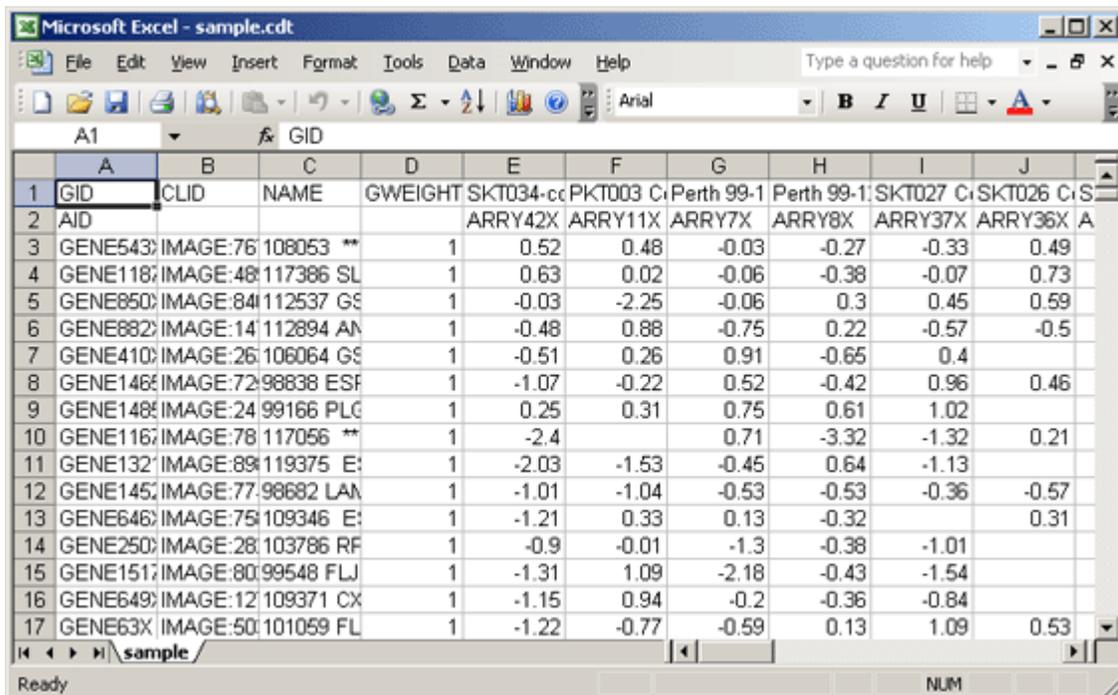
1. Start with a tab-delimited text file that contains the required gene expression data.
2. Open the file in a text editor (or spreadsheet editor).
3. Make the necessary format changes.
4. Save the file as a tab-delimited text file with the appropriate file extension.

### Tutorial: Creating a gct File

This section provides a brief tutorial that walks you through the process of converting a cdt file ([sample.cdt](#)) to a gct file ([sample.imputed.gct](#)).

To convert the cdt file to a gct file:

1. Save the ([sample.cdt](#)) file to your local drive and open it in Microsoft Excel.



	A	B	C	D	E	F	G	H	I	J
1	GID	CLID	NAME	GWEIGHT	SKT034-cc	PKT003 C	Perth 99-1	Perth 99-1	SKT027 C	SKT026 C
2	AID				ARRY42X	ARRY11X	ARRY7X	ARRY8X	ARRY37X	ARRY36X A
3	GENE543	IMAGE:76	108053 **	1	0.52	0.48	-0.03	-0.27	-0.33	0.49
4	GENE118	IMAGE:48	117386 SL	1	0.63	0.02	-0.06	-0.38	-0.07	0.73
5	GENE850	IMAGE:84	112537 GS	1	-0.03	-2.25	-0.06	0.3	0.45	0.59
6	GENE882	IMAGE:14	112894 AN	1	-0.48	0.88	-0.75	0.22	-0.57	-0.5
7	GENE410	IMAGE:26	106064 GS	1	-0.51	0.26	0.91	-0.65	0.4	
8	GENE146	IMAGE:72	98838 ESP	1	-1.07	-0.22	0.52	-0.42	0.96	0.46
9	GENE148	IMAGE:24	99166 PLC	1	0.25	0.31	0.75	0.61	1.02	
10	GENE116	IMAGE:78	117056 **	1	-2.4		0.71	-3.32	-1.32	0.21
11	GENE132	IMAGE:89	119375 E	1	-2.03	-1.53	-0.45	0.64	-1.13	
12	GENE145	IMAGE:77	98682 LAN	1	-1.01	-1.04	-0.53	-0.53	-0.36	-0.57
13	GENE646	IMAGE:75	109346 E	1	-1.21	0.33	0.13	-0.32		0.31
14	GENE250	IMAGE:28	103786 RF	1	-0.9	-0.01	-1.3	-0.38	-1.01	
15	GENE151	IMAGE:80	99548 FLJ	1	-1.31	1.09	-2.18	-0.43	-1.54	
16	GENE649	IMAGE:12	109371 CX	1	-1.15	0.94	-0.2	-0.36	-0.84	
17	GENE63X	IMAGE:50	101059 FL	1	-1.22	-0.77	-0.59	0.13	1.09	0.53

2. Delete the CLID and GWEIGHT columns. The gct file format allows for only two columns of annotations.

1	A	B	C	D	E	F	G	H	I	J
1	GID	NAME	SKT034-cc	PKT003 C	Perth 99-1	Perth 99-1	SKT027 C	SKT026 C	SKT023 C	SKT011 C
2	AID		ARRAY42X	ARRAY11X	ARRAY7X	ARRAY8X	ARRAY37X	ARRAY36X	ARRAY34X	ARRAY26X
3	GENE543	108053 **	0.52	0.48	-0.03	-0.27	-0.33	0.49	0.16	0.27
4	GENE118	117386 SL	0.63	0.02	-0.06	-0.38	-0.07	0.73	0.2	0.23
5	GENE850	112537 GS	-0.03	-2.25	-0.06	0.3	0.45	0.59	1.17	
6	GENE882	112894 AN	-0.48	0.88	-0.75	0.22	-0.57	-0.5	-0.47	-0.24
7	GENE410	106064 GS	-0.51	0.26	0.91	-0.65	0.4		-0.32	-0.21
8	GENE146	98838 ESF	-1.07	-0.22	0.52	-0.42	0.96	0.46	-0.46	-0.18
9	GENE148	99166 PLC	0.25	0.31	0.75	0.61	1.02		-0.3	-0.69
10	GENE116	117056 **	-2.4		0.71	-3.32	-1.32	0.21	-1.84	0.27
11	GENE132	119375 E	-2.03	-1.53	-0.45	0.64	-1.13		-1.63	-0.42
12	GENE145	98682 LAN	-1.01	-1.04	-0.53	-0.53	-0.36	-0.57	-0.88	-1.32
13	GENE646	109346 E	-1.21	0.33	0.13	-0.32		0.31	-1.22	-0.32
14	GENE250	103786 RF	-0.9	-0.01	-1.3	-0.38	-1.01		-1.6	-0.28
15	GENE151	99548 FLJ	-1.31	1.09	-2.18	-0.43	-1.54		-0.9	0.53
16	GENE649	109371 CX	-1.15	0.94	-0.2	-0.36	-0.84		0.05	0.45
17	GENE63X	101059 FL	-1.22	-0.77	-0.59	0.13	1.09	0.53	-0.21	-0.63

3. Delete the second row, which contains array identifiers (AID). The gct file format allows for only one row of identifiers.

1	A	B	C	D	E	F	G	H	I	J
1	GID	NAME	SKT034-cc	PKT003 C	Perth 99-1	Perth 99-1	SKT027 C	SKT026 C	SKT023 C	SKT011 C
2	GENE543	108053 **	0.52	0.48	-0.03	-0.27	-0.33	0.49	0.16	0.27
3	GENE118	117386 SL	0.63	0.02	-0.06	-0.38	-0.07	0.73	0.2	0.23
4	GENE850	112537 GS	-0.03	-2.25	-0.06	0.3	0.45	0.59	1.17	
5	GENE882	112894 AN	-0.48	0.88	-0.75	0.22	-0.57	-0.5	-0.47	-0.24
6	GENE410	106064 GS	-0.51	0.26	0.91	-0.65	0.4		-0.32	-0.21
7	GENE146	98838 ESF	-1.07	-0.22	0.52	-0.42	0.96	0.46	-0.46	-0.18
8	GENE148	99166 PLC	0.25	0.31	0.75	0.61	1.02		-0.3	-0.69
9	GENE116	117056 **	-2.4		0.71	-3.32	-1.32	0.21	-1.84	0.27
10	GENE132	119375 E	-2.03	-1.53	-0.45	0.64	-1.13		-1.63	-0.42
11	GENE145	98682 LAN	-1.01	-1.04	-0.53	-0.53	-0.36	-0.57	-0.88	-1.32
12	GENE646	109346 E	-1.21	0.33	0.13	-0.32		0.31	-1.22	-0.32
13	GENE250	103786 RF	-0.9	-0.01	-1.3	-0.38	-1.01		-1.6	-0.28
14	GENE151	99548 FLJ	-1.31	1.09	-2.18	-0.43	-1.54		-0.9	0.53
15	GENE649	109371 CX	-1.15	0.94	-0.2	-0.36	-0.84		0.05	0.45
16	GENE63X	101059 FL	-1.22	-0.77	-0.59	0.13	1.09	0.53	-0.21	-0.63
17	GENE968	114119 FA	-1.17	-0.8	-0.92	-0.94	-1.75	0.35	-0.88	-0.81

4. Add two header rows at the top of the file:
  - In the first row, first cell, enter: #1.2
  - In the second row, first cell, enter the number of data rows: 1553
  - In the second row, second cell, enter the number of data columns: 44

	A	B	C	D	E	F	G	H	I	J
1	#1.2									
2	1553	44								
3	GID	NAME	SKT034-cc	PKT003 C	Perth 99-1	Perth 99-1	SKT027 C	SKT026 C	SKT023 C	SKT011 C
4	GENE543	108053 **	0.52	0.48	-0.03	-0.27	-0.33	0.49	0.16	0.27
5	GENE118	117386 SL	0.63	0.02	-0.06	-0.38	-0.07	0.73	0.2	0.23
6	GENE850	112537 G	-0.03	-2.25	-0.06	0.3	0.45	0.59	1.17	
7	GENE882	112894 AN	-0.48	0.88	-0.75	0.22	-0.57	-0.5	-0.47	-0.24
8	GENE410	106064 G	-0.51	0.26	0.91	-0.65	0.4		-0.32	-0.21
9	GENE146	98838 ES	-1.07	-0.22	0.52	-0.42	0.96	0.46	-0.46	-0.18
10	GENE148	99166 PL	0.25	0.31	0.75	0.61	1.02		-0.3	-0.69
11	GENE116	117056 **	-2.4		0.71	-3.32	-1.32	0.21	-1.84	0.27
12	GENE132	119375 E	-2.03	-1.53	-0.45	0.64	-1.13		-1.63	-0.42
13	GENE145	98682 LAN	-1.01	-1.04	-0.53	-0.53	-0.36	-0.57	-0.88	-1.32
14	GENE646	109346 E	-1.21	0.33	0.13	-0.32		0.31	-1.22	-0.32
15	GENE250	103786 RF	-0.9	-0.01	-1.3	-0.38	-1.01		-1.6	-0.28
16	GENE151	99548 FL	-1.31	1.09	-2.18	-0.43	-1.54		-0.9	0.53
17	GENE649	109371 CX	-1.15	0.94	-0.2	-0.36	-0.84		0.05	0.45

5. Save the modified file as a text (tab delimited) file with the name `sample.gct`.
6. Your original `cdt` file contained cells that were missing data. Most GenePattern modules require that all cells in a `gct` file contain data. In the Java Client, use the GenePattern analysis module `ImputeMissingValues.KNN` to add the missing data to your `gct` file:
  1. Select *Analysis>Missing Value Imputation>ImputeMissingValues.KNN*. The Java Client displays the parameters for the selected module.
  2. For the *data filename* parameter, enter your `sample.gct` file.

**ImputeMissingValues.KNN** 9

Imputes missing values (Windows only)

Show Parameter Descriptions

**data filename:** C:\sample.gct Browse...  
Data file (missing values are stored as NA) - .gct

**k:** 10  
Number of neighbors to be used in the imputation

**rowmax:** 0.5  
The maximum percent missing data allowed in any row

**colmax:** 0.8  
The maximum percent missing data allowed in any column

**output file:** |filename\_basename|.imputed.gct  
The name of the output file - .gct

Run Reset Help

Generate Code

3. Click *Run* to run the analysis.

The analysis results file, `sample.imputed.gct`, can be used with any GenePattern module that requires a `gct` input file.

# Understanding Version Numbers

GenePattern uses Life Science Identifiers (LSIDs) to uniquely identify objects, such as analysis modules, visualization modules, pipelines, and suites. When you create an object, GenePattern automatically assigns the object an LSID with a version number of one (1). When you update an object, GenePattern automatically updates the version number of the object's LSID. As an example, consider pipeline version numbers.

When you create a pipeline, GenePattern assigns it a new LSID. If you modify that pipeline, GenePattern increments the version number of its LSID. Similarly, each task within the pipeline has its own LSID. If you modify the task, GenePattern increments the version number of its LSID. By using LSIDs, GenePattern preserves the originally created pipeline and every modified version of that pipeline. This careful versioning process allows you to accurately reproduce any previous analysis, even if you have modified the pipeline used for that analysis.

When you view and edit pipelines, you can see the pipeline's version number (the version number of the pipeline's LSID). Typically, you update the latest version of a pipeline, which increments its version number. For example, editing version 1 of a pipeline creates version 2 of that pipeline. At times, you may need to edit an older version of a pipeline, which creates a "point version" of that pipeline. For example, if you have versions 1 and 2 of a pipeline, editing version 1 of the pipeline creates version 1.1 of that pipeline.

## Menu Bar

This section briefly describes each item in the GenePattern menu bar.

### File

- *Open Project Directory*: Adds a directory to the Projects pane. For more information, see [Opening a Project Directory](#).
- *Import Module*: Imports an analysis module, a visualization module, or a pipeline module. For more information, see [Running Analysis and Visualization Modules](#) and [Working with Pipelines](#).
- *Import Suite*: Imports a suite. For more information, see [Working with Suites](#).
- *Alert On Job Completion*: If checked, GenePattern displays a pop-up window when a job completes. Select the menu item to toggle the check mark.
- *Server*: Displays the Server Settings window, which identifies the GenePattern server on which your jobs are running. To select a different GenePattern server, enter the information for that server (the URL, port, and your user name for that server) and click *OK*. You are now using that server; you do not need to restart the Java Client.
- *Refresh*: Updates the GenePattern Java Client to reflect changes that have been made on the GenePattern server. If analysis modules, visualization modules, pipelines, or suites have been installed, imported, or removed, refresh Modules; if analysis jobs have been run or analysis results files have been deleted, refresh Jobs.
- *Quit*: Exits the Java Client.

### Projects

For information about projects, see [Working with Projects](#).

- *Refresh*: Refreshes the selected project directory.
- *Close Project*: Removes the selected project directory from the Projects pane.
- *Send To>parameter name*: When you run an analysis, you must enter values for each parameter. *Send To* makes this file the input value for the selected parameter.
- *Open With*: Displays the content of the selected file in one of two applications:
  - *Text Viewer*: Displays the file in the Text Viewer. You can use the Text Viewer to copy text to the clipboard, but not to modify the file.
  - *Default Application*: Displays the file in the default application (if any) associated with this file type. For information about associating an application with a file type, see your operating system documentation.
- *Show File Location*: Displays the selected file and its parent directory in a file browser.
- *Modules>module name*: When you run an analysis, you must enter values for each parameter. *Modules* runs the selected analysis and makes this file the input value for the first parameter that accepts this file type.

## Results

For information about running analyses and working with analysis results, see [Running Analysis and Visualization Modules](#) and [Working with Analysis Results](#).

- *Reload*: Displays the analysis module and parameter values used to run the selected analysis. You can rerun the same analysis or modify the parameters and run the modified analysis.
- *Delete Job*: Removes the analysis from the Results pane and deletes all of its generated files from the GenePattern server.
- *Delete All Jobs*: Removes all analyses from the Results pane and all of the associated analysis results files from the GenePattern server.
- *Terminate Job*: Cancels a currently executing analysis job.
- *View Code*: Displays the code (Java, MATLAB, or R) used to run the selected analysis.
- *Send To*>*parameter\_name*: When you run an analysis, you must enter values for each parameter. *Send To* makes this file the input value for the selected parameter.
- *Save To*: Copies the selected results file from temporary storage on the GenePattern server to the selected directory.
- *Delete File*: Deletes the selected results file from the GenePattern server.
- *Open With*: Displays the content of the selected file in one of two applications:
  - *Text Viewer*: Displays the file in the Text Viewer. You can use the Text Viewer to copy text to the clipboard, but not to modify the file.
  - *Default Application*: Displays the file in the default application (if any) associated with this file type. For information about associating an application with a file type, see your operating system documentation.
- *Modules*>*module\_name*: When you run an analysis, you must enter values for each parameter. *Modules* runs the selected analysis and makes this file the input value for the first parameter that accepts this file type.
- *Create Pipeline*: Creates a pipeline that contains the analysis modules used to create the selected results file. For more information, see [Working with Pipelines](#).

## Analysis

The Analysis menu lists the analysis modules available on your GenePattern server. For more information, see [Running Analysis and Visualization Modules](#).

## Visualization

The Visualization menu lists the visualization modules available on your GenePattern server. For more information, see [Running Analysis and Visualization Modules](#).

## Pipelines

For information about pipelines, see [Working with Pipelines](#).

- *New*: Creates a pipeline.
- The remainder of the Pipeline menu lists the pipeline modules available on your GenePattern server. Select a pipeline to run, view, edit, or export the pipeline.

## Suites

For information about suites, see [Working with Suites](#).

- *New*: Creates a suite.
- *Filter*: Applies or removes a filter. When you apply a filter and select one or more suites, the Java Client updates its menus to show only the modules listed in the selected suites.
- The remainder of the Suites menu lists the suites available on your GenePattern server. Select a suite to view, edit, or export the suite.

## History

- The first items in the History menu are the most recently run analysis jobs. Select an item to display the analysis module and parameter values used to run that analysis job.
- *View All*: Displays the History window, which lists all analysis jobs that you have run. From the History window, you can select a job and then select *Reload* or *Purge*:
  - *Reload*: Displays the analysis module and parameter values used to run the selected analysis.
  - *Purge*: Removes the selected analysis from the Results pane and deletes all of its generated files from the GenePattern server.

To close the History window, click the Close (X) icon in the upper right corner.

## Help

- *About*: Displays the GenePattern release number, build number and date, and authors.
- *GenePattern Web Site*: Opens a web browser and displays the GenePattern web site. You must be connected to the internet to use this feature.
- *GenePattern Tutorial*: Opens a web browser and displays the GenePattern tutorial. You must be connected to the internet to use this feature.
- *GenePattern Server*: Opens a web browser and displays the GenePattern Web Client.
- *Getting Started*: Displays the welcome message in the main window.
- *Module Color Key*: Displays a legend for the colors used to color-code pipeline names; the colors indicate who authored the module.
- *Errors*: Displays a window that allows you to send a bug report to the GenePattern Help Desk. You must be connected to the Internet to use this feature.

When you create a bug report, please enter as much information as possible including the text of any error messages and the release and build numbers listed in the *Help>About* box.