

User Guide

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.

If you are new to GenePattern, begin with the basics:

- The Concepts Guide provides a brief introduction to GenePattern: its primary objects (modules, pipelines, suites) and its client-server architecture. All other GenePattern documentation assumes that you are familiar with these concepts.
- GenePattern. Many users find that this is all they need to begin using GenePattern.

This User Guide contains the following sections:

Getting Started	Describes how to start and exit from GenePattern. It provides an overview of the user interface, including the navigation bar and pop-up menus.
Running Modules and Pipelines	Describes how to run GenePattern analyses and how to check job status.
Working with Analysis Results	Describes how to display, save, and delete analysis results.
Working with Modules	Describes how to install, create, edit, and delete modules.
Working with Pipelines	Describes how to install, create, edit, and delete pipelines.
Working with Suites	Describes how to install, create, edit, and delete suites.
Managing Modules, Pipelines, and Suites	Provides detailed instructions for installing and deleting modules, pipelines, and suites. The previous sections summarize these actions and provide links to the detailed instructions provided in this section.
Managing the GenePattern Server	Provides information for the GenePattern server administrator.

Getting Started

This section describes how to start and exit from GenePattern. It provides an overview of the user interface, including the navigation bar and pop-up menus.

- Starting GenePattern
- Exploring the User Interface
- Exiting from GenePattern
- Getting Help

Starting GenePattern

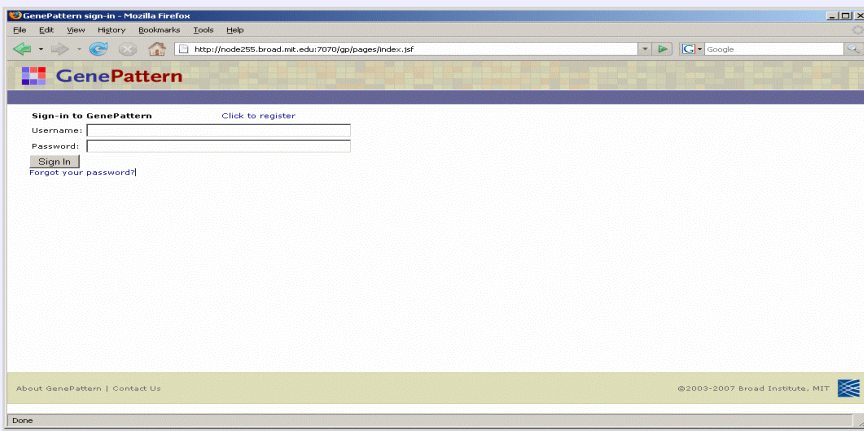
Note: If you are using a local server, you must first start the server as described in Starting a GenePattern Server. If you are using a networked server and you are the server administrator, you must start the server as described in Starting a GenePattern Server.

To start GenePattern:

1. Open a web browser, such as Mozilla Firefox, Internet Explorer or Safari.
2. Enter the URL of the GenePattern server that you want to use:

Server	URL
Broad-hosted server	http://genepattern.broadinstitute.org/gp/
Local server	http://localhost:8080/gp/
Networked server	The URL for the networked server, for example: http://mycompany.com:8080/gp/

GenePattern prompts you to login:



3. Enter your user name and password. If you do not have a GenePattern account, select *Click to register*.

Whether a GenePattern server requires passwords depends on how it is configured. The Broad-hosted server requires passwords. By default, a local server does not.

4. Click *Sign In*. The GenePattern home page appears.

Note: If web browser cannot connect to the server, it displays a message such as “Unable to connect” or “Cannot display the webpage.”

- If you are using the Broad-hosted server or a networked server, wait a minute and try again.
- If you are using a local server, start the server and then try again.

Starting a GenePattern Server

You must install the GenePattern server before you can start it. To install a GenePattern server, follow the instructions provided on the [Download GenePattern](#) page. You use the same installation regardless of whether you are installing a local GenePattern server for personal use or installing a networked server for use by an institution. The difference is in how you configure the server. If you are installing a local GenePattern server, use the default server settings. If you are installing a networked GenePattern, consider reconfiguring the GenePattern server as discussed in [Managing the GenePattern Server](#).

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



StartGenePatternServer

Windows: While the server is starting, the cursor displays as an hourglass. The server is ready when the cursor returns to normal. For Windows 7, you must run this application as an administrator: to start the GenePattern server, right-click on StartGenePatternServer.exe and select *Run as administrator*.

Mac OS X: While the server is starting, the server icon bounces in the Dock. The server is ready when the icon stops bouncing.

Linux: The server starts silently.

When the server has started, open the web interface to your GenePattern server by clicking the GenePatternHome.html shortcut icon. By default, installing GenePattern places this icon on your desktop. If you did not install icons in your task bar or on your desktop, GenePatternHome.html can be found at the top level of your GenePattern install directory (for example, in C:\GenepatternServer\GenePatternHome.html or /Users/JDoe/Applications/GenePatternServer/GenePatternHome.html).

Exploring the User Interface

When first started, GenePattern displays the home page. To return to this page at any time, click the GenePattern icon in the title bar or the *Modules & Pipelines* item in the navigation bar.

- The title bar includes:
 - GenePattern icon: Click the icon to return to this page.
 - **My Settings**: Click this link to modify your GenePattern account settings (see My Settings).
 - **Sign Out**: Click this link to exit from GenePattern.
 - Your username: The username of the person currently logged into GenePattern.
- The navigation bar provides access to other GenePattern pages (see Navigation Bar).
- The Modules & Pipelines pane lists the analysis modules and pipelines that you can run. Modules and pipelines not installed from the Broad repository are shown in red. Use the radio buttons to organize the modules and pipelines as you prefer:
 - **Category** (default): Organizes analysis modules by functional category; pipelines are in the pipeline category. Each module is assigned to a category when it is created. To change the category, you must edit the module definition.
 - **Suite**: Organizes modules and pipelines by suite. Suites are arbitrary collections of modules and pipelines. You can install suites from the Broad repository or create your own.
 - For more information about a suite, click **Suites>Manage** to display all available suites. Click a suite name to display its definition. Click the icon next to a suite name to display a menu of commands for working with the suite (see Suites Menu).
 - **All**: Organizes modules and pipelines alphabetically by name.
- The center pane is the main display pane. GenePattern uses this pane to display information and to prompt you for input. Initially, the center pane contains a welcome page, which provides instructions for running common analysis protocols. To redisplay the welcome page, click the GenePattern icon in the title bar.
- The Recent Jobs pane lists the most recent analyses that you have run and the result files generated by those analyses, as well as any currently running analyses. Click the icon next to a job or file name to display a menu of commands for working with the job or file (see Job Menu and File Menu). Click My Settings to modify the History setting, which controls that number of analyses displayed in the Recent Jobs pane.

Note: The URL of the web browser points to the GenePattern server that you are using. The modules, pipelines and suites displayed in the browser are those installed on the server. When you run a module or pipeline in GenePattern, it runs the analysis on the server and stores the analysis result files on the server.

Navigation Bar

The navigation bar provides access to GenePattern pages and operations. Click a link in the table to go to the section of this guide that describes that operation.

Modules & Pipelines	Display the GenePattern home page.
New Pipeline	Create a pipeline.
New Module	Create a module.
Install from repository	Install a module or pipeline from the Broad repository.
Install from zip	Install a module or pipeline from a zip file.
Manage	Display installed modules or pipelines; delete modules or pipelines.
Suites	Display the Manage Suites page.
New	Create a suite.
Install from repository	Install a suite from the Broad repository.
Install from zip	Install a suite from a zip file.
Manage	Display installed suites; delete suites.
Job Results	Display the Results Summary page.
Results Summary	Display jobs run on the server; delete jobs.
Resources	Display an overview of the resources.
Mailing List	Display the form you use to join a low-traffic GenePattern mailing list.
Report Bugs	Display the form you use to contact the GenePattern team to report bugs, provide feedback, or ask questions.
User Forum	Display information about the GenePattern user forum.
Contact Us	Display a form, which you can use to send questions and comments to the GenePattern team.
Downloads	Display an overview of the available downloads.
Programming Libraries	Download and install GenePattern libraries for use with Java, MATLAB, or R.
Public Datasets	Download sample datasets for use with GenePattern.
Administration	Display the Server Settings page.
Server settings	Modify settings that affect the GenePattern server.
Help	Display the GenePattern home page.
Tutorial	Display the Tutorial, which provides a hands-on tour of GenePattern.
Concepts	Display the Concepts Guide, which provides a brief introduction to GenePattern. Other GenePattern documentation assumes familiarity with these concepts.
User Guide	Display this guide, which describes how to use GenePattern.
Programmer Guide	Display the Programmer Guide, which provides guidelines for writing modules and instructions for accessing GenePattern from the Java, MATLAB, and R programming environments.
Module Documentation	Display a list of the modules and pipelines installed on your server, with brief descriptions and links to the module/pipeline documentation.
File Formats	Display the File Formats Guide, which describes all file formats and provides instructions for creating input files.
Release Notes	Display the Release Notes, which describes new features and known issues in this release.
FAQ	Display the GenePattern list of Frequently Asked Questions.
About	Display the release date and build number of the GenePattern server.


Job Menu

When GenePattern displays an analysis job and its results, click the  icon next to the job name to display a menu of commands for working with that job. For more information, see Working with Analysis Results.

Download	Download a zip file containing all analysis result files for this job.
Terminate	Stop the job. This menu item appears only while the job is running.


Reload	Display the analysis and its parameters in the center pane, with the parameters set to the values used for this analysis job.
Delete	Delete the analysis job and its analysis result files from the GenePattern server.
Info	Display the parameter values and the analysis result files for this job.
View Java Code	Display the command line that you would use to run this job in the Java, MATLAB, or R programming
View MATLAB Code	environments. These commands are useful for programmers who want to access GenePattern from one of these
View R Code	programming environments or from their own applications.

File Menu

When GenePattern displays an analysis job and its results, click the  icon next to the file name to display a menu of commands for working with that file. For more information, see [Working with Analysis Results](#).

Delete	Delete the file from the GenePattern server.
Save	Download the file from the GenePattern server.
Create Pipeline	Create a GenePattern pipeline that includes the modules and parameters necessary to reproduce this result file.
List of modules	List modules that commonly use this type of file as an input parameter. Select an analysis to display its parameters in the center pane, with this result file specified as the first input parameter.

Suites Menu

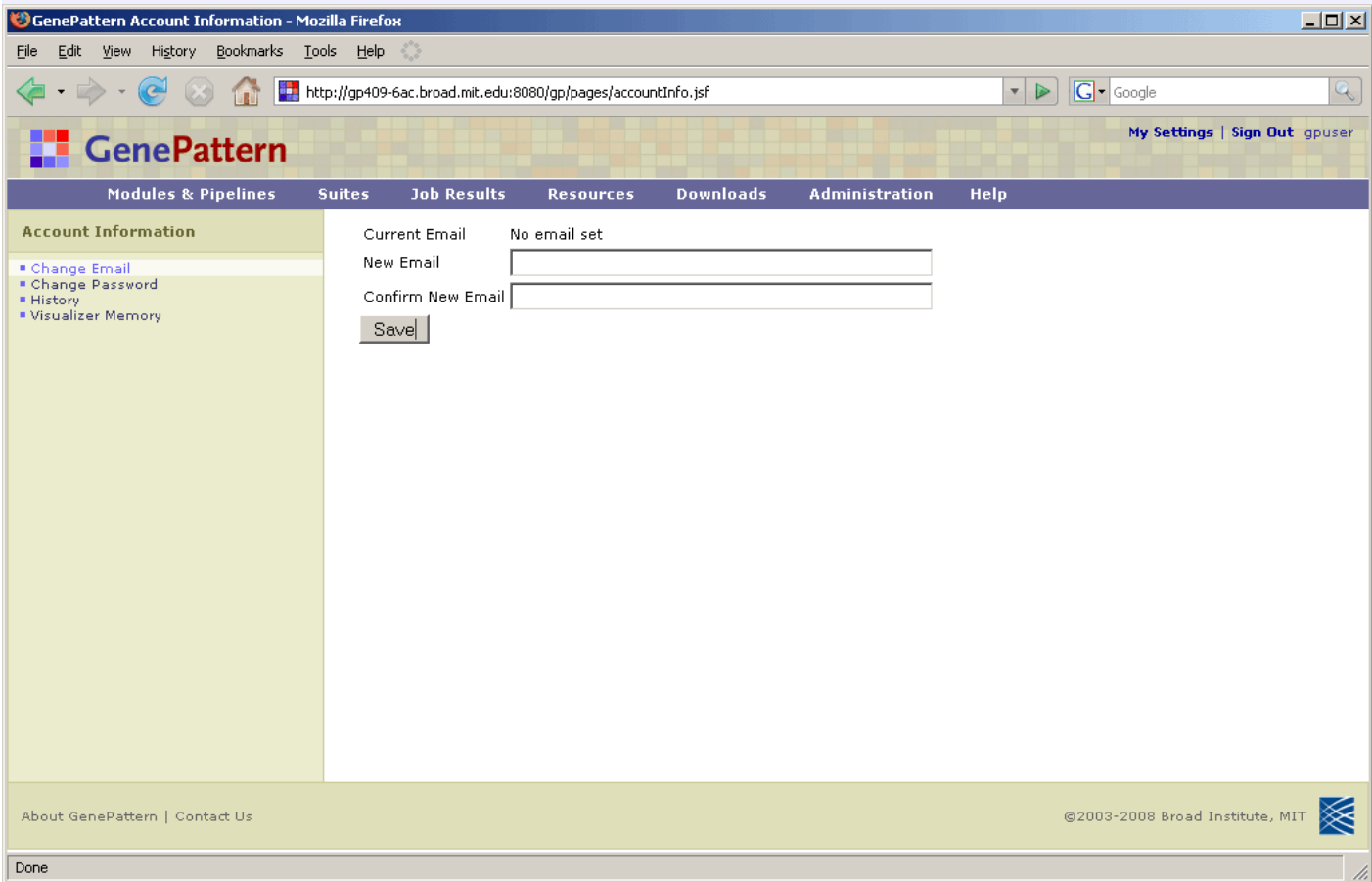
When GenePattern displays a suite, click the  icon next to the suite name to display a menu of commands for working with that suite. For more information, see [Working with Suites](#).

Edit	Available only for suites that you have created. Display the Edit GenePattern Suite page, which you can use to modify your suite.
Delete	Delete the suite from the GenePattern server.
Export excluding dependents	Create a zip file that contains the definition of the suite, but not the modules or pipelines in the suite. The zip file can be used to install the suite on another GenePattern server (see Exporting and Installing Suites Using Zip Files). Installing the suite from this zip file will not install any modules or pipelines in the suite; they must already be installed on the GenePattern server or be installed separately.
Export including dependents	Create a zip file that contains the definition of the suite, as well as the modules and/or pipelines in the suite. The zip file can be used to install the suite on another GenePattern server (see Exporting and Installing Suites Using Zip Files). Installing the suite from this zip file will also install the modules and pipelines in the suite (unless they are already installed on the GenePattern server).

My Settings

Use **My Settings** to change your GenePattern account information:

1. Click **My Settings** (in the GenePattern title bar). GenePattern displays the Account Information page.
2. From the left pane, select the information that you want to modify. GenePattern displays the related page.
3. Modify and save your account information.
4. Optionally, return to step 2 to make additional changes.



- **Change Email:** Change the email address for your GenePattern account on this server. The GenePattern server uses your email address for the following:
 - When you run a module or pipeline, GenePattern sends the job to the GenePattern client and displays the Job Status page. On that page, you can click the *email notification* check box to have GenePattern send you email when the job completes.
 - If you forget your password and request a new password, GenePattern sends your new password to this email address.
- **Change Password:** Change the password for your GenePattern account on this server. **Note:** By default, GenePattern servers are installed without password protection. For information about enabling password protection, see *Securing the Server*.
- **History:** Use this option to specify the number of recent analyses listed in the Recent Jobs pane on the GenePattern home page.
- **Visualizer Memory:** Specify the Java virtual machine configuration parameters (such as VM memory settings) to be used when running visualization modules. By default, this option is used to specify the amount of memory to allocate when running visualization modules (-Xmx512M).

Exiting from GenePattern

To exit from GenePattern, click *Sign Out* in the top right corner of the title bar.

To shutdown a GenePattern server, double-click the Stop GenePattern Server icon, shown below or close the console window.



Windows: When you shutdown the server, the GenePattern console window closes.

Mac OS X: When you shutdown the server, the GenePattern server icon disappears from the Dock.

Linux: The GenePattern server exits silently.

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 Concepts Guide provides a brief introduction to GenePattern: its primary objects (modules, pipelines, suites) and its client-server architecture. All other GenePattern documentation assumes that you are familiar with these concepts.
- The Quick Start (10 minutes) provides a quick introduction to GenePattern and the Broad-hosted server.
- The Tutorial (90 minutes) provides a hands-on tour of GenePattern.
- The User Guide, [this guide](#), describes how to use GenePattern.
- The Programmers Guide provides guidelines for writing modules and instructions for accessing GenePattern from the Java, MATLAB, and R programming environments.
- The Integration Guide provides notes for administrators integrating GenePattern into an analysis environment.
- The Modules page lists the modules and pipelines in the Broad repository, with links to their documentation.
- The File Formats Guide describes all file formats and provides instructions for creating input files.
- The Release Notes describe new features and known issues in this release.
- Frequently Asked Questions answers common questions about GenePattern.


Running Modules and Pipelines

An analysis module runs a single analysis. A pipeline runs a series of analysis modules. If you are unfamiliar with GenePattern modules and pipelines, see the Concepts Guide.

- [Running a Module or Pipeline](#)
- [Parameters Page](#)
- [Rerunning an Analysis](#)

Running a Module or Pipeline

To run a module or pipeline:

1. Select the module or pipeline from the Modules & Pipelines pane. The parameters appear in the center pane.
Alternative: Modules can also be run from the File menu of an analysis result file. Click the  icon next to an analysis result file and select the module to run. The parameters appear in the center pane with the first input file parameter set to the analysis result file. Using this method chains the selected module to the module that created the analysis result file, which is useful for creating pipelines (see [Creating Pipelines](#)).
2. Enter values for the parameter fields (see [Parameters Page](#)).
3. Click **Run**. GenePattern sends the job to the server and displays the Job Status page. How long a job takes to complete depends on the size of your dataset and the analysis that you are running.
4. Click **Return to Modules & Pipelines Start** to return to the GenePattern home page. The Recent Jobs pane shows the job that you just ran.
5. Click the id number of that job to redisplay its Job Status page. The Job Status page provides complete information about the job, including its parameters, input files, output files, and current status (see [Working with Analysis Results](#)).

Parameters Page

When you select a module or pipeline, GenePattern displays its parameters:

- 1 Most modules require one or more input files. There are several ways to choose an input file:
- **Upload a file.** Select the *Upload File* radio button. Click *Browse* to display the file selection window. Navigate to the desired file and select it.
 - **Use a URL.** Select the *Specify URL* radio button. Copy the URL to the entry field.
 - **Use a result file.** Select the *Specify URL* radio button. From the Recent Jobs pane, click the icon next to an analysis result file of the desired type and select *Send To parameter-name*. This has the benefit of chaining this analysis to the previous analysis, which can be useful for creating pipelines (see [Creating Pipelines](#)).

- 2 3 Specify other parameter values using the drop-down lists and entry fields:
- **Drop-down list.** Click a value in the list. Ctrl-click (Mac: Command-click) to select multiple values (if allowed).
 - **Entry field.** Enter a value in the box. Valid values for the field depend on the module and should be listed in the module documentation.

- 4 Hide/show the brief descriptions below each parameter.
- 5 Version of the module. If multiple versions of the module are installed on the server, GenePattern displays the latest version by default. Click the icon next to the version number to select a different version.
- 6
- *Run* button: Start the analysis.
 - *Reset* button: Reset all parameters to their default values.
 - *Properties*: Display the definition of the module or pipeline. The module definition form lists the program that implements the analysis. The pipeline definition form lists the modules run by the pipeline.
 - *Export*: Create a zip file that contains the module or pipeline definition. The zip file can be used to install this module or pipeline on another GenePattern server.
 - *Edit* (not shown): Available only for modules that you have created.
 - *Help*: Display the module or pipeline documentation.



Display 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.

Rerunning an Analysis

To rerun an analysis:

1. Display your analysis jobs in one of two ways:
 - Click **Modules & Pipelines** to display the GenePattern home page, where your most recent jobs are listed in the Recent Jobs pane.
 - Click **Job Results>Results Summary** to display the Job Results page, which lists all of your analysis jobs.
2. Click the icon next to the job that you want to rerun and select **Reload**. The parameters appear in the center pane set to the values that were used for this job.
3. Optionally, modify the parameter values.
4. Click **Run**.



Working with Analysis Results

When you run a module or pipeline, GenePattern runs the analysis job on the GenePattern server. Analysis results are stored on the GenePattern server for a period of time (by default, one week) and then deleted. If you are unfamiliar with how GenePattern runs modules and pipelines, see the Concepts Guide.

- Basic Operations
- Analysis Result Files
- Job Status Page
- Job Results Summary Page
- Sharing Analysis Results

Basic Operations

The following table summarizes ways to work with analysis results:

Display analysis results	Click a job ID number to display the <i>Job Status</i> page, which lists the input parameters and analysis results for that analysis job. (Recent jobs are listed on the GenePattern home page. To display all jobs, click <i>Job Results>Results Summary</i> .)
Share analysis results	By default, analysis results are private. To share results with other GenePattern users, click the Edit Sharing icon on the <i>Job Status</i> page.
Save analysis results	To save results persistently (beyond the period of time they are stored on the server), download the analysis result files to a more permanent location: <ul style="list-style-type: none"> • To save a single file, click the  icon next to a file and select Save. GenePattern saves the file to the location that you select. • To save all of the files, click the  icon next a job and select Download. GenePattern creates a zip that contains all of the result files.
Delete analysis results	If you no longer need your analysis results, you can delete the files from the server: <ul style="list-style-type: none"> • To delete a single file, click the icon next to a file and select Delete. • To delete all of the files, click the icon next to a job and select Delete.

Analysis Result Files

When you run a module or pipeline, the files generated by the module/pipeline are stored on the GenePattern server. The module author determines the content and format of the generated files; however, by convention, each module generates the following files:

- **Analysis result files** are typically formatted text files that contain the results of the analysis. Most analysis result files are intended to be used as input to subsequent analyses. Although these files can be viewed in a text viewer, the amount of information in the files may make them difficult to read. If an analysis module generates an analysis result file that requires examination and interpretation, there is generally a corresponding visualization module that

you can use to display the results. Visualization modules have "Viewer" in the title. For example:

- Results from ComparativeMarkerSelection are viewed using the ComparativeMarkerSelectionViewer
 - Results from modules in the Prediction category are viewed using the PredictionResultsViewer
 - Results from SOMClustering are viewed using the SOMClusteringViewer
 - Gene expression datasets are often viewed using the HeatMapView
- **Execution log files** are text files that describe how the analysis was run. A module might generate 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.txt**: contains "standard output" messages; that is, comments generated as the analysis module runs.
 - **stderr.txt**: contains "standard error" messages; that is, information about errors (if any) that occurred during the analysis.

Log files can be viewed using a text viewer.

Job Status Page

When you run a module or pipeline, GenePattern sends the analysis job to the server and displays the Job Status page. This page displays complete information for an analysis job, including its status, input files, parameter values, and (when the job completes) result files. After starting an analysis, you can continue working. You do not have to leave the Job Status page displayed.

GenePattern offers several ways to redisplay a Job Status page:

- Click *Modules & Pipelines* to display the GenePattern home page, where your most recent jobs are listed in the Recent Jobs pane. Click a job ID number to display the corresponding Job Status page.
- Click *Job Results>Results Summary* to display all of your analysis jobs. Click a job ID number to display the corresponding Job Status page.
- Enter the URL of the Job Status page in your browser: `http://<server>/gp/jobResults/<job number>`
On the Broad-hosted server for example: `http://genepattern.broadinstitute.org/gp/jobResults/111`
On a local server for example: `http://localhost:8080/gp/jobResults/111`

The following figure shows the Job Status page for a pipeline job. The Job Status page for a module job is similar.

The screenshot shows the Job Status page for a pipeline job. The job name is "26.CMSPipeline". The status bar is blue, indicating the job is running, with a red "8" icon. Below the job name, there are several sections: "pipelineErrors.log (0.0 KB)" with a file icon and a red "1" icon; "step 1: PreprocessDataset[id: 27]" with a folder icon, a red "6" icon, and a "Show Input Parameters" button with a red "7" icon; "input filename: all_aml_test.gct" with a file icon and a red "2" icon; "all_aml_test.preprocessed.gct (1.4 MB)" with a file icon and a blue checkmark icon; "step 4: HeatMapView[id: 30]" with a folder icon and a "Show Input Parameters" button with a blue checkmark icon; "dataset: all_aml_test.preprocessed.comp.marker.filt.gct" with a file icon and a red "3" icon; and "Open Visualizer" with a visualizer icon and a red "3" icon. At the top right, there is a "Show Execution Logs" button with a red "4" icon and a share icon with a red "5" icon.

- 1 Click the menu icon to next to the job name or a file name to display the Job or File menu.
- 2 Icons indicate whether this is an input file or an output file.
- 3 Click the visualizer icon to open a visualizer, if the job includes visualizers.
- 4 Show/hide execution log files.
- 5 Click the share icon to share analysis results with other users.
An open share icon indicates that the results are shared.
- 6 Click the parameters icon to display the job's input parameters. To run the job with different parameters, select Reload from the Job menu, enter the desired parameters, and rerun the job.
- 7 Icons indicate whether the job is running, complete, or **ERR** halted due to an error.

8 For a pipeline, each section of the colored line beneath the job name represents a step in the pipeline. As each step completes, its section of the line changes from green to blue.

An *Email Reminder* check box is visible while the job is running. For long running jobs, select the check box to have GenePattern send you email when the job completes. Continue working in GenePattern or exit from GenePattern. When you receive the email indicating that the job is finished, display the Job Status page to review the analysis results.

Job Results Summary Page

The GenePattern home page lists your most recent jobs. The Job Results Summary page lists all of your analysis jobs.

To display the Job Results Summary page, click *Job Results>Results Summary*.

Status	Job	delete	Module Name	Total Size	Submit Date	Complete Date	Job Owner	Your Access	Sharing
✓	18	<input type="checkbox"/>	▼ CMSPipeline	3.5 MB	May 08 09:31:22 AM	May 08 09:31:42 AM	hkuehn@broad.mit.edu	Read, Write	Private
✓	19		↳ 1. PreprocessDataset		May 08 09:31:23 AM	May 08 09:31:24 AM			
		<input type="checkbox"/>	all_ami_test.prep.res	1.8 MB		May 08 09:31:24 AM			
✓	20		↳ 2. ComparativeMarkerSelection		May 08 09:31:25 AM	May 08 09:31:39 AM			
		<input type="checkbox"/>	all_ami_test.prep.cms.odf	1.6 MB		May 08 09:31:39 AM			
✓	21		↳ 3. ExtractComparativeMarkerResults		May 08 09:31:40 AM	May 08 09:31:41 AM			
		<input type="checkbox"/>	all_ami_test.prep.cms.filt.bt	2.0 KB		May 08 09:31:41 AM			
		<input type="checkbox"/>	all_ami_test.prep.cms.filt.res	36.0 KB		May 08 09:31:41 AM			

To sort the job results, click a column header. You can sort jobs by status, job ID, module name, submit date, or completion date. Within jobs, you can sort files by file size or file output date.

1 Filter the display:

- **My job results** (default) lists all analysis jobs that you have run.
- **All job results** lists analysis jobs that you have access to, including jobs that you have run, jobs shared with groups to which you belong, and jobs shared as public. If you are an administrator, you have access to all jobs run on the server so all jobs are displayed.
- **Public job results** lists all jobs shared as public.
- **In group: group-name** lists all jobs share with the named group. There is an **In Group** entry for each group to which you belong.

2 Show/hide the execution log files.

3 Icons indicate whether the job is running, complete, or **ERR** halted due to an error.

4 Click the job ID to display the Job Status page.

5 Delete jobs and/or files: (1) select the check boxes of the jobs and/or files to delete and (2) click the *delete* link in the column header to delete them. Selecting a job selects all of its files. Selecting the check box in the column header selects/clears all check boxes.

6 Name of the module that was run and the name of each result file. Click the arrow next to the Module Name header to hide/show all result files. Click the arrow next to a module name to hide/show its result files.



7 Size of each file and the total job size (combined file size).

8 Time the job was submitted.

9 Time the job was completed and time each file was last saved.

10 Name of the person who ran the job. Or, more precisely, the GenePattern user name of the account that ran the job.



11 Your access to the job. You have read, write access to jobs that you have run. You have either read or read, write access to shared jobs. Write access gives you permission to delete a job or any of its result files.

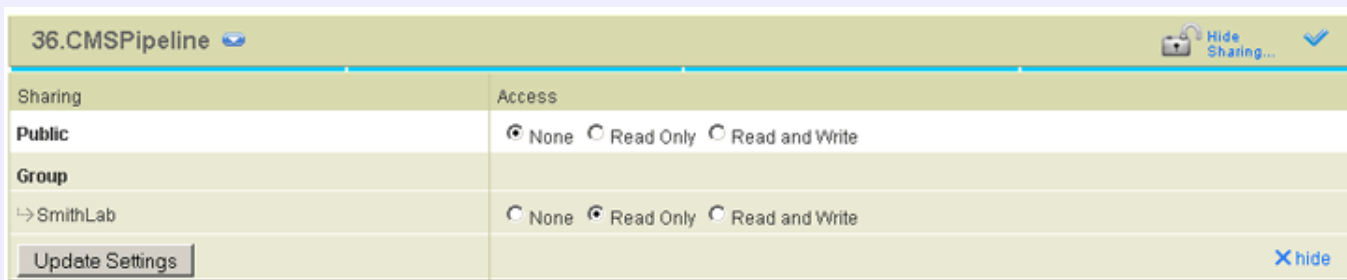
12 Share status:  private or shared . By default, jobs that you run are private: only you (and GenePattern administrators) can view or delete them. To share a job, click its job ID. When GenePattern displays the associated Job Status page, click the share icon to share the job.

Sharing Analysis Results

When you run an analysis job, by default, it is private: only you (and GenePattern administrators) can view or delete the job. Sharing job results gives other GenePattern users access to the job, including its input files, parameter values, and result files.

To share job results or modify the share status of a job:

1. Display the Job Status page for the job. The share icon indicates whether the job is private  or shared .
2. Click the share icon. GenePattern displays the share options for the job:



The screenshot shows the '36.CMSPipeline' job status page. At the top right, there is a 'Hide Sharing...' button with a checkmark. Below this is a table with two columns: 'Sharing' and 'Access'. The 'Public' row has radio buttons for 'None', 'Read Only', and 'Read and Write'. The 'Group' row is set to 'SmithLab' and also has radio buttons for 'None', 'Read Only', and 'Read and Write'. An 'Update Settings' button is at the bottom left, and an 'X hide' button is at the bottom right.

3. Modify the options by choosing which groups should have access and what access they should have.
 - o **Which groups have access.** Jobs can be shared with all GenePattern users (Public) or groups of GenePattern users. You must be a member of group to share a job with that group; therefore, the options include only groups of which you are a member.
 - o **What access.** Groups can be given three levels of access: None, Read Only, Read and Write. Users with read access to a job can download it, rerun it, and view its input parameters, input files, and result files. Users with write access to a job can delete the job or any of its result files.
4. Click *Update Settings* to save your changes.

Sharing input files: In GenePattern, you can specify the output file from one analysis as the input file for a subsequent analysis. For example, you might use the output file from PreprocessDataset as the input file for ComparativeMarkerSelection. In this case, if you share the ComparativeMarkerSelection job, the other user can view the result files but cannot view the input file (which is from the PreprocessDataset job) or rerun the job. To share the ComparativeMarkerSelection job and its input file, either (1) share both the ComparativeMarkerSelection and PreprocessDataset jobs or (2) save the output file from PreprocessDataset, rerun ComparativeMarkerSelection using the saved file, and share the resulting ComparativeMarkerSelection job.

Creating groups: To create a group or add members to a group, contact the GenePattern administrator. If you are an administrator, see [Creating Groups and Administrators](#) for more information.

Working with 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. If you are unfamiliar with GenePattern modules and pipelines, see the [Concepts Guide](#).

- [Basic Operations](#)
- [Displaying Module Definitions](#)
- [Creating Modules](#)
- [Editing Modules](#)

Basic Operations

The following table summarizes the different ways you can work with GenePattern modules.

Run a module	Select a module, enter its parameters and click <i>Run</i> . For more information, see Running Modules and Pipelines .
--------------	--

Display module definitions	A module's definition includes the author, the command line used to invoke module, and the programs used to execute module. To display a module's definition, click <i>Modules & Pipelines</i> and select the module. When GenePattern displays the module parameters, click <i>Properties</i> .
Send module to other users	Zip files provide a convenient way to send modules to other GenePattern users: <ul style="list-style-type: none"> To export a module to a zip file, click <i>Modules & Pipelines</i> and select the module to export. When GenePattern displays the module parameters, click <i>Export</i>. To install a module from a zip file, click <i>Modules & Pipelines>Install from zip</i>. For more information, see Exporting and Installing Modules & Pipelines Using Zip Files .
Install modules from the repository	The Broad Institute maintains a repository of modules, pipelines, and suites. To install modules from the Broad repository, click <i>Modules & Pipelines>Install from Repository</i> . For more information, see Installing Modules & Pipelines from the Repository .
Create modules	An analysis module invokes a program that executes the desired function. To create a module, you must write the program that implements the analysis and then create the GenePattern module that invokes that program. For more information, see Creating Modules .
Edit modules	You can edit a module that you have created or copy a public module and edit your copy of the public module. For more information, see Editing Modules .
Delete modules	To delete a module from your GenePattern server, click <i>Modules & Pipelines>Manage</i> . For more information, see Managing Modules & Pipelines .

Displaying Module Definitions

To display the definition of a module:

1. Click *Modules & Pipelines* to display the GenePattern home page.
2. Select the module to display.
3. When GenePattern displays the module parameters, click *Properties*. GenePattern displays the module definition:

The screenshot shows the GenePattern web interface in a Mozilla Firefox browser window. The page title is 'ExtractComparativeMarkerResults version 3 - Mozilla Firefox'. The browser address bar shows the URL: 'http://gp409-6ac.broad.mit.edu:8080/gp/addTask.jsp?name=urn:lsid:broad.mit.edu:cancer.software.genepal'. The GenePattern logo is visible at the top left, and navigation tabs include 'Modules & Pipelines', 'Suites', 'Job Results', 'Resources', 'Downloads', 'Administration', and 'Help'. The 'My Settings | Sign Out gpuser' link is at the top right.

The main content area displays the module definition for 'ExtractComparativeMarkerResults version 3'. The version is selected in a dropdown menu. Below the title, there is a list of fields:

- Name:** ExtractComparativeMarkerResults (with 'Run' and 'Clone...' buttons)
- LSID:** urn:lsid:broad.mit.edu:cancer.software.genepattern.module.analysis:00046:3
- Description:** Creates a derived dataset and feature list file from the results of ComparativeMarkerSelection
- Author:** Joshua Gould
- Privacy:** public
- Quality level:** production
- Documentation:** [ExtractComparativeMarkerResults.pdf](#)
- Command line:** `<java> <java_flags> -cp <libdir>gp-modules.jar<path.separator><libdir>ExtractComparativeMarkerResults.jar edu.mit.broad.modules.marker.CreateMarkersFileMo <comparative.marker.selection.filename> <dataset.filename> -f<field> -n<min> -x<max> -s<number.of.neighbors> -o<base.output.name>`
- Module Category:** Gene List Selection
- CPU type:** any (if compiled for a specific one)
- Operating system:** any (if operating system-dependent)
- Language:** Java **min. language version:** 1.4
- Version comment:** Updated to work with latest ComparativeMarkerSelection module
- File format(s):** gct;res;Dataset;bt
- Current files:** broad-cg.jar colt.jar ExtractComparativeMarkerResults.jar ExtractComparativeMarkerResults.pdf gp-modules.jar Jama-1.0.1.jar manifest trove.jar
- Parameters:** The names of these parameters will be available for the command line (above) in the form <name>.

The status bar at the bottom of the browser window shows 'Done'.

The module definition is a read-only version of the page used to create and edit the module. The field descriptions below are presented in two parts: a brief description of the field, which is generally sufficient if you are viewing the module definition, and additional details, which are necessary if you are creating or editing the module definition.

- **Name.** Name of the module.

The name of the module is used to identify the module in the user interface of the GenePattern clients. The name should be a short but descriptive, without spaces or punctuation, and may include both upper- and lower-case characters.

Example: ConsensusClustering

- **LSID.** The Life Science Identifier (LSID) used to uniquely identify a GenePattern module.

You cannot create or edit LSIDs. The GenePattern server automatically assigns an LSID to each version of a module. If you are unfamiliar with GenePattern versioning, see the [Concepts Guide](#).

- **Description.** Brief description of the module.

GenePattern displays the description, sometimes in abridged form, in forms and drop-down lists in the clients and in generated code when creating scripts from pipelines. The description should be a sentence or short paragraph that documents succinctly what your module does and why someone would want to use it.

Example from ConsensusClustering: Resampling-based clustering method

- **Author.** The author's name and affiliation (company or academic institution).

GenePattern displays the author name as part of the module definition. This is a comment-only field. If you make this module public, the author field allows other users to credit the author when citing the module and to contact the author with questions, suggestions, or enhancement ideas.

Example from ConsensusClustering: Stefano Monti, Broad Institute

- **Privacy.** Modules may be marked as public or private:
 - Public modules may be accessed by anyone using the GenePattern server.
 - Private modules may be accessed only by the person who installed or created the module (or by an administrator).

When you create a module, by default, it is marked private. When you are ready for others to use your module, change the privacy to public.

- **Quality level.** One of three terms that indicates the author's confidence in the robustness of the module: development, preproduction, and production.

When you create a module, by default, its quality level is development. Although these terms have no strict definitions, they are useful for setting user expectations. If you make this module public, set the quality level appropriately.

- **Command line.** Command line used to launch the module. Values enclosed in angle brackets are replaced by specific values before the command executes.

When you create/update a module, the command line is critical and must be platform-independent. You define the command line as a combination of fixed and dynamic text. GenePattern resolves the dynamic text to build the command line that executes the module. For more information, see [Defining the Module Command Line](#).

- **Module category.** Category name assigned to this module, which is used to organize modules and pipelines. Pipelines are always assigned to the category name *pipeline*.

When you create/update a module, you can choose an existing category name or create a new category name. If your module fits into an existing category, such as Preprocess & Utilities, select that category from the drop-down list; otherwise, click the **New** button to add a new category. GenePattern creates the drop-down list of categories dynamically based on the categories of the modules installed on your GenePattern server. If you delete the last module in a given category, that category is removed from the drop-down list.

- **CPU type.** Indicates the type of CPU required to run the module, or *any* if the module runs on any type of CPU.

When you create/update a module, if your code is compiled for a specific platform (Intel, Alpha, PowerPC, etc.), select that platform from the drop-down list. GenePattern enforces CPU requirements when it runs the module.

- **Operating system.** Indicates the operating system required to run the module, or *any* if the module runs on any operating system.

When you create/update a module, if your code requires a specific operation system (Windows, Linux, MacOS, etc.), select that operating system from the drop-down list. GenePattern enforces operating system requirements when it runs the module.

- **Language.** Indicates the programming language used to implement the module.
- **min. language version.** Indicates the version of the programming language used to implement the module.

GenePattern does not enforce programming language requirements. However, including the language version information in the module definition gives prospective users a hint concerning system requirements.

- **Version comment.** Describes changes made to the module in this version.

When you update a module, briefly describe the changes that you have made. When GenePattern clients display a drop-down list of versions, the comments for each version are visible in the drop-down list.

- **File format(s).** Lists the file formats used by the module.

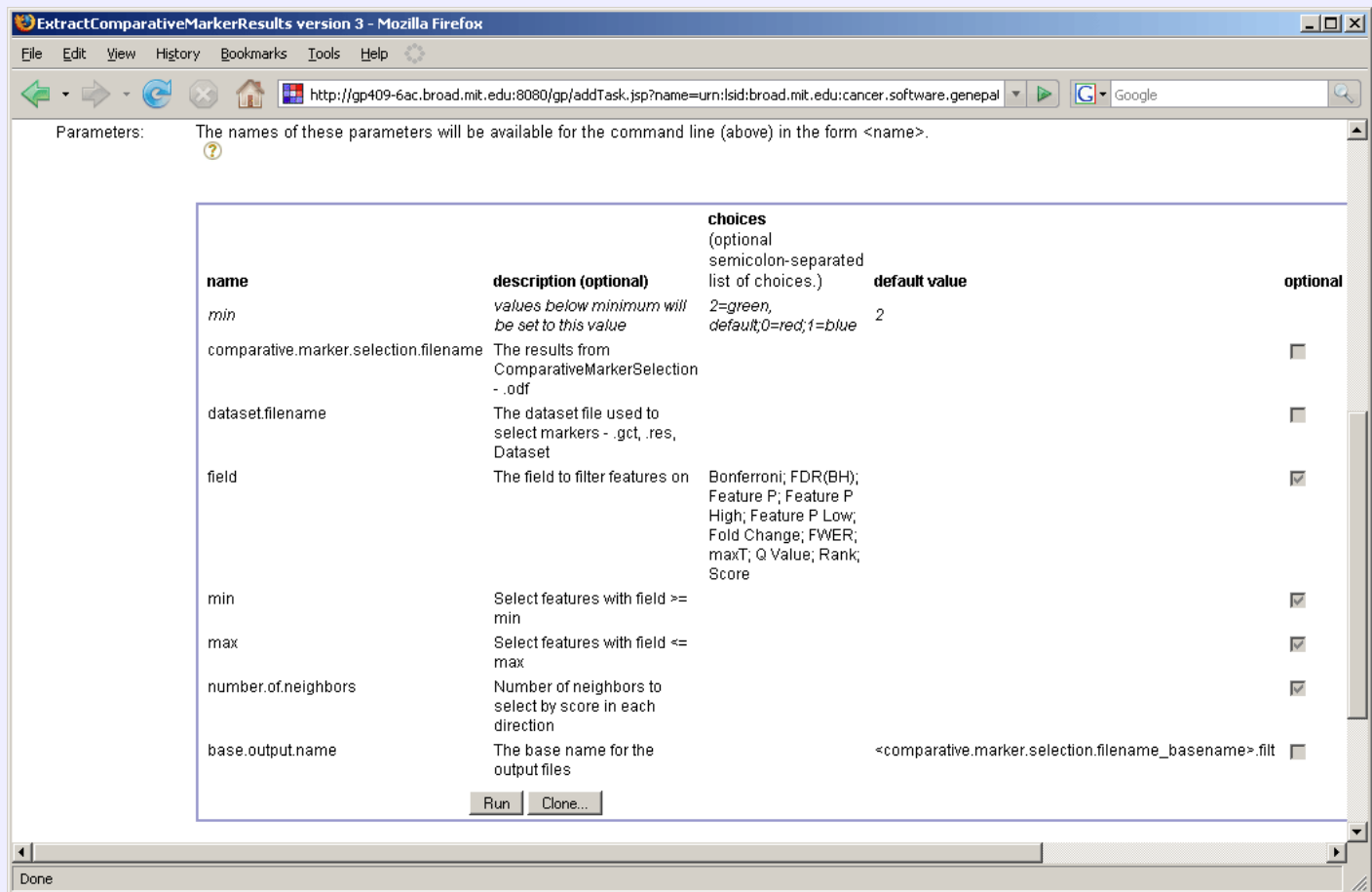
When you create/update a module, use this field to select all file formats used by the module. To select multiple file formats from the list, use ctrl-click or shift-click. If the module uses a file format not included in the list, click the *New* button to add that file format.

- **Current files:** Lists the support files packaged with the module.

When you create/update a module, you must specify all files used in your module, including scripts, libraries, property files, DLLs, executable programs, documentation, and so on. For more information, see Adding Module Support Files.

- **Parameters:** Lists the module parameters.

When you create/update a module, you must define each parameter in the module command line. For more information, see Defining the Module Parameters.



Defining the Module Command Line

Defining the command line that launches your module is a critical piece of defining your GenePattern module. The text of the command line must be variable to address different parameter values, CPU platforms, and operating systems. To ensure that your module runs under different conditions, the command line that you enter will be a combination of fixed and variable text. You specify the variable text in the form of substitution variables enclosed in angle brackets. GenePattern replaces the substitution variables with their assigned values before invoking the command.

Use substitution variables to reference:

- Module parameters. To define module parameters, use the *Parameters* section of the module definition form. To reference a module parameter in the command line, enter the parameter name enclosed in angle brackets; for example, <filename>. Each parameter defined in the *Parameters* section must be mentioned in the command line, unless you have specified that it is an optional parameter. If you provide a default value for a parameter, the default value is used when a user invokes the module and fails to specify a value for the parameter.
- Server configuration properties. To set server configuration properties or create new server configuration properties, select *Administration>Server Settings* and edit the Custom settings, as described in *Modifying Server Settings*. For a complete list of server configuration properties, see the GenePattern /resources/genepattern.properties file. To reference a configuration property in the command line, enter the property name enclosed in angle brackets; for example, <libdir>.
- Java system properties. For more information about these properties, see *Java system properties*. To reference a Java system property in the command line, enter the property name enclosed in angle brackets.
- Programming environment path names. GenePattern provides three server configuration properties, <java> <perl> <R>, that resolve to the full path names for Java, Perl, and R. These are the three languages used to build the modules currently provided in the Broad repository. If your program is run by another interpreter, such as Python or LISP, you can use an absolute pathname (for example, /usr/bin/python) or create a configuration property for the interpreter pathname (for example, python=/usr/bin/python). If your program is an executable file, you do not need an interpreter, so do not need such a property. To create a server configuration property, select *Administration>Server Settings* and edit the Custom settings, as described in *Modifying Server Settings*.

The following table lists the substitution variables for the most commonly used server configuration properties:

<java>	path to Java; usually the one running the GenePattern server
<perl>	path to Perl, installed with GenePattern server on Windows, otherwise the one already installed on your system
<R>	path to a program that runs R and takes as input a script of R commands. R is installed with the GenePattern server on Windows and MacOS
<java_flags>	Java virtual machine configuration parameters (such as VM memory settings) from the Server Settings page (Java Flag settings)
<libdir>	directory where the module's support files are stored
<job_id>	job number
<name>	name of the module being run
<filename_basename>	for each input file parameter, the filename without the file extension or directory
	Note: In the property name, <i>filename</i> is the name of the input file parameter. For example, if you have an input file parameter named input.filename, the substitution property name is <input.filename_basename>. The next two properties are similar.
<filename_extension>	for each input file parameter, the extension without the filename or directory
<filename_file>	for each input file parameter, the input filename without the directory
<path.separator>	Java classpath delimiters (: or ;), useful for specifying a classpath for Java-based modules
<file.separator>	/ or \ for directory delimiter
<line.separator>	newline, carriage return, or both for line endings
<user.dir>	current directory where the job is executing
<user.home>	user's home directory
<parameter_name>	value of the named parameter; for example, if you have a parameter named arg1, use the substitution property <arg1> to include the value of that parameter on the command line

The following example uses substitution variables for two configuration properties, <java> and <libdir>, and one module parameter, <arg1>:

```
<java> -cp <libdir>mymodule.jar com.foo.MyModule <arg1>
```

To execute the module, GenePattern locates the Java runtime and asks to execute the MyModule class using code from the module support file mymodule.jar.

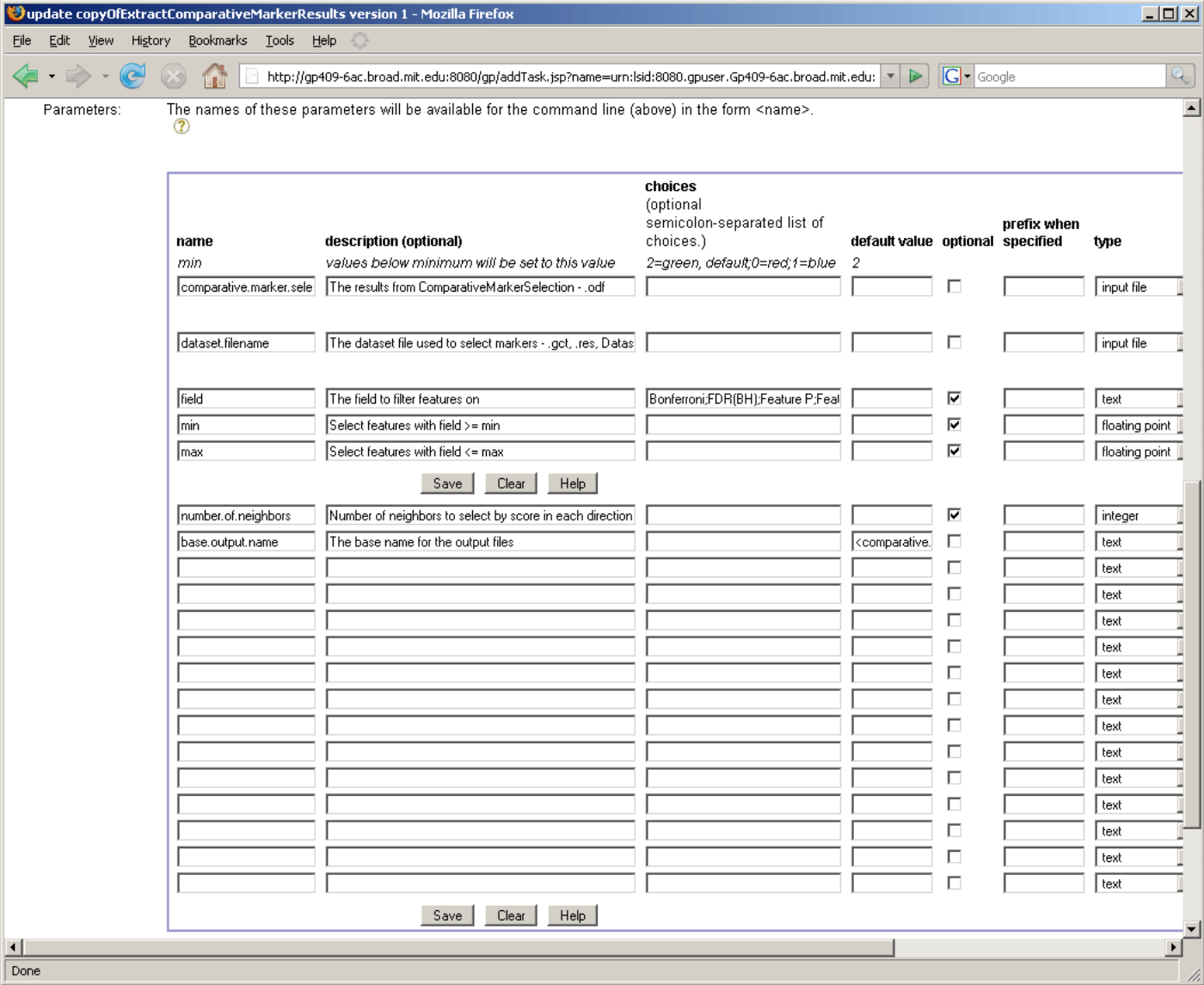
The value of the arg1 parameter is passed as an argument to MyModule.

Note: if your module is designed to accept a standard input stream and/or write to a standard output stream, you can use redirection syntax when describing the command line. To redirect a file to the input stream, enter the text `\<` followed by the input file parameter. To redirect the standard output or standard error streams to a named file, enter the text `\>` or `\>&` followed by the name of the output file. In the following example, the LogTransform module reads its input from the standard input stream and writes its output to the standard output stream:

```
<perl> <libdir>log_transform.pl \< <input.filename> \> <output.file>
```

Defining the Module Parameters

When you create/update a module, you must define each module parameter. GenePattern uses the definition that you supply to prompt users for input when they run your module. Use the Parameters section of the Create/Update module form to enter each parameter. Every parameter entered in this section must appear in the command line, unless you mark the parameter as optional. The following example shows the Parameters section for the Create/Update module form for ExtractComparativeMarkerResults:



Following are descriptions of each field. The first line of the Parameters section provides an example parameter.

- **Name.** Parameter names may include upper- or lower-case characters, numbers, and periods as a separator character between "words". In the command line, reference the parameter name in angle brackets (<name>) to indicate that the value of that parameter should be substituted at that position. The GenePattern clients use the parameter name to prompt the user for the parameter value. When using the programming language environments, parameter names are used to identify the parameters in the function calls.
- **Description.** The description field is optional, but very useful for users of your module. When the GenePattern clients prompt users for module parameters, they display the description of each parameter. It is particularly helpful to provide information such as what the parameter is used for, whether it interacts with other parameters, and any reasonable range of values.

- **Choices.** Some parameters are best represented as a drop-down list of choices. By constraining input to those from the list, the user is saved typing and cannot make a mistake by choosing an invalid setting (unless there is a dependency on some other parameter). The text for the drop-down list can be settings accepted by the program or text that helps the user to understand those settings.
Enter the list of choices as a semi-colon delimited list. For each choice, specify the setting accepted by the program, an equal sign, and the text to be displayed. For example, enter this list of choices:

hierarchical=Hierarchical clustering;SOM=Self-organizing map;NMF=Non-negative Matrix Factorization;=nothing specified;3.14159265=pi

to create a drop-down list that contains the following choices:

Choices	Associated command-line component
Hierarchical clustering	hierarchical
Self-organizing map	SOM
Non-negative Matrix Factorization	NMF
nothing specified	
pi	3.14159265

- **Default value.** Specify a default value to be supplied on the module's command line in cases where the user does not supply a value when invoking the module. This is not the same as the module's own internal defaults. Instead, this allows the GenePattern module declaration author to create a default, even when none exists internally within the module.
Default values for parameters that have a choice list must be either blank or one of the values from the choice list. Any other setting results in an error message. If no default for a choice list is provided, the first entry on the list will be the default.
The default value may use substitution variables, just like the rest of the command line. So a valid default for an output file might be <input.filename_basename>.foo, meaning that the output file will have the same stem as the input.filename parameter, but will have a .foo extension.
- **Optional.** Select the check box to indicate that the parameter is optional. Optional parameters do not have to be specified on the command line. When a user fails to enter a value for an optional parameter, nothing is added to the command line for that parameter.
- **Prefix when specified.** If an optional parameter requires a text prefix on the command line, use this field to specify the prefix. For example, you might need to write "-F *filename*" to pass in a filename; however, if the filename is optional, you do not necessarily want to specify "-F". To solve this problem, set the prefix to "-F " (note that the ending space is included). If the optional filename is provided, then the prefix is also added.
- **Type.** Declaration of the type of an input parameter allows the GenePattern client to present the value to the user in the appropriate format. Parameter type choices are: text, integer, floating point, and input file. (If you specify an input file parameter, when a user runs the module and specifies the input file, GenePattern uploads the entire file to the server, not just the file name.)
- **File Format.** When you select a parameter type of input file, a drop-down list of file formats appears in the file format column. Select the valid file format(s) for this parameter. To select multiple file formats, use CTRL-click. If your module requires an input file format not included in the list, scroll back to the **File format (s)** field and click *New* to add that format to the list.

Adding Module Support Files

When you create/update a module, you must specify all files used in the module, including scripts, libraries, property files, DLLs, executable programs, documentation, and so on. All files are copied to the GenePattern server and may be referenced in the command line field using the syntax: <libdir>filename. You can specify as many files as needed, provided you have the space available on your GenePattern server.

Adding and Removing Support Files

To specify a file, use the *Support files* field:

1. Click the Browse button next to an entry field to select a file. The *Support files* field provides five (5) entry fields, so you can enter at most five files at a time.
2. Click **Save**, near the bottom of the form. GenePattern saves all changes you have made to the module since you last saved, adds the files to the module and copies them to the server, creates a new version of the module, and displays the status message:
Installation of your *ModuleName* module (version *x*) is complete.
Run *ModuleName*
3. To continue editing the module, click your *ModuleName* in the first sentence. In the Create/Update Module form, the files that you added are now listed in the **Current files** field.
4. Repeat this procedure to add additional files.

To remove files that you have already added, use the **Current files** field:

1. Select the file from the *Current files* list.
2. Click **Delete**. Deleting a file from the Current files list deletes all changes since the last time you saved the module. GenePattern displays a warning message reminding you of this fact.
3. Click **OK** to delete the file (and any other changes made since the last time you saved).

Including Documentation

Public modules should always include documentation that provides instructions for using the module, a detailed description of each input parameter, a detailed description of each output file (both its format and content), and explain the algorithm or reference the paper, journal, or book that explains it.

The documentation that you provide with your module is automatically available to GenePattern users. As a GenePattern user, when you select a module, GenePattern displays a form that includes the module parameters and a **Help** button. When you click the **Help** button, GenePattern examines the list of support files for the module and displays the first file that has a standard documentation file extension. If no documentation file was provided, GenePattern displays a message indicating that no information is available. (By default, the standard documentation file extensions are html, htm, xhtml, pdf, rtf, and txt. You can modify this list of extensions by editing the files.doc property in the GenePattern /resources/genepattern.properties file.)

Creating Modules

Creating a GenePattern module is a two-step process:

1. Write a program that executes the desired function. You can write the program in the language of your choice; for example, you can use a compiled language, such as C, to create an executable or use a scripting language, such as Perl, to create a script that is run by an interpreter. The GenePattern Programmer's Guide provides guidelines for writing programs that will be run as GenePattern modules.
2. Use GenePattern to create a module that invokes the program that you have written. It takes just a few minutes to enter the necessary information. Once you have done so, you can run the module. You can decide which parameters from the algorithm to expose to the user and can replace command line parameter names that are hard to remember with names that are self-explanatory. You can also create drop-down list choices for parameters to reduce the possibility of invoking the module with incorrect values.

To create a module that invokes the program that you have written (or otherwise obtained):

1. Click **Modules & Pipelines>New Module**. GenePattern displays the Create Module form:

The screenshot shows the 'add GenePattern module' form in Mozilla Firefox. The browser address bar shows 'http://gp409-6ac.broad.mit.edu:8080/gp/addTask.jsp'. The GenePattern logo is at the top left, and 'My Settings | Sign Out gpuser' is at the top right. The navigation menu includes 'Modules & Pipelines', 'Suites', 'Job Results', 'Resources', 'Downloads', 'Administration', and 'Help'. The form title is 'Create Module'. It contains several input fields with help icons (question marks):

- Name*: (required, no spaces)?
- LSID:
- Description: ?
- Author: (name, affiliation)?
- Privacy: private ?
- Quality level: development ?
- Command line: *
- Module Category: ?
- CPU type: any (if compiled for a specific one) ?
- Operating system: any (if operating system-dependent) ?
- Language: any min. language version: ?
- Version comment: ?
- File format(s): ? (Actual Peaks, Actual Peaks to EM-matched Peaks, adj)
- Support files: (jar, dll, exe, pl) The actual program plus any required libraries will be accessible to your command line as <libdir>filename? (Browse...)

2. Enter values for the fields. For descriptions of the fields, see [Displaying Module Definitions](#) or click the [help](#) icons.
3. Click **Save** to create the module. GenePattern checks the following:
 - o Every parameter not marked as optional is included in the command line.
 - o Every parameter in the command line is a parameter, environment variable, or system property.
 - o Module parameter names are valid.

If no errors are found, GenePattern copies the support files to the server and makes the module available to the GenePattern clients.

Malicious code: By adding a module, a user can execute arbitrary code on the GenePattern server. Because arbitrary code may include malicious code, take precautions to protect your server: for example, employ virus scanner software and restrict access to appropriately privileged (non-root) users. For more information about securing your server, see [Securing the Server](#).

Tutorial: Creating a Module

Following is a brief tutorial that creates a module named `log_transform`. The program invoked by this module is a Perl script, `gp_tutorial_files/log_transform/log_transform.pl`, which log-transforms all positive values in a data set and sets all negative or zero values to zero. This Perl script is part of the GenePattern tutorial data set, which is downloaded during a full installation of GenePattern and is also available on the GenePattern web site.

To create the `log_transform` module:

1. Click **Modules & Pipelines > New Module**. GenePattern displays the Create Module form. (The module definition form, with the tutorial data filled in, appears at the end of the procedure.)
2. Enter the following information in the first few fields:
 - o **Name:** LogTransform
 - o **Description:** Log transform a data (gct) file.
 - o **Author:** Your name and affiliation.
 - o **Privacy:** Select private. This means that only you (or an administrator) can view and run the module. Public allows all users connected to this server to view and run the module.

- o **Quality level:** Select preproduction. This indicates that you have finished development, but are not yet ready for production.

3. Enter the following text in the *Command line* field (if you are working online, cut and paste this text into the field):

```
<perl> <libdir>log_transform.pl -F <input.filename> -o <output.file>
```

Typically, you enter the command line as a combination of fixed text and variables defined by GenePattern. This allows the command line to be independent of the operating environment and allows different values to be specified at different invocations of the command. This command line uses the following variables:

- o <perl> represents the full path to the Perl installation used by GenePattern.
- o <libdir> represents the full path to the directory that contains the files for this module, including the program file.
- o The Perl script, **log_transform.pl**, expects two parameters, an input file and an output file name: **-F <input.filename> -o <output.file>**. When your program has parameters, you include them in the command line and also define them in the *Parameters* field, as described below.

4. Enter the following information in the next few fields:

- o **Module category:** Preprocess & Utilities
- o **CPU type:** Any
- o **Operating system:** Any
- o **Language:** Perl
- o **Version comment:** Initial version
- o **File format(s):** Select gct from the list of formats.

5. Use the *Support files* field to upload your program, any other files needed to execute the module, and documentation for the module:

1. Click the *Browse* button at the end of the first line in the *Support files* field. GenePattern displays the File Upload window.
2. Navigate to the **gp_tutorial_files/log_transform** directory, select the file **log_transform.pl**, and click *Open*. This is the script that implements the module.
3. Click the *Browse* button at the end of the second line in the *Support files* field. GenePattern displays the File Upload window.
4. Select the file **LogTransform.pdf** and click *Open*. This is the documentation for the module. When a GenePattern user displays your module and clicks the *Help* button, GenePattern displays the support file that has a standard text extension (see Adding Module Support Files).

6. Use the *Parameters* field to describe your two program parameters: input.filename and output.file. The parameter names and descriptions that GenePattern displays when a user runs your module are the parameter names and descriptions that you provide here.

In the first row of the *Parameters* field, enter the following information for input.filename:

- o **name:** input.filename
- o **description:** The dataset to be transformed (gct format).
- o **type:** choose input file
- o **file format:** choose gct

In the second row, enter the following information for output.file:

- o **name:** output.file
- o **description:** The name of the new transformed file.
- o **type:** choose text (this is a name that the user enters)

7. Click *Save*. GenePattern displays a message informing you that the module has been saved.

8. Run the module to confirm that it has been added to the GenePattern server correctly.

1. Click **Modules & Pipelines** to display the GenePattern home page.
2. Display a module definition form in one of the following ways:
 - o Select a module that you created. When GenePattern displays the module parameters, click **Edit**.
 - o Select a public module. When GenePattern displays the module parameters, click **Properties**. On the module definition form, click **Clone** to create a copy of the module. You created the copy, so you can edit it.
3. Edit the module definition form. Click **Help** for descriptions of the fields.
4. Click **Save** to create a new version of the module.

Working with Pipelines

A GenePattern pipeline defines a sequential series of modules to be run. Modules run from a pipeline work exactly the same as those run directly from GenePattern. If you are unfamiliar with GenePattern pipelines, see the Concepts Guide.

- Basic Operations
- Displaying Pipeline Definitions
- Creating Pipelines
- Editing Pipelines
- Pipeline Designer Form

Basic Operations

The following table summarizes the different ways you can work with GenePattern pipelines.

Run a pipeline	Select a pipeline, enter its parameters and click Run . For more information, see Running Modules and Pipelines.
Display pipeline definitions	A pipeline's definition lists the pipeline's author, the modules to be run and their parameters. To display the pipeline definition, click Modules & Pipelines and select the pipeline. When GenePattern displays the pipeline parameters, click Properties .
Send pipelines to other users	<p>Zip files provide a convenient way to send pipelines to other GenePattern users.</p> <ul style="list-style-type: none"> • To export a pipeline to a zip file, click Modules & Pipelines and select the pipeline to export. When the pipeline parameters (if any) appear in the center pane, click Export. • To install a pipeline from a zip file: click Modules & Pipelines>Install from zip. <p>For more information, see Exporting and Installing Modules & Pipelines Using Zip Files.</p>
Install pipelines from the repository	The Broad Institute maintains a repository of modules, pipelines, and suites. To install pipelines from the Broad repository, click Modules & Pipelines>Install from Repository . For more information, see Installing Modules & Pipelines from the Repository.
Create pipelines	You can create an empty pipeline and add modules to it, or you can start with an analysis result file and have GenePattern create a pipeline that recreates that analysis result file. For more information, see Creating Pipelines.
Edit pipelines	You can edit a pipeline that you have created or clone a public pipeline and edit your copy of the public pipeline. For more information, see Editing Pipelines.
Delete pipelines	To delete a pipeline, click Modules & Pipelines>Manage . For more information, see Managing Modules & Pipelines.

Displaying Pipeline Definitions

To display the definition of a pipeline:

1. Click **Modules & Pipelines** to display the GenePattern home page.
2. Select the pipeline to display.
3. When GenePattern displays the pipeline parameters, click **Properties**. GenePattern displays the pipeline definition:

The screenshot shows a web browser window displaying the GenePattern interface. The title bar reads 'Golub.Slonim.1999.Nature.all.aml.pipeline - Mozilla Firefox'. The address bar shows the URL 'http://gp409-6ac.broad.mit.edu:8080/gp/viewPipeline.jsp?name=urn:lsid:broad.mit.edu:cancer.software.genet...'. The GenePattern logo is at the top left, and navigation links like 'My Settings', 'Sign Out', and 'gpuser' are at the top right. A main navigation bar contains 'Modules & Pipelines', 'Suites', 'Job Results', 'Resources', 'Downloads', 'Administration', and 'Help'. The main content area displays the pipeline 'Golub.Slonim.1999.Nature.all.aml version 1' with 'Clone...' and 'Run' buttons. Below the pipeline name, it says 'ALL/AML methodology, from Golub and Slonim et al., 1999' and 'Author: GenePattern'. There are links for 'Documentation: Golub.Slonim.1999.Nature.all.aml.pdf' and 'open all | close all'. A list of 8 modules is shown, each with a right-pointing arrow and a version number: 1. PreprocessDataset (version 2), 2. PreprocessDataset (version 2), 3. SOMClustering (version 2), 4. SOMClusterViewer (version 3), 5. ClassNeighbors (version 1), 6. GeneListSignificanceViewer (version 2), 7. WeightedVotingXValidation (version 2), and 8. FeatureSummaryViewer (version 1). A 'Done' status bar is at the bottom left.


On this page, you can:

- Click *open all* and *close all* to show and hide all module parameters.
- Click the arrow icon next to a module to show/hide its parameters.
- Click *Clone* to create your own copy of this pipeline, which you can then edit.
- Click *Run* to run the pipeline.
- Click *Edit* to edit the pipeline. The *Edit* button, not shown here, is only available for pipelines that you have created.

Creating Pipelines

You can create a pipeline in several ways: from an analysis result file, from an existing pipeline, or from scratch (beginning with an empty pipeline).

To create a pipeline from an analysis result file:

1. Click the  menu icon next to the analysis result file and select *Create Pipeline*. GenePattern displays a pipeline designer form and automatically fills in the fields to create a pipeline that will reproduce the analysis results file. GenePattern adds modules to the pipeline based on the following logic: add the module that created the result file; check the module's input file parameters; if the input file for the module was the output file of a previous module, add the previous module; check that module's input file parameters; continue to walk back through the chain of modules, adding modules to the pipeline, until reaching the initial input file.
2. Edit the pipeline designer form as desired.
3. Click *Save* to create the pipeline.

To create a new copy of an existing pipeline:

1. Click *Modules & Pipelines* and select the pipeline. GenePattern displays the pipeline parameters.
2. Click *Properties*. GenePattern displays the pipeline definition page.
3. Click *Clone* to create a copy of the pipeline. GenePattern prompts you to name the new pipeline.

4. Enter a name for the pipeline and click **OK**. GenePattern displays the pipeline definition for the new pipeline.
5. Click **Edit** to edit pipeline. GenePattern displays the pipeline designer form.
6. Edit the pipeline designer form as desired.
7. Click **Save** to create the pipeline.

To create a pipeline from scratch:

1. Click **Modules & Pipelines>New Pipeline**. GenePattern displays an empty pipeline designer form.
2. Edit the pipeline designer form as desired.
3. Click **Save** to create the pipeline.

Editing Pipelines

To edit a pipeline:

1. Click **Modules & Pipelines** to display the GenePattern home page.
2. Select the pipeline that you want to edit. GenePattern displays the pipeline parameters.
3. Display the pipeline designer form in one of two ways:
 - o Click the **Edit** link, if it is available. GenePattern displays the pipeline designer form.
This link is visible only if you created this pipeline on this GenePattern server.
 - o Otherwise, create a copy of the pipeline to edit:
 1. Click **Properties**. GenePattern displays the pipeline definition form.
 2. Click **Clone** (at the top of the form). GenePattern prompts you to name the new pipeline.
 3. Enter a name for the pipeline and click **OK**. GenePattern displays the pipeline definition form for the new pipeline.
 4. Click **Edit** (at the top of the form). GenePattern displays the pipeline designer form.
4. Edit the pipeline designer form as desired.
5. Click **Save** to save the pipeline.

Pipeline Designer Form

GenePattern displays the pipeline designer form when you create or edit a pipeline. Unless you are creating a pipeline from scratch, the pipeline designer form is already populated:

The screenshot shows the GenePattern Pipeline Designer interface in a Mozilla Firefox browser. The page title is "job11 - Mozilla Firefox". The URL is "http://gp409-6ac.broad.mit.edu:8080/gp/pipelineDesigner.jsp?name=urn%3AIsid%3A8080.gpuser.Gp409-6ac.". The GenePattern logo is visible at the top left, and navigation links for "My Settings" and "Sign Out gpuser" are at the top right. A menu bar contains "Modules & Pipelines", "Suites", "Job Results", "Resources", "Downloads", "Administration", and "Help".

The main content area is titled "Pipeline Designer - job11 version 1". It contains a form with the following fields:

- Pipeline name*: job11
- Description: describe it here
- Author: gpuser (name, affiliation)
- Privacy: private
- Version comment: 0
- Documentation: (with a "Browse..." button)

Below the form, the LSID is displayed: "urn:lsid:8080.hkuehn.Gp409-6ac.broad.mit.edu.genepatternmodules:4:1".

A section titled "1. ComparativeMarkerSelection:" is expanded, showing a version dropdown set to "3.9 [latest]". Below this, a table lists parameters for the module:

prompt when run	parameter name	value	description
<input type="checkbox"/>	input file:	<input type="text"/> Browse...	The input file - .res, .gct, .odf type=Dataset
<input type="checkbox"/>	cls file:	<input type="text"/> Browse... <GenePatternURL>getFile.jsp?task=<LSID>&file=all_amf_train.res	The class file - .cls
<input type="checkbox"/>	confounding variable cls file:	<input type="text"/> Browse... <GenePatternURL>getFile.jsp?task=<LSID>&file=all_amf_train.cls	The class file containing the confounding variable.
<input type="checkbox"/>	test direction:	2 Sided	The test to perform

The status bar at the bottom of the browser window shows "Done".

When you create a pipeline from scratch, the form is initially empty:

The screenshot shows the GenePattern Pipeline Designer interface in a Mozilla Firefox browser. The page title is "new pipeline - Mozilla Firefox". The URL is "http://gp409-6ac.broad.mit.edu:8080/gp/pipelineDesigner.jsp". The GenePattern logo is visible at the top left, and navigation links for "My Settings" and "Sign Out gpuser" are at the top right. A menu bar contains "Modules & Pipelines", "Suites", "Job Results", "Resources", "Downloads", "Administration", and "Help".

The main content area is titled "Pipeline Designer". It contains an empty form with the following fields:

- Pipeline name*: (empty)
- Description: (empty)
- Author: (empty) (name, affiliation)
- Privacy: private
- Version comment: (empty)
- Documentation: (with a "Browse..." button)

Below the form, the LSID is displayed: "urn:lsid:8080.hkuehn.Gp409-6ac.broad.mit.edu.genepatternmodules:4:1".

At the bottom of the form, there are two dropdown menus for "Module Category & Name:" and a "Save" button.

The footer of the page contains "About GenePattern | Contact Us" on the left and "©2003-2007 Broad Institute, MIT" on the right, along with the Broad Institute logo.

The status bar at the bottom of the browser window shows "Done".

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

- Editing the Pipeline Description
- Adding Modules

- Specifying Parameter Values
- Removing Modules
- Reordering Modules
- Passing Parameters to a Pipeline

Editing the Pipeline Description

The fields at the top of the pipeline definition form describe the pipeline:

The screenshot shows the GenePattern Pipeline Designer interface in a Mozilla Firefox browser window. The page title is 'copyOfGolub.Slonim.1999.Nature.all.aml - Mozilla Firefox'. The URL is 'j:8080/gp/pipelineDesigner.jsp?name=urn:lsid:8080.gpuser.Gp409-6ac.broad.mit.edu:genepatternmodules:5:1'. The GenePattern logo is visible in the top left, and navigation links for 'My Settings' and 'Sign Out gpuser' are in the top right. A menu bar contains 'Modules & Pipelines', 'Suites', 'Job Results', 'Resources', 'Downloads', 'Administration', and 'Help'. The main content area is titled 'Pipeline Designer - copyOfGolub.Slonim.1999.Nature.all.aml version 1'. The form includes fields for 'Pipeline name*', 'Description', 'Author', 'Privacy', and 'Version comment'. The 'Pipeline name*' field contains 'copyOfGolub.Slonim.1999.Nature.all.aml'. The 'Description' field contains 'ALL/AML methodology, from Golub and Slonim et al., 1999'. The 'Author' field contains 'GenePattern' with a small 'name, affiliation' label. The 'Privacy' field is a dropdown menu set to 'private'. The 'Version comment' field contains '1.0'. Below these fields is a 'Documentation' section with a link to 'Golub.Slonim.1999.Nature.all.aml.pdf' and a 'Remove Golub.Slonim.1999.Nature.all.aml.pdf from pipeline' button. There is also an 'Add doc file' field with a 'Browse...' button. The 'LSID' is 'urn:lsid:8080.hkuehn.Gp409-6ac.broad.mit.edu:genepatternmodules:5:1'. Below the main form is a section for '1. PreprocessDataset:' with a version dropdown set to '2'. This section contains a table of parameters:

prompt when run	parameter name	value	description
<input type="checkbox"/>	input filename:	<input type="text" value="http://ftp.broad.mit.edu/pub/genepattern/all_aml/all_aml_train.res"/> Browse...	input filename - .res, .gct, .odf
<input type="checkbox"/>	output file:	<input type="text" value="all_aml_train_filt"/>	Output file with preprocessed dataset
<input type="checkbox"/>	output file format:	<input type="text" value="res"/>	output file format
<input type="checkbox"/>	filter flag:	<input type="text" value="filter"/>	Variation filter and thresholding flag
<input type="checkbox"/>	preprocessing flag:	<input type="text" value="no disc or norm"/>	Discretization or normalization flag

- **Pipeline name:** The name of the pipeline. When naming pipelines, note the following:
 - Pipeline names can include alphanumeric characters, periods (.), and underscores (_).
 - Pipeline names must not include spaces or special characters 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. If you are unfamiliar with GenePattern versioning, see the Concepts Guide.
- **Description:** A brief description of the pipeline, which is displayed when a user runs the pipeline or views the pipeline definition.
- **Author:** The author's name and affiliation (company or academic institution). This is a comment-only field. If you make this pipeline public, the author field allows other users to credit the author and to contact the author with questions, suggestions, or enhancement ideas.
- **Privacy:** Select Private (default) or Public. A private pipeline can be accessed only by the person who created or installed the pipeline (or by an administrator); a public pipeline can be seen and run by all users.
- **Version comment:** A brief description of this version. When GenePattern clients display a drop-down list of versions, the comments for each version are visible in the drop-down list.
- **Documentation:** The full path name of the file that contains the pipeline documentation. Documentation is strongly encouraged for public pipelines. Click **Browse** to select the (previously created) pipeline documentation file.

The documentation that you provide with your pipeline is automatically available to GenePattern users. As a GenePattern user, when you select a pipeline, GenePattern displays a form that includes the pipeline parameters and a *Help* button. When you click the *Help* button, GenePattern examines the list of support files for the module and displays the first file that has a standard documentation file extension. If no documentation file was provided, GenePattern displays a message indicating that no information is available. (By default, the standard documentation file extensions are html, htm, xhtml, pdf, rtf, and txt. You can modify this list of extensions by editing the files.doc property in the GenePattern /resources/genepattern.properties file.)

- **LSID:** The Life Science Identifiers (LSIDs) for this pipeline. You cannot create or edit LSIDs. The GenePattern server automatically assigns an LSID to each version of a pipeline. If you are unfamiliar with GenePattern versioning, see the Concepts Guide.

Adding Modules

To add a module:

1. Save any changes that you have made to the pipeline.

When you add a new module in the middle of a pipeline, file name parameters for subsequent modules may be lost if they have not yet been saved. Saving the pipeline avoids this potential problem.

2. Display the module selection fields:

- o If the pipeline designer form is empty, these fields are displayed at the top of the form:

Module Category & Name: [-Category-] []

- o Otherwise, find the module **before** the one you want to add and click its *Add Another Module* button. GenePattern displays the selection fields below the selected module:

3. ComparativeMarkerSelectionViewer: version 4

Views the results from ComparativeMarkerSelection
documentation: ComparativeMarkerSelectionViewer.pdf

prompt when run	parameter name	value	description
<input type="checkbox"/>	comparative marker selection filename:	[] Browse...	The output from ComparativeMarkerSelection
		or use output from [2. ComparativeMarkerSelection] [Comparative Marker Selection]	
<input type="checkbox"/>	dataset filename:	[] Browse...	The dataset file used as input to ComparativeMarkerSelection
		or use output from [Choose module] []	

Module Category & Name: [-Category-] []

3. Select a category of module from the *Category* list. GenePattern displays a drop-down list of modules in that category:

Module Category & Name: [Clustering] []

[- Module -]

- Module -
- ConsensusClustering - Resampling-based clustering method
- HierarchicalClustering - Hierarchical Clustering
- NMFConsensus - Non-negative Matrix Factorization (NMF) Consensus Clust...
- SDMClustering - Self-Organizing Maps algorithm

4. Select a module from the drop-down list of modules. GenePattern displays the definition form for that module:

Self-Organizing Maps algorithm
documentation: [SOMClustering.pdf](#)

version 2

prompt when run	parameter name	value	description
<input type="checkbox"/>	dataset filename:	<input type="text"/> <input type="button" value="Browse..."/>	Dataset (res, gct, or odf dataset)
<input type="checkbox"/>	output stub:	<input type="text" value="<dataset.filename_base"/>	The base output file name - .odf
<input type="checkbox"/>	cluster range:	<input type="text" value="2-3"/>	Range of values can be entered and the program will automatically run the algorithm for each number of clusters in the range.
<input type="checkbox"/>	seed range:	<input type="text" value="42"/>	The seed for the random number generator, is exposed to allow the user to recreate a given session at a later time, (as opposed to allowing the program to generate a random initial seed which could result in different outcomes even if all the other parameters are identical).
<input type="checkbox"/>	iterations:	<input type="text" value="50000"/>	How many times the algorithm should try to refine the clusters. Initially, this value can be set low for faster exploration, but should be set high (e.g., 50,000) for good convergence.
<input type="checkbox"/>	cluster by:	<input type="text" value="rows"/>	Whether to cluster by rows or columns
<input type="checkbox"/>	som rows:	<input type="text" value="0"/>	Setting this and som.cols to a non zero value will override cluster.range and the computation will be for the specified geometry
<input type="checkbox"/>	som cols:	<input type="text" value="0"/>	Setting this and som.rows to a non zero value will override cluster.range and the computation will be for the specified geometry
<input type="checkbox"/>	initialization:	<input type="text" value="Random_Vectors"/>	How to select initial random centroids
<input type="checkbox"/>	neighborhood:	<input type="text" value="Bubble"/>	Neighborhood function determining how centroids near the target centroid are updated
<input type="checkbox"/>	alpha initial:	<input type="text" value="0.1"/>	Initial learning weight for centroid updates
<input type="checkbox"/>	alpha final:	<input type="text" value="0.005"/>	Final learning weight for centroid updates
<input type="checkbox"/>	sigma initial:	<input type="text" value="5.0"/>	Initial sigma to determine update neighborhood size around the target centroid
<input type="checkbox"/>	sigma final:	<input type="text" value="0.5"/>	Final sigma to determine update neighborhood size around the target centroid

- The title bar of the module definition form includes the module name and version. By default, you are using the most recent version of the module. If more than one version of the module is installed on the GenePattern server, the title bar includes a drop-list of the installed versions. To have the pipeline run a different version of the module, select that version from the drop-down list.
- Enter the parameter values for the module as described in [Specifying Parameter Values](#).

The pipeline definition form has no mechanism for adding a module **before** an existing module; therefore, to add a module at the **beginning** of the pipeline: add the `new_first` module below the `old_first` module, delete the `old_first` module, and (if necessary) recreate the `old_first` module below the `new_first` module.

Specifying Parameter Values

For most parameters, you enter a value, select a value from a drop-down list, or use the default value supplied by GenePattern.

For input file parameters, you can select a file or use an output file from a previous module:

- To select a file, click the **Browse** button to the right of the parameter field and select the file.
- To select an output file generated by a previous module, select the file from the drop-down lists next to the **use output from** label. From the first drop-down list, select the module that generated the output file. From the second drop-down list, select the desired output file.

The drop-down list of output files lists ordinal numbers (1st, 2nd, 3rd, 4th), which allows you to select the output file based on the order in which it was generated, and may also lists data types (for example, gct or cls), which allows you to select the first output file of the selected type. Note that an output file of type **dataset** indicates an **odf file of type dataset**, not a gct or res file.

Prompt for value: Rather than specifying a parameter, you can have GenePattern prompt the user for a value when the pipeline is run:

- Select the **Prompt when run** box to the left of the parameter field. GenePattern removes the parameter value from the pipeline definition.

▼ 1. ComparativeMarkerSelection:

Computes significance values for features using several metrics, including FDR(BH), Q-Value, maxT, FWER, Feature-Specific P-Value, and Bonferroni.
documentation: [ComparativeMarkerSelection.pdf](#)

prompt when run	parameter name	value
<input checked="" type="checkbox"/>	input file:	set prompt when run display settings...
<input checked="" type="checkbox"/>	cls file:	set prompt when run display settings...
<input type="checkbox"/>	confounding variable cls file:	<input type="text"/> <input type="button" value="Browse..."/>

- Optionally, click *set prompt when run display settings* to modify the parameter name and/or description. By default, the parameter name and description used for this pipeline parameter will be identical to those used for this module parameter.

Removing Modules

To remove a module, click the delete button for that module definition:

▼ 4. SOMClusterViewer: version 5

Visualize clusters created with the SOM algorithm
documentation: [SOMClusterViewer.pdf](#)

prompt when run	parameter name	value	description
<input type="checkbox"/>	som cluster filename:	<input type="text"/> <input type="button" value="Browse..."/>	SOM Properties ODF file
	or use output from	3. SOMClustering	1st output

Reordering Modules

The pipeline definition form has no mechanism for reordering module. You must delete the module that you want to move and recreate it in its new position.

Passing Parameters to a Pipeline

When you add a module to a pipeline, you specify the parameter values for that module. Optionally, you can have GenePattern prompt the user for one or more parameter values when the pipeline is run. As described in [Adding Modules](#), you select the *Prompt when run* box to the left of a parameter and, optionally, click *set prompt when run display settings* to modify the description of the parameter. GenePattern prompts the user for these parameter values each time the pipeline is run.

Occasionally, a pipeline requires that the same input file be specified for multiple parameters. For example, consider the following scenario:

- Run ComparativeMarkerSelection. Prompt the user for the *input filename* parameter (an expression dataset file) and the *cls filename* parameter (a class file).
- Run ExtractComparativeMarkerSelection. Set the *comparative marker selection filename* parameter to the output file from ComparativeMarkerSelection and prompt the user for the *dataset filename* parameter (the same expression dataset file used for ComparativeMarkerSelection).

Ideally, you want to prompt the user for an expression dataset file and use that file for both the *input filename* (ComparativeMarkerSelection) and *dataset filename* (ExtractComparativeMarkerSelection) parameters. To do that, add the ConvertLineEndings module to the start of your pipeline:

- Run ConvertLineEndings. Prompt the user for the *input filename* parameter. ConvertLineEndings takes an input file and creates an output file that is similar (it converts the line endings in the file to those used by perl on the host operating system).
- Run ComparativeMarkerSelection. Set the *input filename* parameter to the output file from ConvertLineEndings and prompt the user for the *cls filename* parameter (a class file).
- Run ExtractComparativeMarkerSelection. Set the *comparative marker selection filename* parameter to the output file from ComparativeMarkerSelection and set the *dataset filename* to the output file from ConvertLineEndings.

Working with Suites

Use suites to group modules and pipelines into packages that have related functionality; for example, you might create a suite that contains the modules that you most commonly use to analyze new data files. Suites help you to organize and work with modules and pipelines. If you are unfamiliar with GenePattern suites, see the Concepts Guide.

- Basic Operations
- Displaying Suite Definitions
- Creating Suites
- Editing Suites

Basic Operations

The following table summarizes the different ways you can work with GenePattern suites.

Display suite definitions	To display the suite definition: <ol style="list-style-type: none"> 1. Click <i>Suites>Manage</i>. 2. Click the name of the suite that you are interested in.
Send suites to other users	Zip files provide a convenient way to send suites to other GenePattern users. <ul style="list-style-type: none"> • To export a suite to a zip file, click <i>Suites>Manage</i>, click the icon next to a suite to display the suite menu, and select <i>Export Including dependents</i> or <i>Export excluding dependents</i>. • To install a suite from a zip file: click <i>Suites>Install from zip</i>. <p>For more information, see Exporting and Installing Suites Using Zip Files.</p>
Install suites from the repository	The Broad Institute maintains a repository of modules, pipelines, and suites. To install suites from the Broad repository, click <i>Suites>Install from Repository</i> . For more information, see Installing Suites from the Repository .
Create suites	To create a suite, click <i>Suites>New</i> . For more information, see Creating Suites .
Edit suites	You can edit a suite that you have created or copy a public suite and edit your copy of the public suite. For more information, see Editing Suites .
Delete suites	To delete a suite from your GenePattern server, click <i>Suites>Manage</i> . For more information, see Managing Suites .

Displaying Suite Definitions

To display the definition of a suite:

1. Click *Suites>Manage*.
2. Click the name of the suite that you want to display. GenePattern displays the suite definition:

ClusteringSuite

Name:* ClusteringSuite

Description: Clustering modules partition a gene expression dataset into clusters such that the gene expression data in each cluster share common expression traits based on the distance measure used.

Author: GenePattern


Privacy: public

Support Files: open all | close all

module name	version
ConsensusClustering	1
HierarchicalClustering	2
NMFConsensus	3
SOMClustering	2

- **Name:** The name of the suite.
- **Description:** A brief description of the suite.
- **Author:** The author's name and affiliation (company or academic institution). This is a comment-only field.
- **Privacy:** Select Private (default) or Public. A private suite may be accessed only by the person who created or installed the suite (or by an administrator); a public suite can be accessed by all users.
- **Support files:** Any files included with the suite. Public suites generally include documentation.
- **Modules & Pipelines:** The remainder of the definition form lists all of the modules and pipelines that are on your GenePattern server by category. The ClusteringSuite in this figure contains modules from only one category, the Clustering category.

From this page, you can:

- Click the  menu icon next to the suite name to display the suite menu.
- Click **open all** and **close all** to show and hide the modules and pipelines in all categories.
- Click the arrow icon next to a category to show/hide the modules and pipelines in that category.

Creating Suites

To create a suite:

1. Click **Suites>New**. GenePattern displays an empty suite definition form.
2. Define the suite by entering values for each field. For descriptions of these fields, see Displaying Suite Definitions.
3. To add modules and pipelines to the suite:
 - Select the check box next to a module/pipeline to add that module/pipeline to the suite.
 - Select the check box next to a category name to add all modules and pipelines in that category to the suite.
4. Click **Save**. GenePattern creates the suite.

Create GenePattern Suite

Name:*

Description:

Author:

Privacy:

Any documentation or data files you wish to bundle with the suite (pdf, doc, data files, etc.)

Support Files:

[open all](#) | [close all](#)

Annotation	
<input type="checkbox"/> module name	version
<input type="checkbox"/> GeneCruiser	<input type="text" value="latest"/>

Clustering	
<input type="checkbox"/> module name	version
<input type="checkbox"/> ConsensusClustering	<input type="text" value="latest"/>
<input type="checkbox"/> HierarchicalClustering	<input type="text" value="latest"/>
<input type="checkbox"/> KMeansClustering	<input type="text" value="latest"/>

Projection	
<input type="checkbox"/> module name	version
<input type="checkbox"/> NMF	<input type="text" value="latest"/>
<input type="checkbox"/> PCA	<input type="text" value="latest"/>

Proteomics	
<input type="checkbox"/> module name	version
<input type="checkbox"/> AreaChange	<input type="text" value="latest"/>
<input type="checkbox"/> LocatePeaks	<input type="text" value="latest"/>

Editing Suites

To view or edit a suite:

1. Click **Suites>Manage**. GenePattern displays the Manage Suites page, which lists the suites on your GenePattern server.
2. To view the suite definition, click the suite name.
3. To edit the suite definition:
 1. Click the menu icon next to the suite name and select **Edit**. This option is available only if you created the suite.
 2. Edit the suite definition.
 3. Click **Save** to update the suite definition.

Managing Modules, Pipelines, and Suites

An analysis module runs a single analysis. A pipeline runs a series of analysis modules. Suites group modules and pipelines into packages that have related functionality, which helps you to organize and work with modules and pipelines. If you are unfamiliar with GenePattern modules, pipelines and suites, see the Concepts Guide.

- Installing Modules & Pipelines from the Repository
- Exporting and Installing Modules & Pipelines Using Zip Files
- Managing Modules & Pipelines
- Installing Suites from the Repository
- Exporting and Installing Suites Using Zip Files
- Managing Suites

Installing Modules & Pipelines from the Repository

The Broad Institute maintains a repository of modules and pipelines that are freely available to the public. To install these modules and pipelines on your GenePattern server:

1. Click **Modules & Pipelines>Install from Repository**. GenePattern displays the Install Modules & Pipelines from Repository page, as shown below.

2. Select the modules and pipelines to install.

The search fields at the top of the page allow you to select any version of a module that is not installed on your server or more recent versions of any module already installed on your server. To install an older version of a module already installed on your server, delete the current version(s) installed on your server.

You can then use this page to install any version of the module (which is no longer installed on your server).

3. Click **Install Checked**.

The screenshot shows the GenePattern web interface. At the top, there's a navigation bar with 'Modules & Pipelines', 'Suites', 'Job Results', 'Resources', 'Downloads', 'Administration', and 'Help'. Below this is the 'Install Modules & Pipelines From Repository' section. It contains several search options: 'Search for new modules to install' (checked), 'Search for updates of the currently installed modules' (checked), and 'Search for up to date modules' (unchecked). Under 'Operating System', 'Platform Independent' and 'Windows' are checked. A 'Search' button is present. Below the search options, it says 'Found 20 modules' and 'Install Checked' button. A table lists the results:

<input checked="" type="checkbox"/>	Name (Version)	Module Type	Details
<input checked="" type="checkbox"/>	CompareSpectra 1	Proteomics	author: D.R. Mani, Broad Institute size: 2.8 MB requirements: R 2.0.1, any OS Compares two spectra to determine similarity documentation download zip
<input checked="" type="checkbox"/>	HierarchicalClustering.MATLAB 0	Clustering	author: Gad Getz, gp-help@broad.mit.edu size: 175.0 KB required patches: MCRInstaller4.0 (81.4 MB) requirements: MATLAB, Windows OS Hierarchical Clustering

Use the top section of the form to find the modules to install. To update the list of modules/pipelines, select the modules/pipelines to search for and click **Search**:

- **Search for new modules to install**: Displays modules and pipelines that are in the Broad repository and not on your server.
- **Search for updates of the currently installed modules**: Displays modules and pipelines where the Broad repository contains a more recent version than the version installed on your server.
- **Search for up to date modules**: Displays modules and pipelines where the Broad repository contains versions not installed on your server.
- **Operating system**: Filters the search results to display only modules and pipelines that run the selected operating system platform(s).

For each module and pipeline, GenePattern displays similar information:

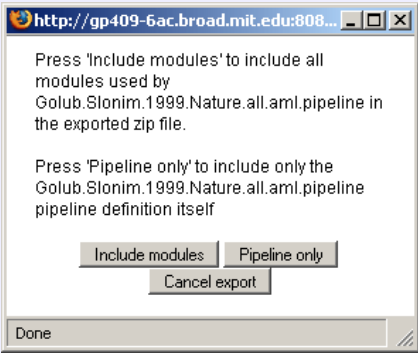
- In the **Name (Version)** column:
 - Module/pipeline name.
 - Drop-down list of available versions. By default, the most recent version is selected.
 - Brief description of the module/pipeline.
 - **documentation** link, which displays the module/pipeline documentation.
 - **download zip** link, which downloads a zip file of the module/pipeline. Downloading the zip file allows you to examine the source files before installing them on your server. You can then install the module/pipeline from the zip file, as described in Exporting and Installing Suites Using Zip Files.
- In the **Module type** column, the category label assigned to the module/pipeline.
- In the **Details** column, the author, size, and operating system requirements for the module/pipeline.

Exporting and Installing Modules & Pipelines Using Zip Files

Zip files provide a convenient means of sending your modules and pipelines to other GenePattern users. You can export a module or pipeline to a zip file. The zip file can then be used to install the module or pipeline on another GenePattern server.

To export a module or pipeline to a zip file:

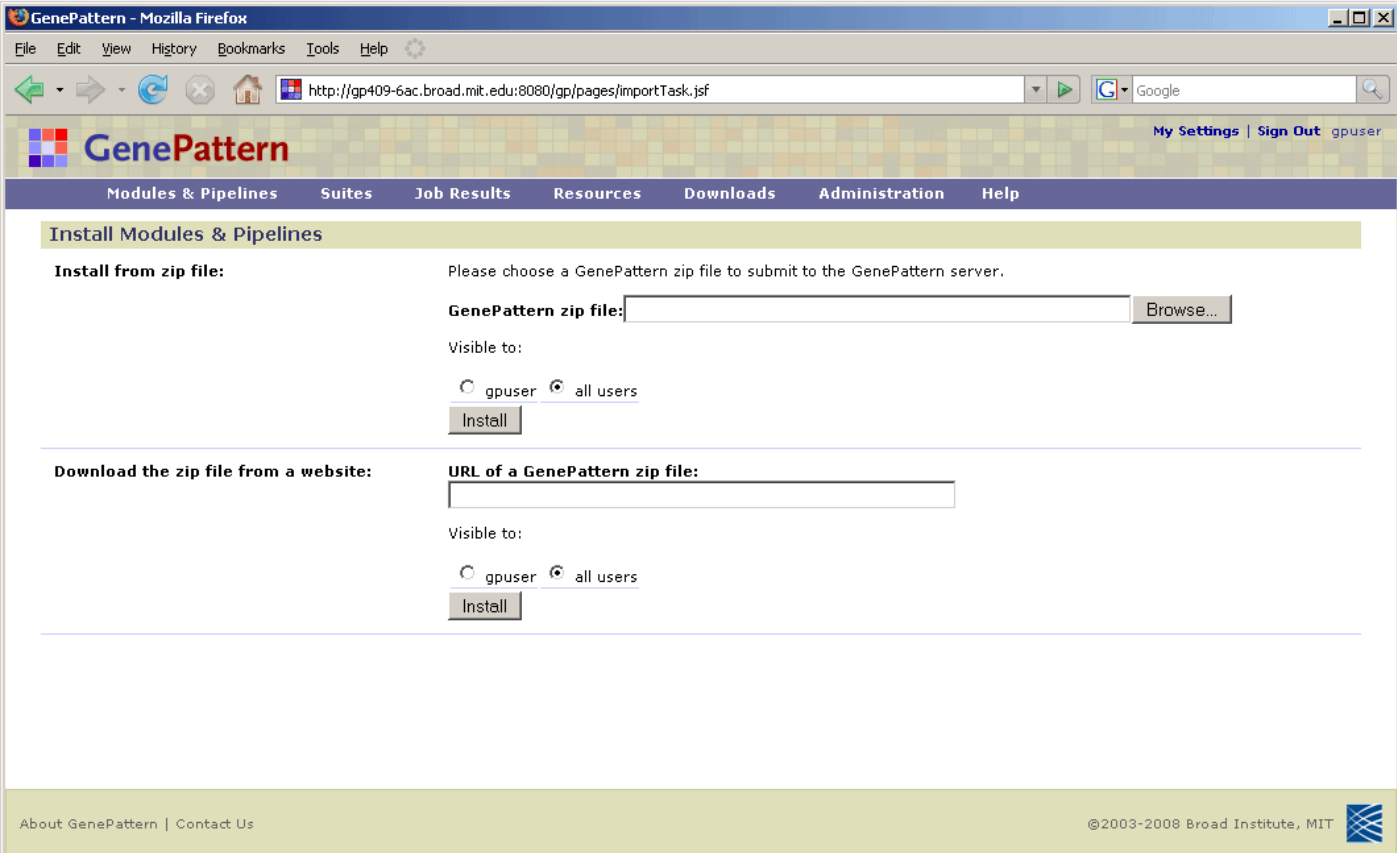
1. Click **Modules & Pipelines** to display the GenePattern home page.
2. In the Modules & Pipelines pane, select the module or pipeline to export. GenePattern displays the analysis parameters (if any).
3. Click **Export**.
4. If you are exporting a pipeline, GenePattern displays the following dialog:



- o Click **Include modules** to create a zip file that includes all modules in the pipeline.
 - o Click **Pipeline only** to create a zip file that contains only the pipeline definition. (If you install a pipeline without installing the modules that it uses, when you run the pipeline, GenePattern lists the missing modules and prompts you to edit the pipeline or install the modules.)
5. GenePattern creates a zip file that contains the module/pipeline and prompts you to save the file to your local drive.

To install a module or pipeline from a zip file:

1. Click **Modules & Pipelines>Install from zip**. GenePattern displays the Install Modules & Pipelines page, as shown below.
2. Select the file to install in one of two ways:
 - o To select a zip file from a directory, click **Browse**.
 - o To select a zip file from a web site, enter the URL for the zip file.
3. Use the **Visible to** radio buttons to select a private or public installation:
 - o Your user name: Choose this option for a private installation. Only you (or an administrator) can view and run the module/pipeline.
 - o All users (default): Choose this option for public installation. Any user connected to this server can view or run the module/pipeline.
4. Click **Install**.

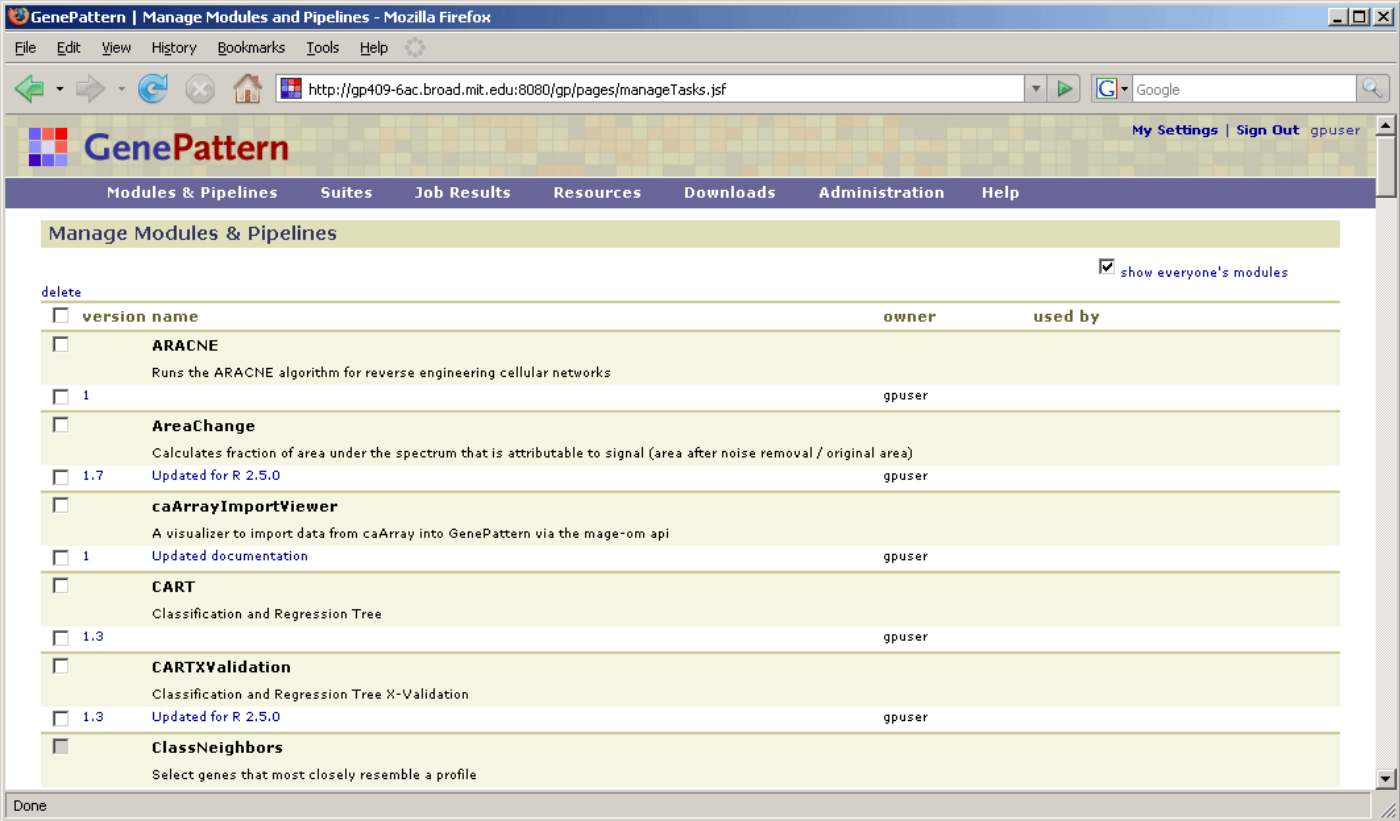


Managing Modules & Pipelines

Click *Modules & Pipelines>Manage* to display the Manage Modules & Pipelines page. From this page, you can

- View the modules/pipelines installed on your GenePattern server. To display only your own modules/pipelines, clear the *show everyone's modules* check box. To display all modules/pipelines, select the check box. Private modules owned by other users are not visible unless you are an administrator.
- Delete modules/pipelines from your GenePattern server.

Note: If the check box next to the module/pipeline is grayed out, you cannot delete it. Typically, this occurs when the module/pipeline is used by another module, pipeline, or suite. To delete the module/pipeline, you must first delete the module, pipeline, or suite that uses it or remove it from that module, pipeline, or suite.



Installing Suites from the Repository


The Broad Institute maintains a repository of suites that are freely available to the public. To install these suites on your GenePattern server:

1. Click *Suites>Install from Repository*. GenePattern displays the Install Suites from Repository page, as shown below.
2. Select the suites to install.
Installing a suite installs its modules, if they are not yet installed on the server.
3. Click *Install Checked*.

Use the top section of the form to find the suites to install. To update the list of suites, select the suites to search for and click **Search**:

- **Search for new suites to install:** Displays suites that are in the Broad repository and not on your server.
- **Search for updates of the currently installed suites:** Displays suites where the Broad repository contains a more recent version than the version installed on your server.
- **Search for up to date suites:** Displays suites where the Broad repository contains versions not installed on your server.


For each suite, GenePattern displays similar information:

- In the **Name (Version)** column:
 - Suite name. Click the  menu icon next to the name to display the suites menu.
 - Drop-down list of available versions. By default, the most recent version is selected.
 - One-line description of the suite.
 - **download zip** link, which downloads a zip file of the suite. Downloading the zip file allows you to examine the source files before installing them on your server. You can then install the suite from the zip file, as described in [Exporting and Installing Suites Using Zip Files](#).
- In the Modules column, a list of the modules and pipelines in the suite.

Exporting and Installing Suites Using Zip Files

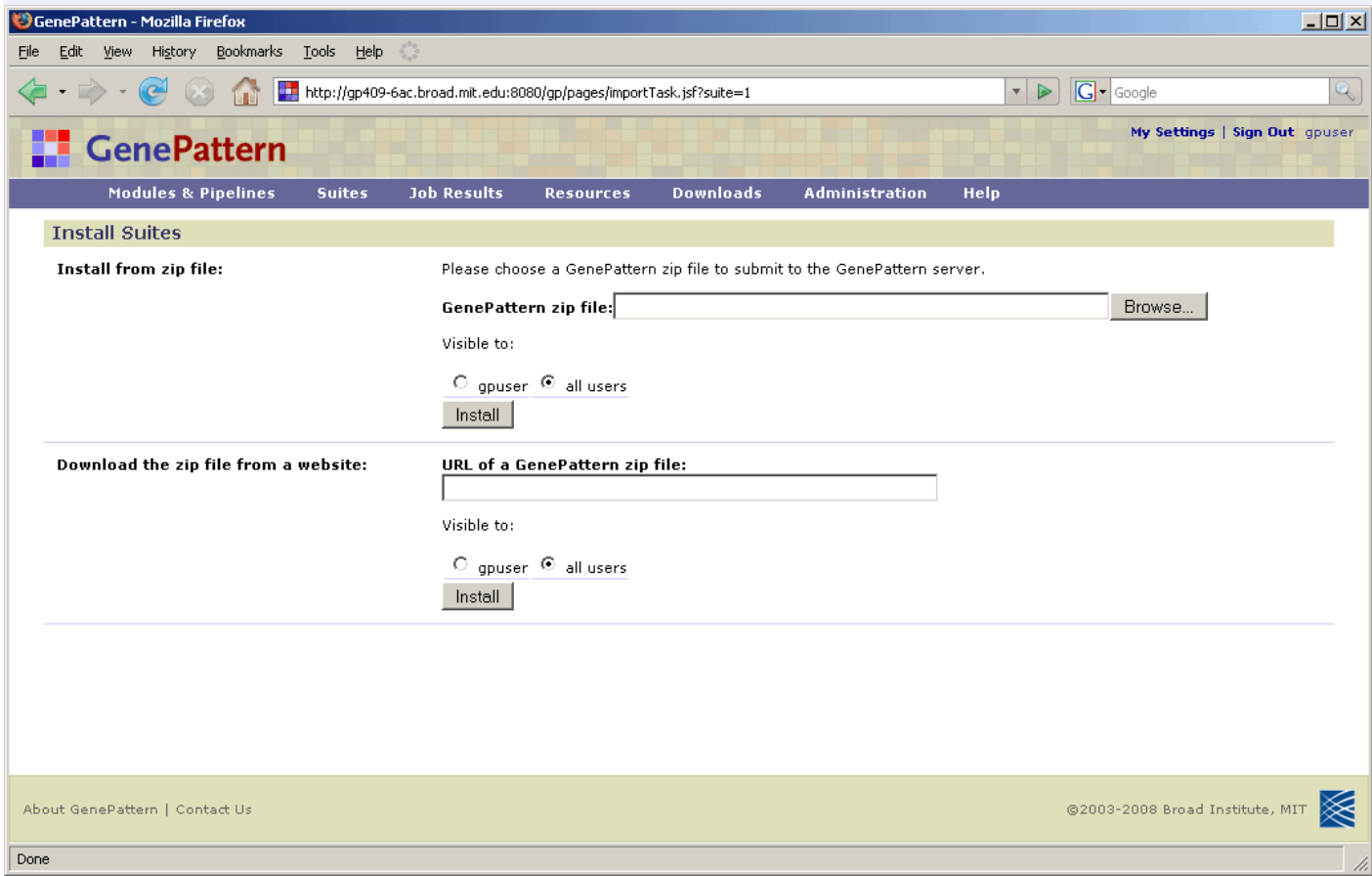
Zip files provide a convenient means of sending your suites to other GenePattern users. You can export a suite to a zip file. The zip file can then be used to install the suite on another GenePattern server.

To export a suite to a zip file:

1. Click **Suites>Manage**.
2. Click the  menu icon following the name of the suite that you want to export. GenePattern displays the suite menu.
3. Click one of the following:
 - **Export excluding dependents:** Creates a zip file that contains the definition of the suite, but not the modules or pipelines in the suite. Installing the suite from this zip file will not install any modules or pipelines in the suite; they must already be installed on the GenePattern server or be installed separately.
 - **Export Including dependents:** Creates a zip file that contains the definition of the suite, as well as the modules and/or pipelines in the suite. Installing the suite from this zip file will also install the modules and pipelines in the suite (unless they are already installed on the GenePattern server).


To install a suite from a zip file:

1. Click **Suites>Install from zip**.
2. Select the file to install. You can identify the zip file using a file specification or a URL.
3. Use the **Visible to** radio buttons to select a private or public installation:
 - o Your user name: Choose this option for a private installation. Only you (or an administrator) can view and run the module/pipeline.
 - o All users (default): Choose this option for public installation. Any user connected to this server can view or run the module/pipeline.
4. Click **Install**.



Managing Suites

Click **Suites>Manage** to display the Manage Suites page. From this page, you can

- View the suites installed on your GenePattern server.
- Delete suites from your GenePattern server. If the check box next to the module/pipeline is grayed out, you cannot delete it. Deleting a suite does not delete its modules.
- Click the  menu icon next to a suite name to display the suites menu.

Managing the GenePattern Server

You can use the GenePattern server hosted at the Broad Institute, install a local GenePattern server for your own use, or install a networked GenePattern server to be used by several people. The [Concepts Guide](#) explains the benefits of each approach.

If you are using the Broad-hosted GenePattern server at <http://genepattern.broadinstitute.org/gp/>, you do not need to manage the server; the GenePattern team does it for you. If you are installing a local GenePattern server, you will most likely use the default server settings; however, you are the GenePattern server administrator for your server and have full access to configuration options described in this section. If you are installing a networked GenePattern server for use by several users, read this section carefully. You are the GenePattern server administrator and will want to configure the server appropriately for your site.

GenePattern can be run standalone on a small machine or separated into its client and server components to take advantage of a more powerful compute server. When you install a GenePattern server, you set basic server configuration options. If you are installing a local GenePattern server for your own use, you generally do not need to modify the server configuration. If you are the server administrator for a networked GenePattern server, you generally want to modify several of the GenePattern configuration options described in this section:

- Creating Groups and Administrators
- Modifying Server Settings
- Setting the Java Version
- Using a Queuing System
- Using Different Versions of R
- Increasing Memory Allocation
- Securing the Server
- Changing the GenePattern Database (HSQL to Oracle)

Creating Groups and Administrators

The GenePattern configuration file `GenePatternServer/resources/userGroups.xml` defines groups and group membership. The [Users and Groups server settings page](#) lists all registered users and the groups to which they belong.

The GenePattern installation defines one group, administrators, which includes all GenePattern users:

```

<!-- map of users to groups -->
<userGroups>
<group name="administrators">
    <user name="*" />
</group>
</userGroups>

```

Administrators Group

Members of the administrators group have full access to the GenePattern server and all jobs run on the server. Therefore, when all users are administrators, GenePattern has no concept of "private" data. **Initially, all users are administrators.**

```

<!-- map of users to groups -->
<userGroups>
<group name="administrators">
    <user name="*" />
</group>
</userGroups>

```

To maximize data privacy, minimize the number of users in the administrators group. For example, add exactly one person to the administrators group and only that one administrator can view all jobs run on the server. Other users can view their own jobs and jobs that have been explicitly shared.

Editing Groups

To create groups or change group membership, edit the `userGroups.xml` file. The XML syntax is simple but must be followed carefully. The rules are as follows:

- Use the `<group>` element to create a group. You can create any number of groups. The group names must be unique. They should include only alphanumeric characters, periods (`.`), and underscores (`_`).
- Use the `<user name>` element to add members to a group. You can add any number of users to a group. A user may be in any number of groups. Setting `user name = "*"` adds all users to a group.
- **Warning:** Do not delete the administrators group. GenePattern requires it.

The following edited `userGroups.xml` file adds exactly two users to the administrators group and creates a new group, `mjones_lab`:

```

<!-- map of users to groups -->
<userGroups>
<group name="administrators">
    <user name="j smith" />
    <user name="mj ones" />
</group>
<group name="mjones_lab">
    <user name="mj ones" />
    <user name="j doe" />
    <user name="sfederan" />
</group>
</userGroups>

```

Renaming a group does not update shared analysis results. Members of a group can share analysis results. If you rename a group, from `old_name` to `new_name` for example, the users in the `old_name` group are now in the `new_name` group. Analysis results that they shared however were shared with the `old_name` group. Each user who shared job results with the `old_name` group should edit the share options for the job and share the job results with the `new_name` group.

Modifying Server Settings

To modify the configuration of your GenePattern server, use the Server Settings page:

1. Click *Administration>Server Settings* to display the Server Settings page.
2. From the Server Settings pane, select the server setting that you want to modify. GenePattern displays a page of related server configuration options.
3. Modify and save the server configuration options.
4. Optionally, return to step 2 to change additional settings.

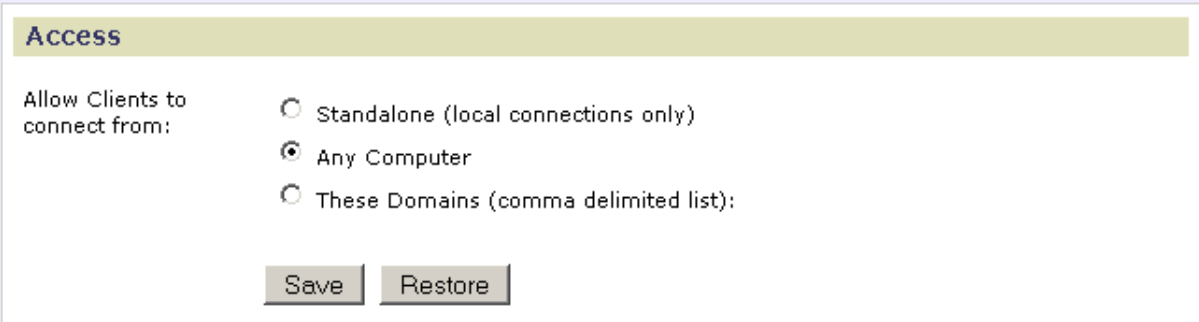
The following table summarizes the server settings. For more detail, click a link in the table.

Access	Specify which clients have access to the server.
Advanced	Specify software source directories and other low-level configuration options.
Command Line Prefix	Specify commands and qualifiers to be prepended to the command line used to invoke a module or pipeline.
Custom	Create new server configuration options.
Database	Specify configuration options for the GenePattern database.
File Purge	Specify how long files remain on the server before being deleted.
GenePattern Log	Display the log file for the GenePattern server.
Programming Languages	Specify the root directories for the programming languages used by GenePattern and the Java flags to be added to Java command lines executed by the server.
Proxy	If your organization has a web proxy between the GenePattern server and the internet, specify the proxy information required to access the internet.
Repositories	Specify the URL used to access the module repository and the suite repository.
Shut Down Server	Shutdown the GenePattern server.
System Message	Broadcast a message to all users logged into the GenePattern server.
Users and Groups	Display account information for all users, including the groups to which they belong.
Web Server Log	Display the log file for the web server used by the GenePattern server.

Access

Use the Access page to define which GenePattern clients have access to the GenePattern server. The localhost (127.0.0.1) computer cannot be denied access to the locally installed GenePattern server. This prevents you from inadvertently denying yourself access to the server.

Using the Access page to control which computers have access to the GenePattern server is the simplest way to secure your server. You can also control access to your server based on user authentication and user permissions, as described in *Securing the Server*. The Access page filters are applied before any user-specific authentication or permissions are checked. If your computer cannot access the server, you cannot access the server regardless of your username/ password or permissions.



- Click *Standalone* to allow only local clients to connect to the server; that is, you can access this GenePattern server only from the computer that it is running on.
- Click *Any Computer* (default) to allow any client to connect to the server.
- Click *These Domains* to allow only clients from specific domains to connect to the server. Enter a comma-separated list of domains or IP addresses in the text box, for example: `broadinstitute.org, dfci.harvard.edu, mit.edu`.

GenePattern scans all incoming connection attempts. If they match in whole or in part any domain name or IP address in this list, the server allows access; otherwise, the server redirects the connection to a page indicating that the server does not allow access.

Click **Save** to save your changes. Click **Restore** to return to the value set at installation.

Advanced

The Advanced page contains directory specifications for the GenePattern source files and other low-level configuration options. You rarely need to modify these options.

Advanced Configurations

Patch Qualifiers:	<input type="text" value="build.tag,os.arch,os.name,os.version,java.version,sv"/>
Location for Patches:	<input type="text" value="/Applications/GenePatternServer/patches"/>
Location for Ant:	<input type="text" value="<java> -cp <tomcatCommonLib>/tools.jar -jar <tc"/>
Directory for Resources:	<input type="text" value=" ../resources"/>
Directory for Modules:	<input type="text" value=" ../taskLib"/>
Directory for TomcatCommonLib:	<input type="text" value="/Applications/GenePatternServer/Tomcat/common/"/>
Directory for Webapp:	<input type="text" value="/Applications/GenePatternServer/Tomcat/webapps/"/>
Directory for Genepattern Log:	<input type="text" value="logs/genepattern.log"/>
Classpath for Pipelines:	<input type="text" value="<webappDir> <file.separator>WEB-INF <file.separator>"/>
Main Class for Pipelines:	<input type="text" value="org.genepattern.server.webapp.RunPipelineSoap"/>
VM args for Pipelines:	<input type="text" value="-DGenePatternURL=<GenePatternURL>"/>
Installed Patch LSIDs:	<input type="text" value="urn:lsid:broad.mit.edu:cancer.software.genepattern."/>
Number of Threads:	<input type="text" value="20"/>
	<input type="button" value="Save"/> <input type="button" value="Restore"/>

Click **Save** to save your changes. Click **Restore** to return to the values set at installation.

Command Line Prefix

Use the Command Line Prefix page to define commands and qualifiers to be prepended to the command line used to invoke a module or pipeline. For example, use this page to prepend commands and qualifiers that execute modules and pipelines on a cluster farm, as described in Using a Queuing System.

Command Line Prefix

Default Command Prefix:

save
default

The default command prefix is used for all modules that do not have an explicit prefix setting below.

Add New Prefix:

name: content:

add
prefix

name	content
default	

module	command prefix name
--------	---------------------

For Modules ...

ComparativeMarkerSelection	▲
ComparativeMarkerSelectionViewer	
ConvertLineEndings	
ExtractComparativeMarkerResults	
HeatMapView	
MyComparativeMarker	▼

... Use Prefix...

default ▼

add
mapping

To prepend text to all (or most) command lines executed by the GenePattern server:

1. Enter the desired commands and qualifiers in the *Default Command Prefix* field.
2. Click **save default**. GenePattern displays the updated content of the *default* prefix. The *name/content* table in the middle of the form lists the *default* prefix and its content. The previous illustration shows the *default* prefix with no content.

When GenePattern executes a module or pipeline, it constructs the appropriate command line, prepends the *default* prefix to that command line, and then executes the command line.

To prepend text only to command lines that invoke specific modules or pipelines:

1. In the *Add New Prefix* field, enter a name for the prefix and the commands and qualifiers to prepend to the command line.
2. Click **add prefix**. GenePattern creates the new prefix, updates its content, and adds the prefix to the *name/content* table in the middle of the form.
3. At the bottom of the form, select one or more module(s)/pipeline(s), select your new prefix, and click **add mapping**. GenePattern adds the prefix information to the *module/command prefix name* table.

When GenePattern executes a module or pipeline listed in the *module/command prefix name* table, it constructs the appropriate command line, prepends the specified prefix to that command line, and then executes the command line. (When GenePattern executes a module or pipeline *not* listed in that table, it constructs the appropriate command line, prepends the *default* prefix to that command line, and then executes the command line.)

Custom

Use the Custom page to define your own configuration options.

When you create a module, the custom configuration options are available as substitution variables in the module command line. For example, if you define a custom property "foo", you can use <foo> in the command line to pass the value of the custom configuration option to your module. In the Broad repository, for example, the LandmarkMatching and PeakMatch modules use the custom configuration option pepperPrefix.

Custom Settings

Add New Custom Settings:

name: content: [add setting](#)

name	content
------	---------

1. In the *name* field, enter a name for the configuration option.
2. In the *content* field, enter a value for the configuration option.
3. Click *add setting*. GenePattern adds the option to the table at the bottom of the form.

Database

Use the Database Parameters page to set configuration options for the GenePattern database. The following figure shows the HSQL options. You rarely need to change these options.

Database Parameters

Database: **HSQL** Currently, you are using HSQL.

Hibernate Driver Class:

Hibernate URL:

Hibernate User Name:

Hibernate Password:

Hibernate Dialect:

Hibernate Shutdown: true false

HSQL Port:

HSQL Class:

HSQL Args:

HSQL Schema:

Click *Save* to save your changes. Click *Restore* to return to the value set at installation.

File Purge

Use the File Purge page to specify when analysis result files are deleted from the server:

File Purge Settings

Purge Jobs After: days

Purge Time:

File purge settings define how often and, at what time of day, intermediate result files created on the server are deleted (purged). To never delete files set the purge frequency to -1. Note that this could result in large amounts of hard drive space being used over time.

- Use *Purge Jobs After* to specify the number of days the server keeps the analysis result files. To prevent the server from automatically deleting the files, set this value to -1.
- Use *Purge Time* to specify what time of day (24-hour format) the server deletes the files.

Click *Save* to save your changes. Click *Restore* to return to the values set at installation.

GenePattern Log

Use the GenePattern Log page to view warnings and messages generated by the GenePattern server. (Use the Web Server Log page to view messages generated by the web server that GenePattern uses.)

GenePattern Log

Gene Pattern log file from node258.broad.mit.edu on Fri Nov 16 15:28:01 EST 2007

```
2007-11-06 16:44:39.834 ERROR [Thread-43] (GenePatternAnalysisTask.java:3505) - Cannot inst
2007-11-06 16:44:39.843 ERROR [Thread-43] (TaskCatalogBean.java:317) - org.genepattern.serv
2007-11-06 16:59:53.576 ERROR [http-8080-Processor23] (HibernateSessionRequestFilter.java:5
2007-11-07 12:06:31.234 ERROR [Thread-130] (GenePatternAnalysisTask.java:3505) - Cannot ins
2007-11-07 12:06:31.236 ERROR [Thread-130] (TaskCatalogBean.java:317) - org.genepattern.ser
2007-11-07 12:07:28.784 ERROR [http-8080-Processor191] (HibernateSessionRequestFilter.java:5
```

Programming Languages

The Programming Languages page contains two sections. After making changes, click *Save* to save them or *Restore* to return to the value set at installation.

Use **Programming Language Configurations** to specify the root directories for the programming languages used by GenePattern:

Programming Language Configurations

Java:

Perl:

R 2.5 Home:

R 2.0.1 Home:

When you install GenePattern, you install the programming languages used by GenePattern. If you have alternate programming language installations that you prefer to use, use this page to point to those installations. If you would like to use more recent versions of R, see Using Different Versions of R.

Use **Programming Language Options** to increase the memory allocated to modules written in Java and R:

Programming Language Options

Java VMOptions:	<input type="text" value="-Xmx512M"/>
Java Visualizer VMOptions:	<input type="text" value="-Xmx512M"/>
R Options:	<input type="text" value="--no-save --quiet --slave --no-restore"/>
	<input type="button" value="Save"/> <input type="button" value="Restore"/>

Java VMOptions are passed to the VM of most Java language modules. You can use these options to increase the amount of memory allotted to a Java language module if you are experiencing OutOfMemory errors.

For details of java options run
java -X
at a command line.

R Options are passed to the R interpreter of most R language modules. You can use these options to increase the amount of memory allotted to an R language module if you are experiencing vector allocation errors.

For details of R options run
R --help
at a command line.

- In the Java VMOptions field, modify the Java -X flag to increase the memory allocated for Java modules run on the server (non-visualizer modules); for example, -Xmx1024M doubles the amount of memory allocated. You can specify up to the maximum memory size of your server machine. Changes take effect when you stop and restart the server.
- In the Java Visualizer VMOptions field, modify the Java -X flag to increase the memory allocated by default for modules run on the client (visualizer modules). Changes take effect when you next run a visualizer module.
GenePattern users can customize the memory allocated to visualizer modules based on the amount of memory available on their desktop PCs. To customize the memory allocated to visualizer modules run on your own desktop: select My Settings, select Visualizer Memory, and modify the Java -X flag to specify the amount of memory to be allocated to visualizer modules run on your desktop.
- In the R Options field, add the --max-mem-size flag to specify the maximum amount of memory to be used by R modules running on the server; for example, --max-mem-size=1G allocates a gigabyte. You can specify up to the maximum memory size of your server machine. The --max-mem-size flag affects only Windows operating systems. Changes take effect when you stop and restart the server.

You can also increase the amount of memory allocated to the GenePattern server or client. For more information, see [Increasing Memory Allocation](#).

Proxy

If your server is behind a firewall, use the Proxy page to set the HTTP and FTP Proxy information. Without the proxy information, the server cannot download modules, pipelines, or suites from the repository maintained by the Broad Institute. If you do not know the proxy information, contact your systems administrator.

Proxy Settings

Proxy Host:

Proxy Port:

Proxy Username:

Proxy Password:

Proxy settings are needed only to connect to the Module Repository, and only if your organization has a web proxy between you and the internet. Username and password are required only if your proxy requires authentication. The username and password values will not be saved to the config file for security reasons and need to be re-entered after a server restart.

Click **Save** to save your changes. Click **Restore** to return to the values set at installation.

Repositories

Use the Repositories page to identify the location of the repository to be accessed by the GenePattern server when you install modules and pipelines or suites from the repository. By default, it points to the module repository maintained by the Broad Institute. If you would like to implement and maintain a module repository at your site, contact the GenePattern help desk ([gp-help \(at\) broadinstitute.org](mailto:gp-help@broadinstitute.org)).

Module Repositories

Previous Module Repository URLs:

Module Repository URL:

Suite Repositories

Previous Suite Repository URLs:

Suite Repository URL:

Click **Save** to save your changes. Click **Restore** to return to the values set at installation. Click **Remove** to delete the selected URL from the list.

Shut Down Server

You can shutdown the GenePattern server by clicking the link on this page. For easier ways of shutting down the server, see [Exiting from GenePattern](#).

Shut Down Server

Warning: The GenePattern server can NOT be restarted from this page.

[Click here to shut down the GenePattern server.](#)

System Message

Use the System Message page to broadcast a message to all users logged into the GenePattern server.

System Message

Users and Groups

Use the Users and Groups page to view user account information, including the groups to which a user belongs. This page shows only registered users. An administrator can add users to a group (Creating Groups and Administrators) before they register, but the users are not listed on this page until they have created a GenePattern account by clicking the *Registration* link on the GenePattern login page.

Users and Groups

[Reload Users and Groups](#)

Users

user_id	emailgroups	admin?
gpuser		
jdoe	mjones_lab	
jsmith	administrators	✓
mjones	administrators mjones_lab	✓
sfederan	mjones_lab	

Groups

administrators
jsmith
mjones
mjones_lab
jdoe
mjones
sfederan

Web Server Log

Use the Web Server Log page to view messages generated by the web server that GenePattern uses. (Use the GenePattern Log page to view warnings and messages generated by the GenePattern server.)

WebServer Log

Web Server log file from node258.broad.mit.edu on Fri Nov 16 15:28:01 EST 2007

```
2007-11-06 16:44:39.834 ERROR [Thread-43] (GenePatternAnalysisTask.java:3505) - Cannot inst
2007-11-06 16:44:39.843 ERROR [Thread-43] (TaskCatalogBean.java:317) - org.genepattern.serv
2007-11-06 16:59:53.576 ERROR [http-8080-Processor23] (HibernateSessionRequestFilter.java:5
2007-11-06 16:59:53.582 ERROR [http-8080-Processor23] (StandardWrapperValve.java:253) - Ser
```

Setting the Java Version

As of Release 3.2.1, the GenePattern server can be configured to run under Java 5 or Java 6.

Mac OS X 10.6 (Snow Leopard)

When installed on Mac OS X 10.6 (Snow Leopard), the GenePattern server is automatically configured for Java 6.

Mac OS X 10.5 (Leopard)

When installed on Mac OS X 10.5 (Leopard), the GenePattern server is configured for Java 5 by default.

To configure the GenePattern server for Java 6:

1. Confirm that Java 6 is installed.
Tip: Use the 'java -version' command.
2. Stop the GenePattern server.
3. Set your Java Preferences to use Java 6.

Tip: Java Preferences are usually found under Application > Utilities.

4. Right-click StartGenePatternServer.app and Show Package Contents.
5. Edit the Info.plist file and set the JVMVersion to 1.6.
6. Restart the GenePattern server.

Windows

When installed on Windows, the GenePattern server is configured for Java 5 by default.

To configure the GenePattern server for Java 6:

1. Confirm that Java 6 is installed.
Tip: Use the 'java -version' command.
2. Stop the GenePattern server.
3. Edit the StartGenePatternServer.lax file and update the location of the Java executable:

```
# LAX.NL.CURRENT.VM
# -----
# the VM to use for the next launch
lax.nl.current.vm=jre\bin\java.exe
```

4. Edit the StopGenePatternServer.lax file and update the location of the Java executable.
5. Restart the GenePattern server.

Linux

When installed on Linux, the GenePattern server is configured for Java 5 by default.

To configure the GenePattern server for Java 6:

1. Confirm that Java 6 is installed.
Tip: Use the 'java -version' command.
2. Stop the GenePattern server.
3. Edit the StartGenePatternServer.lax file and update the location of the Java executable:

```
# LAX.NL.CURRENT.VM
# -----
# the VM to use for the next launch
lax.nl.current.vm=/tools/pkgs/jdk_1.6.0_12/bin/java
```

4. Edit the StopGenePatternServer.lax file and update the location of the Java executable.
5. Restart the GenePattern server.

Using a Queuing System

Queuing systems such as the Load Sharing Facility (LSF) and the Sun Grid Engine (SGE) allow computational resources to be used effectively. If you have installed a queuing system, you can configure the GenePattern server to use it. On a heavily used server, using a queuing system to execute analysis jobs generally improves performance overall, especially for compute-intensive and long-running jobs; however, short jobs might take slightly longer because they must be dispatched to the queuing system.

To configure the GenePattern server to execute jobs using LSF or SGE:

- Add the GenePatternURL property to the GenePattern configuration file, `GenePatternServer/resources/genepattern.properties`, specifying the URL of your server. For example:

GenePatternURL=http://myserver.company.com:8080/gp/

When you run a pipeline, the GenePattern server uses this URL to construct the links to the output files.

By default, the GenePatternURL property is not set. When you run a pipeline, the GenePattern server derives the URL at run time based on the current IP address of the host server. This is ideal for a user running on a laptop, where the IP address may change at startup. However, if you are using a queuing system, the derived URL is incorrect: it is based on the IP address of the queuing system server rather than the GenePattern server.

- For Sun Grid Engine modify the R2.5 property in the GenePattern configuration file, **GenePatternServer/resources/genepattern.properties**, to quote the <r_flags> options. For example:

```
R2.5=<java> -DR_suppress\=<R.suppress.messages.file> -DR_HOME\=<R2.5_HOME>
-Dr_flags\="<r_flags>" -cp <run_r_path> RunR
```

Modify other similar properties (if any) that were added to support additional versions of R.

- Use the Command Line Prefix options of the Server Settings page to have the GenePattern server add the required options to the command line each time it executes a module.

For example, if you are using LSF, modify the Command Line Prefix options as follows:

1. Click **Administration>Server Settings** to display the Server Settings page.
2. Select Command Line Prefix from the Server Settings pane. GenePattern displays the Command Line Prefix page.
3. Enter the following text in the **Default Command Prefix** field:

```
bsub -K -o lsf_log.txt
```

- The -K flag instructs the bsub command to wait for the job to complete before returning.
 - The -o flag specifies the file to which the job writes standard output and standard error messages.
4. Optionally, set the environment variables BSUB_QUIET and BSUB_QUIET2 to prevent bsub from printing common job messages to standard out:
 - Setting BSUB_QUIET prevents bsub from printing the messages <<Job is submitted to default queue <normal>>> and <<Waiting for dispatch>>.
 - Setting BSUB_QUIET2 prevents bsub from printing the message <<Job is finished>>.

Another alternative is to create a script that sets the environment variables and then executes the job using LSF or SGE. The command prefix would then execute the script. For example:

1. Create the shell script to set the variables and execute the job using LSF. The script executes in the jobResults directory for the job; for example, for job 3248, the script executes in the GenePattern/Tomcat/webapps/gp/jobResults/3248/ directory. The following script sets the environment variables, submits the job to the LSF queue, waits for the job to complete, saves stdout to a new file, stdout.txt, and saves stderr to a new file, stderr.txt. By convention, GenePattern considers a job to fail if there is any output to stderr.

```
#!/bin/bash
#
# Submit the job to LSF
# Save lsf out and err files in the jobResults directory.
# If there is stdout from the job, pipe to stdout of this script.
# If there is stderr from the job, pipe to stderr of this script.
lsf_err=.lsf.err;
cmd_out=cmd.out;
BSUB_QUIET=
BSUB_QUIET2=
export BSUB_QUIET
export BSUB_QUIET2
# submit the job and wait (-K) for the job to complete
bsub -q genepattern -K -o .lsf_%J.out -e $lsf_err "$@" \>$cmd_out
# sleep to allow for NFS delay
sleep 2;
```

```

# If there is stdout from the job, pipe to stdout of this script, then delete the output file
if [ -e $cmd_out ]
then
    cat $cmd_out >&1;
    rm $cmd_out;
fi
# If there is stderr from the job, pipe to stderr of this script then delete stderr file
if [ -e $lsf_err ]
then
    cat $lsf_err >&2;
    rm $lsf_err;
fi

```

2. Click **Administration>Server Settings** to display the Server Settings page.
3. Select Command Line Prefix from the Server Settings pane. GenePattern displays the Command Line Prefix page.
4. Enter the following text in the **Default Command Prefix** field:
`/fully/qualified/path/to/lsf_default.sh`
5. The script shown here saves the lsf log file into the job results directory. In GenePattern, the log files are displayed with the other job result files. If you do not want the log files displayed in GenePattern, edit the `/resources/genepattern.properties` file and set the following property:
`jobs.FileNameFilter=.lsf*`

Using Different Versions of R

In GenePattern, each module definition includes a command line that runs the analysis program. For an R module, the module developer specifies which version of R to use by including the appropriate substitution variable in the command line. For example, the `<R>` variable translates to the full path of the R2.0.1 programming environment and the `<R2.5>` variable translates to the full path of the R2.5 programming environment. Similar variables can be created for other versions of R.

Installing GenePattern (version 3.1 or later) installs R2.5 and sets the `R2.5_HOME` server configuration parameter, which defines the `<R2.5>` variable. If you upgraded from GenePattern 3.0, your GenePattern installation includes R2.0.1 and defines the `<R>` variable.

To add more recent versions of R to your GenePattern installation:

1. Install the new version of R. For example, R2.7. Typically, to install R you will need to:
 1. Go to <http://www.r-project.org/> and select a CRAN mirror.
 2. Locate the source code for the desired version of R.
 3. Download the `<release>.tar.gz` file for that R version.
 4. Untar the file: `tar xvfz <release>.tar.gz`
 5. Configure and install the release:

```

./configure
./make
./make install

```
2. In GenePattern, click **Administration>Server Settings** and go to the Custom page.
3. Add a custom setting to define the `R*_HOME` variable. In this example, the name would be `R2.7_HOME` and the content would be the full path of the R2.7 installation.
4. Add a second custom setting to define the `<R*>` variable. In this example, the name would be `R2.7` (without the brackets) and the content would be:

```

<java> -DR_suppress=<R.suppress.messages.file>
      -DR_HOME=<R2.7_HOME>
      -Dr_flags=<r_flags>
      -cp <run_r_path> RunR

```
5. Exit from GenePattern.
6. Stop and restart the server.
GenePattern can now run modules written for this version of R.

To add R2.0.1 to your GenePattern installation:

1. Install R2.0.1.
2. In GenePattern, click *Administration>Server Settings* and go to the Programming Languages page.
3. Set the R 2.0.1 Home parameter to the full path of the R2.0.1 installation. This defines the <R> variable.
4. Click Save to update the GenePattern server configuration.
5. Exit from GenePattern.
6. Stop and restart the server.

GenePattern can now run modules written for R2.0.1.

Increasing Memory Allocation

GenePattern allocates memory to the server, to the "client" (the computer you are using to access GenePattern), and to individual modules. When a module fails with an out of memory error, you can try increasing the amount of memory allocated to the server, the client, or the module.

To increase the amount of memory allocated to a module written in Java or R, click *Administration>Server Settings*. The Programming Languages page (Programming Language Options) provides several options for increasing Java and R memory options.

To increase the amount of memory allocated to the server and/or the client, follow the instructions for your platform:

Mac OS X

1. Right-click on the file `GenePattern/Tomcat/StartGenePatternServer` (server) or the `GenePatternClient/GenePatternClient` (client).
2. Select *Show Package Contents* from the pop-up menu. The Contents directory should open in the finder.
3. In the Contents directory, double-click the `Info.plist` file. This should open the Property List Editor program.
4. Add the child `VMOptions` under the `Java` node.
5. Change the Class of the added `VMOptions` node to 'Array'.
6. Add the child with Class 'String' with the value - `Xmx512M`. You can replace the value 512 with the maximum amount of memory in MB that you want the GenePattern Client to use.

Windows and Linux

1. Edit the configuration file `GenePatternServer/StartGenePatternServer.lax` (server) or `GenePatternClient/GenePatternClient.lax` (client).
2. In either file, look for the entries noted below and increase these values (for example, double the value) up to the maximum memory size of the machine you are using. (Note: Windows limits the total space available to a process to 2 GB. Some of that is used for overhead, so slightly less is really available to the JRE.)
 - o `lax.nl.java.option.java.heap.size.initial`
 - o `lax.nl.java.option.java.heap.size.max`

Securing the Server

Secure the GenePattern server to control who has access to which operations. Since GenePattern is primarily a web application (including SOAP interfaces) running on a web server, general approaches for securing web servers are applicable to the GenePattern server. In addition, GenePattern provides several security features that can easily be used by non-technical users to control access to the server.

This section describes several ways to secure the GenePattern server:

- Access Filtering
- Password Protection
- User Accounts
- User Permissions
- User Authentication and Authorization
- Secure Sockets Layer (SSL) Support

Access Filtering

Use the Access page to define which GenePattern clients have access to the GenePattern server. This is the simplest way to secure your GenePattern server.

Access filtering prevents users from connecting to the GenePattern server unless they come from a known computer. If your computer cannot access the server, you cannot access the server regardless of your username/password or permissions. The localhost (127.0.0.1) computer cannot be denied access to the locally installed GenePattern server. This prevents you from inadvertently denying yourself access to the server.

To use access filtering (as described in [Modifying Server Settings](#)):

1. Click **Administration>Server Settings**.
2. Use the Access page to determine which clients have access to your GenePattern server:



- o Click **Standalone** to allow only local clients to connect to the server; that is, you can access this GenePattern server only from the computer that it is running on.
- o Click **Any Computer** (default) to allow any client to connect to the server.
- o Click **These Domains** to allow only clients from specific domains to connect to the server. Enter a comma-separated list of domains or IP addresses in the text box, for example: **broadi nsti tute. org, dfci . harvard. edu, mi t. edu**.
GenePattern scans all incoming connection attempts. If they match in whole or in part any domain name or IP address in this list, the server allows access; otherwise, the server redirects the connection to a page indicating that the server does not allow access.

Password Protection

By default, the GenePattern server requires only a user name to authenticate a GenePattern user. You can easily add password protection by modifying the GenePattern server properties.

To add password protection, modify the GenePattern server properties:

1. Edit the GenePattern configuration file, **GenePat ternServer/resources/genepattern. properti es**.
2. Set the requirePassword property to true: `requirePassword=true`.
3. Save the **genepattern. properti es** file.
4. Restart the GenePattern server.

When you add password protection to the server:

- Existing users are assigned a blank password. The first time a user signs in with a blank password, the GenePattern server requires the user to set the password.
- New users are required to register before using the GenePattern server. By registering, the user creates a GenePattern account with an associated username, password, and email address.
- The sign-in screen prompts you for a username and password. If you forget your password, click the **Forgot your password?** link and GenePattern emails you a temporary password.

Assigning passwords to existing user accounts prevents anyone from inadvertently or intentionally logging into and taking control of another user's account.

After adding password protection to the server, set passwords for existing users as follows:

1. Select **Administration>Server Settings>Users and Groups** to list all users registered on the server.
2. Sign into GenePattern using the name of an existing user.

3. When GenePattern prompts you to set a password, select a password for that user.
4. After setting the password, GenePattern displays the Change Email page (My Settings). Set the user's email address if it has not been set. This is the address GenePattern uses to send the user a new password if necessary.
5. Sign out and repeat the process for the next user.
6. After setting passwords for all users, let them know that passwords have been set. You do not need to send the users their passwords. Simply ask users to sign into GenePattern and click the *Forgot your password?* link to have GenePattern send a temporary password.

User Accounts

By default, users create their own accounts by clicking the *Registration* link on the GenePattern login page. To configure GenePattern to allow only administrators to create new accounts:

1. Shut down the server.
2. Edit the file `GenePatternServer/Tomcat/webapps/gp/WEB-INF/web.xml`. Remove `registerUser.jsf` from the `no.login.required.redirect.to.home` parameter value. After the edits, it looks like this:


```
<init-param>
<!-- List of jsf pages that user can access if not logged in. If user requests one of these pages while logged in, he is redirected to the home page. -->
<param-name>no.login.required.redirect.to.home</param-name>
<param-value>login.jsf,forgotPassword.jsf</param-value>
</init-param>
```

 Result: A user cannot access the registration page until she has successfully logged into the server.
3. Edit the file `GenePatternServer/resources/actionPermissionMap.xml`. Add the following line to the `<actionPermissionMap>`:


```
<url link="registerUser.jsf" permission="adminServer"/>
```

 Result: A user must be an administrator to access the registration page.
4. Edit the file `GenePatternServer/Tomcat/webapps/gp/pages/login.xhtml`. Replace the phrase `rendered="#{loginBean.createAccountAllowed and loginBean.showRegistrationLink}"` with `rendered="false">`
 Result: Removes the *Click to register* link from the login page.
5. Restart the server.

To create an account:

1. Login using an administrator account. GenePattern displays the home page.
2. Open the user registration page, `registerUser.jsf`. For example, if the URL for the GenePattern home page is `http://127.0.0.1:8080/gp/pages/index.jsf` change the URL to `http://127.0.0.1:8080/gp/pages/registerUser.jsf`
3. Create the new user account. You are automatically logged in as that new user.
4. To create another new account, logout of the new user account and login using your administrator account.

User Permissions

User permissions determine valid actions for the user. Permissions are based on two configuration files in the `GenePatternServer/resources` directory (the links show the default files):

- `userGroups.xml` defines user groups, as described in [Creating Groups and Administrators](#)
- `permissionMap.xml` defines which user groups have which permissions

A user who belongs to multiple groups is given the most permissive permissions granted to those groups. For example, an administrator who belongs to other groups retains administrator permissions.

To assign or modify user permissions, edit the `permissionMap.xml` file. The XML syntax is simple but must be followed carefully. The rules are as follows:

- To assign permissions to a group, add a `<group>` element to that permission. A `<permission>` element may have any number of `<group>` elements. A `<group>` element may be listed under any number of `<permission>` elements.
- To assign a permission to all groups (and therefore all users), use the syntax `<group name="*" />`. The presence of a group named `*` means that all groups (and therefore all users) have that permission.
- **Warning:** Do not add or remove `<permission>` elements. GenePattern uses them to define the permissions that it requires and implements. The permissions are described in the following table.

By default:

- When you install GenePattern, all groups have the following permissions: `createPrivatePipeline`, `createPublicPipeline`, `createPrivateSuite`, `createPublicSuite`. Administrators have all permissions.
- On the Broad-hosted GenePattern server, all groups have the following permissions: `createPrivatePipeline`, `createPrivateSuite`. Administrators (the GenePattern team) have all permissions.

Note: No explicit permission is required to run public modules/pipelines, or private modules/pipelines that you have created. No explicit permission is required to edit or delete your own modules, pipelines, suites, or jobs.

createModule	Permits creation of a module. Creation refers to any action that adds a module to the server, including create, install from repository, install from zip, and clone.
createPrivatePipeline	Permits creation of a private pipeline (a pipeline visible only to its creator). Creation refers to any action that adds a private pipeline to the server, including create, install from repository, install from zip, and clone. Note: To install the modules in a pipeline, you must have <code>createModule</code> permission.
createPrivateSuite	Permits creation of a private suite (a suite visible only to its creator). Creation refers to any action that adds a private suite to the server, including create, install from repository, install from zip, and clone. Note: To install the modules in a suite, you must have <code>createModule</code> permission.
createPublicPipeline	Permits creation of a public pipeline. Creation refers to any action that adds a public pipeline to the server, including create, install from repository, install from zip, and clone. Note: To install the modules in a pipeline, you must have <code>createModule</code> permission.
createPublicSuite	Permits creation of a public suite. Creation refers to any action that adds a public suite to the server, including create, install from repository, install from zip, and clone. Note: To install the modules in a suite, you must have <code>createModule</code> permission.
adminJobs	Permits viewing and deleting jobs and associated files owned by other users. Users with this permission can delete any job on the server. Typically, only members of the Administrators group are given this permission.
adminModules	Permits viewing and deleting private modules owned by other users. Permits deleting public modules. Note: No explicit permission is required to view public modules.
adminPipelines	Permits viewing and deleting private pipelines owned by other users. Permits deleting public pipelines. Note: No explicit permission is required to view public pipelines.
adminSuites	Permits viewing and deleting private suites owned by other users. Permits deleting public suites. Note: No explicit permission is required to view public suites.
adminServer	Permits access to <i>Administration>Server Settings</i> and all actions on the Server Settings page, including modifying server settings and shutting down the server. Users with this permission are considered to be GenePattern administrators. On the Users and Groups page, a checkmark in the <i>admin?</i> column indicates that a user has this permission. Typically, only members of the Administrators group are given this permission.

User Authentication and Authorization

You can configure the GenePattern server to provide password protection, restrict creation of user accounts, and assign permissions based on groups. Additional or alternative authentication and authorization mechanisms can be added to the server by an administrator with programming experience. The remainder of this section is written for such a programmer. **Note:** The links in this section display the source code for the default GenePattern installation, which should be used as the starting point for any modifications.

Authentication

The authentication filter, `AuthenticationFilter.java`, controls whether a user can log into the server (typically based on username and password). The easiest way to modify GenePattern authentication is by implementing the `IAuthenticationPlugin.java` interface:

1. Implement the `IAuthenticationPlugin` interface. Use the `IAuthenticationPlugin.java` file as the starting point. Comments in the file provide the specification. For example, create a `MyCustomGenePatternAuthenticati on. j ava` interface.
2. Compile the interface and add it to the classpath for the GenePattern server web application.
3. Modify the `authenticati on. cl ass` property in the GenePattern configuration file, `GenePatternServer/resources/genepattern. properti es`, to point to the new interface. For example:
`authenticati on. cl ass=org. genepattern. server. auth. MyCustomGenePat ternAuthenticati on`
4. Restart the GenePattern server for the changes to take effect.

See <ftp://ftp.broadinstitute.org/pub/genepattern/src/gp-custom-auth.zip> for an example project that prepares a custom authentication jar file for deployment to your local GenePattern server.

If the `IAuthenticationPlugin` interface methods do not provide enough flexibility, you can modify the authentication filter.

Authorization

The authorization filter, `AuthorizationFilter.java`, controls which GenePattern operations (web pages) the user can access. As described in `User Permissions`, permissions are based on two configuration files: `userGroups.xml`, which defines user groups, and `permissionMap.xml`, which defines which groups have access to which permissions.

Organizations that have user groups defined in an external system can use those groups rather than using the `userGroups.xml` file. To have the authorization filter use external user groups rather than the `userGroups.xml` file, implement the `IGroupMembershipPlugin.java` interface:

1. Implement the `IGroupMembershipPlugin` interface. Use the `IGroupMembershipPlugin.java` file as the starting point. Comments in the file provide the specification. For example, create a `MyCustomGroupMembershi pPl ugi n. j ava` interface.
2. Compile the interface and add it to the classpath for the GenePattern server web application.
3. Modify the `group. membershi p. cl ass` property in the GenePattern configuration file, `GenePatternServer/resources/genepattern. properti es`, to point to the new interface. For example:
`group. membershi p. cl ass=org. genepattern. server. auth. MyCustomGroupMembershi pPl ugi n`
4. Restart the GenePattern server for the changes to take effect.

To assign permissions to a group authorized through the `IGroupMembershi pPl ugi n` interface, include the group in the `permi ssi onMap. xml` file.

If the `IGroupMembershipPlugin` interface methods do not provide enough flexibility, you can modify the authorization filter.

Modifying the filters

The authentication and authorization filters are servlet filters installed in front of the GenePattern web application in the `GenePatternServer/Tomcat/webapps/gp/WEB-INF/web.xml` file. To implement an alternative authentication (or authorization) filter:

1. Write and compile a new `ServletFilter` that that performs the desired authentication (or authorization).
2. Place the jar file containing the new `ServletFilter` into the following directory:
`*/GenePatternServer/Tomcat/webapps/gp/WEB-INF/lib`
3. Modify the GenePattern server's `web.xml` document.
`*/GenePatternServer/Tomcat/webapps/gp/WEB-INF/web.xml`
Note: It is important to maintain the existing order of the servlet filters in the `web.xml` document as they are used in the order they are defined in the document. The Authentication filter must come before the Authorization filter for the Authorization filter to work.
4. Change the definition of the `AuthenticationFilter` (or `AuthorizationFilter`) to use the class that you have provided.
5. Add any necessary configuration elements that it requires.
6. Restart the GenePattern server for the changes to take effect.

Note: If you look at the code for the default Authentication Filter (`AuthenticationFilter.java`), you will see that it allows requests through that have a parameter

called `jsp_precompile` that have come from the `Local host`. If you do not allow these requests through unauthenticated, you will see a series of errors when you start the GenePattern server as it attempts to precompile the JSP pages. These are **not** fatal errors, but they slow down server response for users the first time that pages are accessed following a server restart.

Secure Sockets Layer (SSL) Support

This section describes how you can modify the GenePattern web application to run on a web server that is configured to use the HTTPS protocol, where essentially the regular http requests are routed through a secure sockets layer (SSL) making them much harder for hackers to access. If you have installed your GenePattern server onto a web server other than the default Tomcat instance it is distributed with, configure your web server according to its instructions and then follow Step 2 below.

Note: When running under SSL, programming language clients and the GenePattern Desktop Client may not be able to connect to your GenePattern server.

Step 1. Configure Tomcat for SSL support

Follow the instructions available at <http://tomcat.apache.org/tomcat-5.5-doc/ssl-howto.html> to configure the Tomcat instance for using SSL. In doing so, you will modify the Tomcat configuration file, which is located in the `GenePatternServer/Tomcat/conf` directory.

Step 2. Configure GenePattern for SSL

Once the Tomcat (or other web server) has been configured for SSL, modify the GenePattern configuration file, `GenePatternServer/resources/genepattern.properties`, to ensure that its properties are in synch with the web server:

1. Add a new key, `java.net.ssl.trustStore=<path to keystore>`.
This should point to the keystore you created when configuring Tomcat (above) or some keystore that GenePattern can use to establish SSL connections.
2. Modify the value for the key `GENEPATTERN_PORT` to use the https port you selected when configuring Tomcat (above).
3. Modify the value for the key `GenePatternURL` to use the https protocol and the https port you selected, for example:
`http://localhost:8080/gp` becomes `https://localhost:8443/gp`

Save the `genepattern.properties` file and restart your server. Any bookmarked links to your GenePattern server must be updated to the new protocol and port.

Changing the GenePattern Database (HSQL to Oracle)

The GenePattern database has been implemented in both HSQL and Oracle. The GenePattern installation builds the HSQL database and sets the GenePattern server properties to reference that database. To use the Oracle implementation instead, build the Oracle database and modify the GenePattern properties to reference that database, as described in the following procedure:

1. Build the Oracle database using the .sql file provided with the GenePattern installation:
`/GenePatternServer/resources/analysis_oracle-30.sql`
2. Edit the `GenePatternServer/resources/genepattern.properties` file:
 1. Locate the Database Parameters section.
 2. Comment out the HSQL statements.
 3. Uncomment the Oracle statements.
 4. Modify the Oracle URL, username, and password for your site. If necessary, modify the Oracle scheme name.
3. Start the GenePattern server.

Documentation Update History

Version Release date Comments

- | | | |
|-------|---------------|--|
| 3.2.1 | November 2009 | Add Setting the Java Version.
Remove Visualizer Modules section (visualizers no longer require Java 1.5). |
| 3.2 | June 2009 | GenePattern 3.2 Release |
| 3.1.1 | July 2008 | Updated Using a Queuing System |

3.1 December 2007 GenePattern 3.1 Release

3.0 April 2007 GenePattern 3.0 Release