User Guide

Getting Started

GenePattern provides access to a broad array of computational methods used to analyze genomic data. Its extendable architecture makes it easy for computational biologists to add analysis and visualization modules. This ensures that GenePattern users have access to a continuously growing repository of new computational methods.

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

- Concepts provides a brief overview of GenePattern, introducing you to its fundamental principles and primary components. All other GenePattern documentation assumes that you are familiar with the basic information provided by Concepts.
- Quick Start provides a 10-minute tutorial on how to run an analysis and review results. Many users find that this is all they need to begin using GenePattern.

Using a GenePattern Server

To use GenePattern, you open a web browser and enter a URL. The URL that you enter is the address of a GenePattern server. For example:

URL	Server
http://genepattern.broadinstitute.org/gp/	**GenePattern public server**. This URL connects you to the GenePattern server hosted at the Broad Institute.
http://localhost:8080/gp/	**GenePattern local server**. You can download and install the GenePattern software,start your own GenePattern server and then use this URL to connect to your own GenePattern local server. For more information, see Starting Your Own GenePattern Server.
The URL for a networked server, for example: http://mycompany.com:8080/gp/	**GenePattern networked server**. A GenePattern administrator at your company or another institution, can install a GenePattern server and make it available for your use. You can then use the URL of the networked GenePattern server to connect to that server.

For example, to use the GenePattern public server:

1. Open a web browser, such as Firefox, Chrome, or Safari.
2. Enter the URL of the GenePattern public server: http://genepattern.broadinstitute.org/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 GenePattern public server requires passwords. By

default, a local server does not.

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

Cannot connect to server: If the 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 GenePattern public server hosted at the Broad Institute, wait a minute and try again.
- If you are using another networked GenePattern server, wait a minute and try again. If the problem persists, contact the GenePattern administrator.
- If you are using your own local server, start the server (see Starting Your Own GenePattern Server) and try again.

Exiting from GenePattern

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

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. Click *Documentation/Tutorials* for a list of GenePattern tutorials and documentation, including:

- Concepts. A brief overview of GenePattern, introducing you to its fundamental principles and primary components.
- Quick Start (10 minutes). A quick hands-on introduction to GenePattern and the GenePattern public server.
- Video Tutorials. A frequently updated series of training videos.
- User Guide (this guide). Detailed instructions for using GenePattern.
- Programmers Guide. Guidelines for writing modules and instructions for accessing GenePattern from the Java, MATLAB, and R programming environments.
- Administrators Guide. Information on how to configure and secure a local or networked GenePattern server.
- File Formats Guide. Descriptions of all file formats and instructions for creating input files.
- In-Depth Articles. Brief articles and tutorials written in response to questions from our users or to share information about new GenePattern features, modules or pipelines.

Starting Your Own GenePattern Server

If you would like, you can install your own local GenePattern server by downloading and installing the GenePattern software. You can install a local server on a standalone machine for your personal use or on a networked machine for use by several people or an entire organization. A local GenePattern server shared by several users is typically referred to as a networked GenePattern server.

For information about the pros and cons of using the GenePattern public server versus your own local GenePattern server, see Concepts:Servers.

Installing Your GenePattern Server

To install a local GenePattern server, follow the instructions provided on the Download GenePattern page.

You use the same installation instructions 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 for personal use, the default server configuration is typically sufficient. If you are installing a networked server for use by an institution, see the Administrators Guide for instructions on how modify the default server configuration.

Starting Your GenePattern Server

After installing a GenePattern server, you must start the server before you can use it.

Windows 7	To start the GenePattern server, right-click on StartGenePatternServer.exe and select *Run as administrator*. While the server is starting, the cursor displays as an hourglass. The server is ready when the cursor returns to normal.
Other Operating Systems	To start the GenePattern server, double-click the Start GenePattern Server icon (shown below). By default, installing GenePattern places this icon on

your desktop.

Windows: While the server is starting, the cursor displays as an hourglass.
The server is ready when the cursor returns to normal.

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.

Using Your GenePattern Server

After starting your local GenePattern server, you can use it by opening a web browser and entering the URL for the server: http://localhost:8080/gp/.

Alternatively, you can click the GenePatternHome.html shortcut icon (shown below), to open a web browser and connect to your server. 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).

Shutting Down Your GenePattern Server

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

Windows: The GenePattern server exits silently.

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

Linux: The GenePattern server exits silently.

User Interface

When first opened in your browser, 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).

Use the **search box** to quickly find a module or pipeline.

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.
- *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 far right pane provides access to several tabs:

- The *Recent Jobs* tab 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 the number of analyses displayed in the Recent Jobs pane.
- The *Uploads* tab lists the files in your uploads directory on the GenePattern server. Click the icon next to a directory or file name to display a menu of commands for working with the uploaded files (see Upload Menu).
- A GenomeSpace-enabled GenePattern installation includes a *GenomeSpace* tab that can be used to help move data files from

GenomeSpace and analysis results to GenomeSpace. For more information, see www.genomespace.org.

Navigation Bar

The navigation bar provides access to GenePattern pages and operations. When you log into a GenePattern server on which GenomeSpace is not enabled, you see the following navigation bar:

If you log into another GenePattern server, such as your own local GenePattern server or a GenomeSpace-enabled GenePattern server, you may see additional items in the navigation bar:

The following table describes all possible options in the navigation bar. Not all of these options are available on the GenePattern public server.

Modules & Pipelines	Display the GenePattern home page.
New Pipeline	Create a pipeline.
New Module[1]	Create a module.
Install from repository[1]	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[1]	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.
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.
GParc[1]	Display the GenePattern Archive web site, which is a repository and community where users can share and discuss their own GenePattern modules.
Administration[1]	Display the Server Settings page.
Server settings[1]	Modify settings that affect the GenePattern server.
Help	Display the GenePattern home page.
Tutorial	Display the Tutorial, which provides a comprehensive hands-on tour of GenePattern.
Video Tutorials	Display the list of GenePattern training videos.
User Guide	Display this guide, which describes how to use GenePattern.
Programmers Guide	Display the Programmers 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.
GenomeSpace	Display the GenomeSpace menu options: *Login* (to GenomeSpace), *GenomeSpace UI*, and *About* (GenomeSpace). A GenomeSpace-enabled GenePattern installation includes several elements that help move data files from GenomeSpace and analysis results to GenomeSpace. For more information, see www.genomespace.org.

[1]Not available on the public GenePattern server.

Job Menu

When you run an analysis, GenePattern displays the job results in the *Recent Jobs* tab:
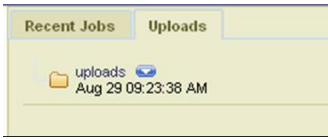

Click the ⬇ icon next to the job name to display a menu of commands for working with that job. For more information, see Working with Analysis Results.

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 View MATLAB Code View R Code	Display the command line that you would use to run this job in the Java, MATLAB, or R programming environments. These commands are useful for programmers who want to access GenePattern from one of these programming environments or from their own applications.

File Menu

When you run an analysis, GenePattern displays the job results in the *Recent Jobs* tab:

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.
Send to <input-file-parameter>	Specify this file as the input file parameter.

Upload Menu

The Uploads pane lists files that you have uploaded to the GenePattern server:

Click the 🔽 icon next to the uploads directory, or a subdirectory or file in that directory, to display a menu of commands for working with that directory or file. For more information, see Running Modules and Pipelines.

Upload Files	Upload a file to the uploads directory on the GenePattern server.
Create Subdirectory	Create a subdirectory under the uploads directory on the GenePattern server.
Delete	Delete the file/directory from the uploads directory on the GenePattern server.
Save	Download the file from the uploads directory on the GenePattern server.
Send to <input-file-parameter>	Specify this file as the input file parameter.
Send Batch to <input-file-parameter>	Specify this directory as the input file parameter. This creates a batch job to analyze each file in the directory.
List of modules	List of modules that accept this type of file as an input parameter. Select an analysis to display its parameters in the center pane. This file is specified as the first input parameter.

Suites Menu

Click *Suites>Manage* to display the currently installed suites:
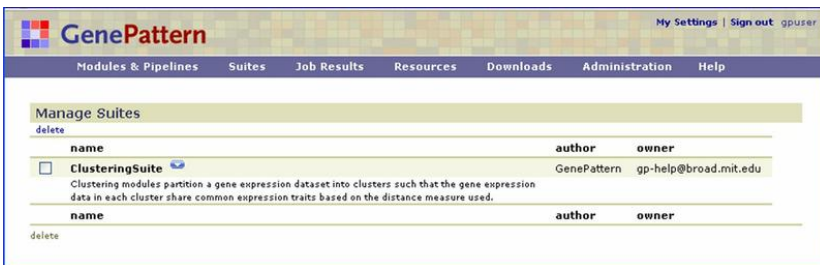
Click the 🔽 icon next to the suite name to display a menu of commands for working with that suite. For more information, see Working with Suites.

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   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
dependents   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.
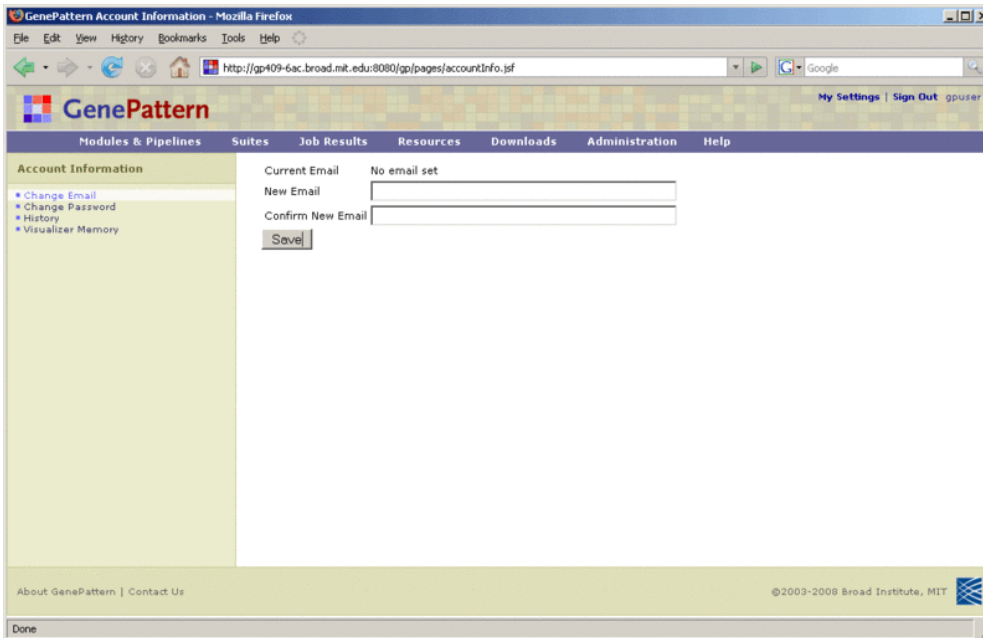
- **Change Email**: Change the email address for your GenePattern account on this server. The GenePattern server uses your email address for the following:
  - When you run a module or pipeline, GenePattern sends the job to the GenePattern client and displays the Job Status page. On that page, you can 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).

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 Concepts.

- Running a Module or Pipeline
- Setting Parameters
- Rerunning an Analysis
- Uploading Files
- Using File Paths

- Batch Processing

Running a Module or Pipeline

To run a module or pipeline:

1. Start typing the module name in the search box at the top of the Modules & Pipelines pane to quickly find the module you need.

2. Select the desired module. The parameters appear in the center pane.
3. Enter values for the parameter fields (see Setting Parameters).
4. 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.
   Tip: You do not have to wait for the job to complete. You can move off of the Job Status page and continue working or even log out of GenePattern.
5. Click *Return to Modules & Pipelines Start* to return to the GenePattern home page. The Recent Jobs tab shows the job that you just ran.
6. 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).

Alternative: A module or pipeline can also be run from an analysis result file or uploaded file. When GenePattern displays such files (for example, in the Recent Jobs tab, the Uploads tab, or the Job Status page), click the 🔽 icon next to the file of interest, as shown below. GenePattern displays the file menu, which lists all modules that accept this type of file as an input file parameter. Select the module or pipeline. GenePattern displays its parameters in the center pane, setting the input file parameter to the file of interest.
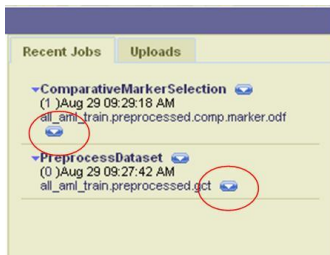
Licensed Modules and Pipelines

Some modules or pipelines may have associated end-user license agreements (EULAs). When you first run a licensed module/pipeline, you will be shown a window containing the license terms, where you can read the terms and choose to accept them or not. (Note: The license terms will differ depending on the module. The terms shown in this image are only an example.)
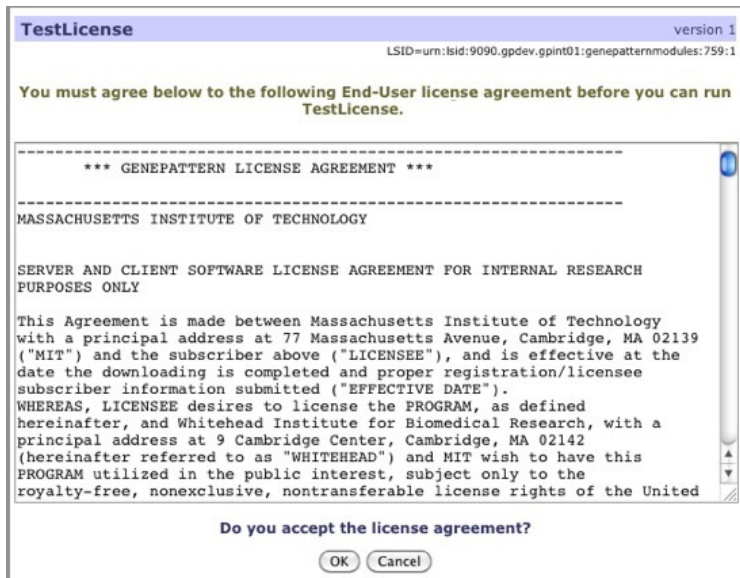
When you click *OK* to accept the license agreement, your acceptance is logged in a database maintained at the Broad. You will not be asked to accept the license terms when you use that particular module/pipeline again until and unless a new version of the module/pipeline is released.

If you choose not to accept the license agreement, you will be unable to run the licensed module or pipeline.

To view a license you have already accepted, click the *properties* link on the licensed module or pipeline page. The modules properties page contains a link to the full-text license.
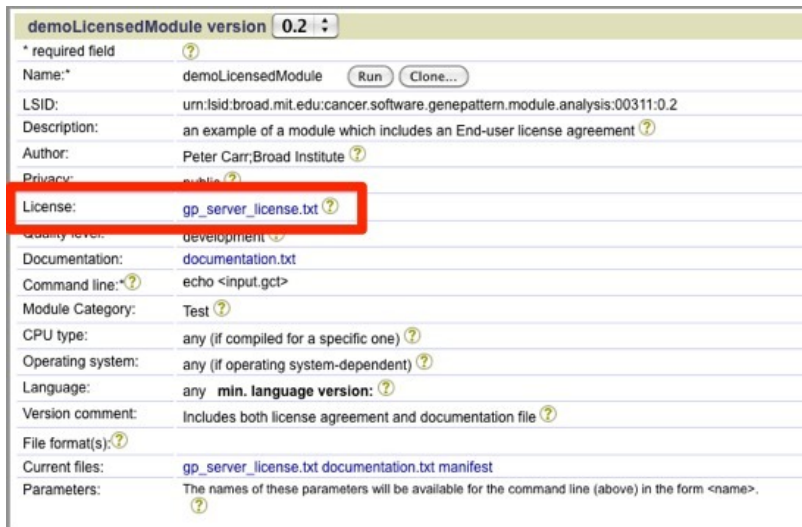
On the pipeline properties page, the link appears above the pipeline listing.
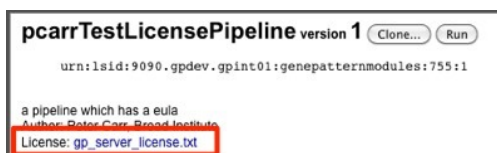
Setting Parameters

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

![Most modules icon] Most modules require one or more input files. There are several ways to choose an input file:

- **Upload a file** (default). Select the *Basic Upload* radio button. Click *Browse* to display the file selection window. Navigate to the desired file and select it. This option **cannot** be used for extremely large files (1.2 GB or more).
  GenePattern uploads the file to the server before running the analysis. The file is stored with the analysis results.
- **Use a previously uploaded file**. Select the *Specify URL* radio button. From the Uploads pane, click the ⬇ icon next to an uploaded file of the desired type and select *Send To parameter-name*. For more information about uploading files, see Uploading Files.
  GenePattern uses the file already stored in the upload directory. A pointer to the file is stored with the analysis results.
- **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 connecting this analysis to the previous analysis, which can be useful for creating pipelines from an analysis result file.
  GenePattern uses the analysis result file stored with the previous job. A pointer to the file is stored with the analysis results.
- **Use a file URL**. Select the *Specify URL* radio button. Copy the URL or FTP address of the file to the entry field.
  GenePattern uploads the file to the server before running the analysis. When the job completes, it deletes the file and stores the URL with the analysis results. (During file transfers, the analysis job is neither PENDING nor RUNNING; deleting a job while it is in this state does not cancel the file transfer.)
- **Use a file path**. Select the *Specify File Path or URL* radio button. Click *Browse Server File System...* to select the file from a local or networked drive. **Note**: For security reasons, file paths are disabled on the GenePattern public server; they can be enabled on a local server. For more information, see Using File Paths.
  GenePattern uses the file identified by the file path. The file path is stored with the analysis results.
- **Use a directory**. Select a directory of files to run the same analysis on several files. For more information, see Batch Processing.

In general, uploading a file using the default method (*Basic Upload*) is fine. However, if you are focused on a particular dataset, it may be faster to upload your files and then analyze the uploaded files. Just be aware that if you delete the uploaded files, you cannot rerun the analyses. If you have extremely large datasets, consider using file paths. You can save a significant amount of time by avoiding file transfers.

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 properties page of the module or pipeline. The module properties page lists the program that implements the analysis. The pipeline properties page lists the modules run by the pipeline.
- *Export*: Create a zip file that contains the module or pipeline. 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.
- *Documentation*: Display the module or pipeline documentation.

7 Display the code (Java, MATLAB, or R) used to run the module with the parameters that you have entered. This can be a useful starting point for programmers writing code to invoke the module.

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 tab.
    - Click *Job Results>Results Summary* to display the Job Results Summary 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*.

Uploading Files

See the video tutorial: Using the GenePattern Uploads Tab

Uploading files to the GenePattern server provides the following benefits:

- Frequently analyzed files can be uploaded once rather than being uploaded each time you run an analysis. This saves you time when running the analyses. It also saves space on the GenePattern server since the file is stored in your upload directory on the server rather than being stored with each analysis.
- Large files (1.2 GB and more) can be uploaded and analyzed. As shown in Setting Parameters, you can specify large input files using an uploaded file, a file URL, or a file path. On the GenePattern public server, you must use an uploaded file or a file URL (file paths are not enabled). If you are running your own local server, we strongly recommend that you use file paths.
- Uploaded files (and analysis result files) can be used as the starting point for an analysis. Click the icon next to the file of interest. GenePattern displays a menu listing all modules that accept this type of file as an input file parameter. Select the module or pipeline. GenePattern displays its parameters in the center pane, setting the input file parameter to the file of interest.

To upload files:

1. Click the Uploads tab to bring it forward.

2. Click the icon next to the Uploads directory. GenePattern displays the uploads menu:
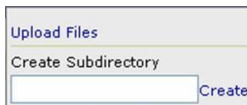
3. Optionally, create a subdirectory:

   1. Enter a subdirectory name and click *Create*.
   2. Click the ⬇ icon next to the subdirectory. GenePattern displays the uploads menu:

4. Click *Upload Files* and, if necessary, grant the upload applet access to run on your local machine. GenePattern opens a Java applet, as shown below, to upload the files.

5. Click ⊕ Add... to browse for the files to upload or drag-and-drop the files into the window.
   Tip: The order in which you add the files to the window determines the order in which the applet uploads the files to the server.

6. Click the ⬆ upload button to begin uploading the files. You can continue working in the applet or in GenePattern while the applet uploads the files to the server.
   Do not close the applet window; it must remain open to complete the upload. The applet displays its progress and estimated time left as it proceeds. Upload time depends on the size of your file, the speed of your network connection, bandwidth available and a variety of other factors. For example, on a T1 connection a 1GB file can take as little as 5-10 minutes to upload, but that same file over a WiFi VPN connection in a room full of other WiFi users can take an hour or more.

7. When the upload completes, **refresh** the GenePattern web browser page. The Uploads tab displays an up-to-date listing of the files in your uploads directory.
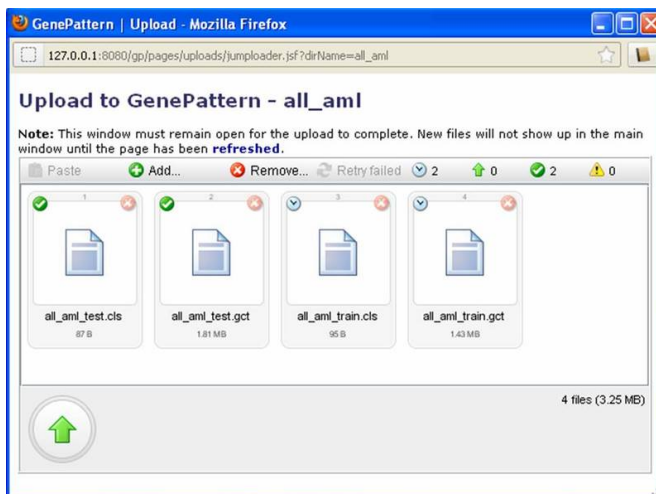
⊕ Add...	Add files to the applet window. Click the icon to display a file browser from which you can select files. Alternatively, drag-and-drop files directly onto the window.
⊗ Remove...	Remove files from the applet window. Click the icon to display a menu from which you can choose to delete selected, pending, finished, failed, or all files from the window.
⊙ 2	Pending. When you have added files to the window, but have not yet clicked Upload, the files are pending.
⬆ 0	Queued. When you click Upload, all pending files are added to the applet's upload queue.

✅ 2	Finished. Files that have been successfully uploaded to the GenePattern server are marked as finished.
⚠ 0	Failed. Files that could not be uploaded to the GenePattern server are marked as failed.
⬆	Upload. Start uploading all pending files.

Deleting Uploaded Files

To delete an uploaded file from the server:

1. Click the Uploads tab to bring it forward.
2. Click the ⬇ icon next to a file or subdirectory. GenePattern displays the uploads menu.
3. Click *Delete.* GenePattern deletes the file or subdirectory from your upload directory on the server.
   - You must delete the files in a subdirectory before you can delete the subdirectory.
   - When you use an uploaded file to run an analysis or as part of a pipeline, the job or pipeline includes a pointer to the file. If you delete the file from your uploads directory, jobs and pipelines that reference the file cannot access the file and, therefore, cannot be (re)run.

Special Considerations

- **Disk space:** Uploading files adds the files to your uploads directory on the server's file system. Your ability to add files to the file system is dependent on the space available and any quotas that may have been set by the server administrators. If you cannot upload files due to disk space, contact the server administrator. If you are using the GenePattern public server, contact gp-help(at)broadinstitute.org.

- **Partial uploads:** If an upload fails to complete (because you stopped the upload or because of a server error or network interruption), the file is marked as '(Uploading)' in the Uploads tab. To reload the file, delete the partially uploaded file and upload it again.

- **Duplicate names:** The server prevents you from uploading multiple files with the same name. If two different files have the same name, rename the file before uploading it to GenePattern.

- **Sharing uploaded files**: As of GenePattern 3.4, you cannot share uploaded files with other users. We plan to implement this feature in a future GenePattern release.

- **Upload window stalls**: On rare occasions, the upload window may remain blank or not fully load. If this occurs, close the applet and click *Upload Files* to reopen the applet. In extremely rare cases, you may need to restart your web browser or reboot your machine. This behavior appears to be caused by known compatibility issues between web browsers (specifically Firefox on Macs) and Java applets.

Using File Paths

Note: For security reasons, file paths are not enabled on the GenePattern public server.

When file paths are enabled on a GenePattern server, you can use file paths to identify input files for modules and pipelines. The GenePattern server can directly access the files stored on your local or networked drives; therefore, the files do not have to be transferred to or stored on the GenePattern server. Avoiding file transfers can save significant upload time and avoiding file storage can save significant amounts of disk space. Therefore:

- If you are running your own local server, we recommend that you use file paths rather uploading files.
- If you are working with large files (2GB or more), consider installing a local GenePattern server and enabling file paths on that server.

When file paths are enabled, the module/pipeline run page includes the *Specify File Path* option. To use a file path as an input file parameter:

1. Select a module or pipeline. The parameters appear in the center pane.
2. Select the *Specify File Path or URL* radio button:

3. Click *Browse Server File System...* to select a file from a local or networked drive.
   GenePattern runs the analysis, using the file identified by the file path. The file path, rather than the file, is stored with the analysis results.

   Tip: You cannot use the *Browse Server File System...* option to select a directory; however, you can use it to select a file in the directory and then edit the path to point to the directory.

To enable file paths on your GenePattern server:

1. Open the `genepattern.properties` file, which is located in the resources directory under your GenePattern server directory.
2. By default, the `allow.input.file.paths` property is set to false. Set it to true:
   `allow.input.file.paths=true`
3. Save the updated `genepattern.properties` file.
4. Restart your GenePattern server.

Optionally, you can also set the `server.browse.file.system.root` property to define a root directory, where the server begins browsing for network files. For example, if you set:

   `server.browse.system.root=/Users/mydata/ngs`

when the user clicks *Browse Server File System...* to select a file from a local or networked drive, GenePattern opens the file selection window to /Users/mydata/ngs.

Batch Processing

See the video tutorial: Batch Execution in GenePattern 3.3.3

Batch processing provides an automated method of running several files through a module or pipeline in parallel. To run a batch job:

1. Place the files to be processed in a subdirectory. You can either upload the files into a subdirectory on the GenePattern server or use a file path to point to a subdirectory on your local machine. See Using an Upload Subdirectory or Using a File Path Subdirectory.
2. Select the module or pipeline to run.
3. Specify the subdirectory as an input file parameter for the module or pipeline. See Using an Upload Subdirectory or Using a File Path Subdirectory.

   To provide multiple input files for a module or pipeline, identify each set of files to be processed by giving them the same name (excluding file extension). For example, to run ComparativeMarkerSelection as a batch job, give each .gct or .res file the same name (excluding file extension) as its matching .cls file. For example: test1.gct, test1.cls, test2.gct, test2.cls. You can place all of the files in one subdirectory or create a separate subdirectory for each input file parameter (for example, one subdirectory for .gct files and another for .cls files). GenePattern processes each set of matching files, ignoring any other files in the directory.

4. Click *Run*. GenePattern starts one analysis job for each input file or valid set of input files. It then displays all of the jobs in the Job Results Summary page.

   Tip: You do not have to wait for the job to complete. You can move off of the Job Results Summary page and continue working or even log out of GenePattern. If you choose to wait for the jobs to complete, periodically **refresh** the Job Results Summary page to update the status of the jobs.

5. Click *Modules & Pipelines* to return to the GenePattern home page. The Recent Jobs pane shows the submitted jobs.
   - Click the id number of a job to display its Job Status page. The Job Status page provides complete information about the job, including its parameters, input files, output files, and current status.
   - Click *Job Results>Results Summary* to redisplay the Job Results Summary page. On the Job Results Summary page, select the batch job from the drop-down list to list only those analysis jobs submitted as part of this batch job.

Using an Upload Subdirectory

To create and run a batch job using an upload subdirectory:

1. Upload the files that you want to run as a batch into a subdirectory of the Uploads directory. For more information, see Uploading Files.
2. Select the module or pipeline to run. GenePattern displays the module run page.
3. For each input file parameter whose input files are in the subdirectory, click the ☁ icon next to the subdirectory and select *Send Batch to <input-file-parameter>*.
4. Enter the remaining parameter values.
5. Click *Run*.

Using a File Path Subdirectory

To create and run a batch job using a file path subdirectory:

1. Install a local GenePattern server and enable the server to accept file paths. For more information, see Using File Paths.
2. Verify that the files that you want to run as a batch are in a subdirectory accessible to your local GenePattern server.
3. Start your local server, open a web browser, and log into GenePattern.
4. Select the module or pipeline to run. GenePattern displays the module run page.
5. For each input file parameter whose input files are in the subdirectory:
   1. Click the *Specify Path or URL* radio button.
   2. Enter the file path of the subdirectory.
      Tip: You cannot use the *Browse Server File System...* option to select a directory; however, you can use it to select a file in the directory and then edit the path to point to the directory.
   3. Click the *Batch Parameter* check box.
6. Enter the remaining parameter values.
7. Click *Run*.

Special Considerations

- **Module restrictions**: To run in a batch job, a module or pipeline must support using a directory name in place of a file name for an input file parameter. To check whether a module can be used in a batch job:
  - Click the ☁ icon next to a directory in the Uploads tab. The modules listed in the menu accept directories as input.
  - When you run the module, the description of the input file parameter should indicate whether it accepts a directory in place of a file name.
  - The module documentation should also indicate whether its input file parameters accepts directories in place of file names.
- **Directory restrictions**: You cannot run batch processing on an empty directory or a directory containing only one file.
- **Errors in batch submission**: If an error occurs while submitting the batch job, GenePattern redisplays the module run page with an error message. Typical causes for errors include:
  - Invalid input folder in batch. The folder is not found.
  - Invalid input folder in batch. You do not have permission to read this folder.
  - Invalid input folder in batch. The input folder contains no files.
  - Invalid input folder in batch. The input folder contains no files which match the input parameter type.
- **Sharing batch jobs**: If you share a batch job with a non-administrator GenePattern user, that user can view the job status pages for the jobs but may not be able to view the input files or successfully reload and rerun the jobs. In particular, if you ran the batch job using an uploads subdirectory, the non-administrator user cannot access the files in your uploads directory and, therefore, cannot view the input files or successfully reload and rerun the jobs. If you ran the batch job using a file path subdirectory, whether the non-administrator user can access the files in that subdirectory determines whether the user can view the input files or successfully reload and rerun the jobs.
- **Programming environments**: GenePattern provides a programming interface (with versions for Java, R, and MATLAB) that allows you to submit jobs in parallel. For more information, see the Programmers Guide.

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 Concepts.

- 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 Job Status 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 *Job Results>Results Summary*.)
Share analysis results	By default, analysis results are private. To share results with other GenePattern users, click the ![Edit Sharing] icon on the Job Status 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: • To save a single file, click the ![icon] 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] 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: • To delete a single file, click the ![icon] icon next to a file and select *Delete*. • To delete all of the files, click the ![icon] 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 HeatMapViewer
    - Large, integrated datasets are often viewed using Integrative Genomics Viewer (IGV) — a high-performance visualization tool that supports a wide variety of data types including sequence alignments, microarrays, and genomic annotations (see the In-Depth Article: Using IGV Through Genepattern)
- **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 GenePattern public 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.

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*. If a job is still running, refresh the page periodically to update its status.

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 shared with the named group. There is an **In Group** entry for each group to which you belong.
- **Batch:** *number* lists all jobs run as a result of submitting this batch process.

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:

3. Modify the options by choosing which groups should have access and what access they should have.
   - **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.

- ○ **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, RNA-seq analysis, flow cytometry, and data preprocessing and conversion. **Visualization modules** display your data and analysis results graphically. If you are unfamiliar with GenePattern modules and pipelines, see Concepts.

- Basic Operations
- Displaying Module Properties
- 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 *Run*. For more information, see Running Modules and Pipelines.
Display module properties	Module properties include the author, the command line used to invoke the module, and the programs used to execute the module. If a module has an end-user license agreement associated with it, you can view the license via the module properties page. To display a module's properties, click *Modules & Pipelines* and select the module. When GenePattern displays the module parameters, click *Properties*.
Send module to other users	Zip files provide a convenient way to send modules to other GenePattern users: • To export a module to a zip file, click *Modules & Pipelines* and select the module to export. When GenePattern displays the module parameters, click *Export*. • To install a module from a zip file, click *Modules & Pipelines>Install from zip*. 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 *Modules & Pipelines>Install from Repository*. 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 *Modules & Pipelines>Manage*. For more information, see Managing Modules & Pipelines.

Displaying Module Properties

To view the definition of a module, display its properties:

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 properties:

Clicking the help icon displays brief descriptions of each field:

- **Name**. Name of the module.
- **LSID**. The Life Science Identifier (LSID) used to uniquely identify a GenePattern module.
- **Description**. Brief description of the module.
- **Author**. The author's name and affiliation (company or academic institution).
- **License.** The End-User license agreement for the module. (This is an optional property and will therefore usually be empty.)
- **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).
- **Quality level**. One of three terms that indicates the author's confidence in the robustness of the module: development, preproduction, and production.
- **Documentation**. Help file that describes the module and provides instructions for its use.
- **Command line**. Command line used to launch the module. Values enclosed in angle brackets are replaced by specific values before the command executes.
- **Module category**. Category under which to list this module. Categories are used to organize modules and pipelines on the GenePattern home page. Pipelines are always assigned to the category name *pipeline*.
- **CPU type**. Indicates the type of CPU required to run the module, or *any* if the module runs on any type of CPU.
- **Operating system**. Indicates the operating system required to run the module, or *any* if the module runs on any operating system.
- **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.
- **Version comment**. Describes changes made to the module in this version.
- **File format(s)**. Lists the file formats of any output files generated by the module.
- **Current files**: Lists the support files packaged with the module, such as executable programs, documentation, and so on.
- **Parameters**: Lists the module parameters, including the file formats of any input files required by the module.

Creating Modules

Only the GenePattern team can create, edit, or install modules on the GenePattern public server. Therefore, to create a module, you must have a local GenePattern server installed (see Starting Your Own GenePattern Server).

Creating a GenePattern module is a two-step process:

1. Find or write a program that executes the desired function. Any program that can be executed from the command line can be run as a GenePattern module. If you are writing the program, you can use any programming language. 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 Programmers 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 module integrator.
2. Define the module.
3. Click *Save* to create the module.

For more information about creating modules and using the module integrator, see the Programmers Guide.

Editing Modules

Only the GenePattern team can create, edit, or install modules on the GenePattern public server. Therefore, to edit a module, you must have a local GenePattern server installed (see Starting Your Own GenePattern Server).

To edit a module:

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

For more information about editing modules and using the module integrator, see the Programmers Guide.

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 Concepts.

- Basic Operations
- Displaying Pipeline Properties
- Creating Pipelines
- Editing Pipelines
- Pipeline Designer

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 properties	Pipeline properties include the pipeline's author, the modules to be run and their parameters. If a pipeline has an end-user license agreement associated with it, you can view the license via the pipeline properties page.

	To display pipeline properties, click *Modules & Pipelines*, select the pipeline, and click *Properties*. For more information, see Displaying Pipeline Properties.
Send pipelines to other users	Zip files provide a convenient way to send pipelines to other GenePattern users. • 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*. For more information, see Exporting and Installing Modules & Pipelines Using Zip Files.
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, create a pipeline by cloning an existing pipeline, or 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.

Back to top

Displaying Pipeline Properties

To view the definition of a pipeline, display its properties:

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

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.

Back to top

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 creates a pipeline that will reproduce the analysis results file and opens the pipeline designer for that pipeline.

   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 as desired. For more information, see Pipeline Designer.
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 (if any).
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 properties for the new pipeline.
5. Click *Edit* to edit pipeline. GenePattern displays the pipeline designer.
6. Edit the pipeline as desired. For more information, see Pipeline Designer.
7. Click *Save* to create the pipeline.

To create a pipeline from scratch:

1. Click *Modules & Pipelines>New Pipeline*. GenePattern displays the pipeline designer.
2. Edit the pipeline as desired. For more information, see Pipeline Designer.
3. Click *Save* to create the pipeline.

Back to top

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 (if any).
3. Open the pipeline designer in one of two ways:
    - Click *Edit*, if it is available. GenePattern displays the pipeline designer.
      This option is visible only if you created this pipeline on this GenePattern server.
    - Otherwise, create a copy of the pipeline to edit:
      1. Click *Properties*. GenePattern displays the pipeline properties.
      2. Click *Clone*. GenePattern prompts you to name the new pipeline.
      3. Enter a name for the pipeline and click *OK*. GenePattern displays the pipeline properties for the new pipeline.
      4. Click *Edit*. GenePattern displays the pipeline designer.
4. Edit the pipeline as desired. For more information, see Pipeline Designer.
5. Click *Save* to save the pipeline.

Back to top

Pipeline Designer

See the video tutorial: Exploring the New GenePattern Pipeline Designer.

When you create or edit a pipeline, GenePattern displays the pipeline designer:
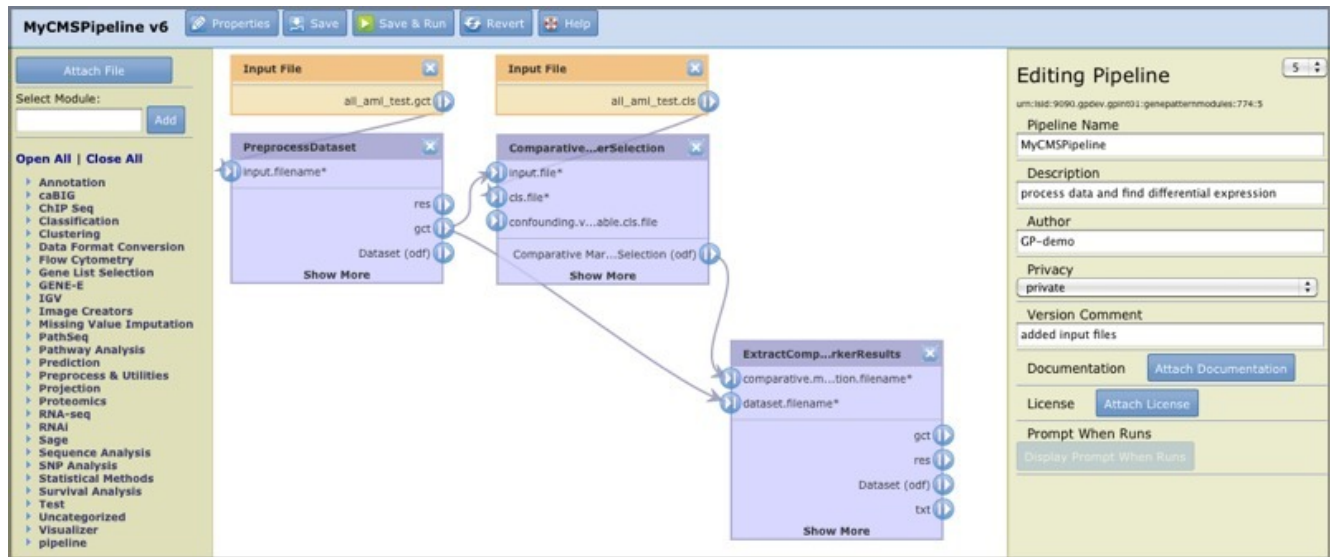
From left to right:

- **Module selection panel**. Use this panel to add a module or a file to the pipeline.
- **Pipeline diagram**. The center panel displays the pipeline diagram. It shows each module in the pipeline, as well as any input files included in the pipeline. Connections between the modules show the flow of data through the pipeline. Use the diagram to reposition modules/files and to add connections between modules/files.
- **Editing panel**. Use the editing panel to view and edit basic pipeline properties (as shown here) or the properties of the object selected in the diagram.

The pipeline diagram toolbar provides the following options:

Displays the basic pipeline properties in the Editing Pipeline panel, as shown here. For more information, see Editing Basic Pipeline Properties.

Saves your changes without closing the designer.

Saves your changes, closes the designer and runs the pipeline.

Loads the last saved version of the pipeline, overwriting any unsaved changes.

Displays this section of the GenePattern documentation.

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

- Editing Basic Pipeline Properties
- Adding Modules/Files
- Editing Module Properties
- Setting Input File Parameters
- Reordering Modules

Editing Basic Pipeline Properties

To edit basic pipeline properties:

1. Click the Edit Properties icon   to display the basic pipeline properties in the editing panel.
2. Edit the pipeline properties.
3. Click *Save* or *Save & Run* to save your changes.

The Editing Pipeline panel displays the following properties:

- **Version**: At the top right 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.
- **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.
- **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.
- **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); ideally, the author's e-mail. 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**: Documentation is strongly encouraged for public pipelines. As a GenePattern user, when you select a pipeline, GenePattern displays a form that includes the pipeline parameters and a *Documentation* button. When you click *Documentation*, GenePattern displays the pipeline documentation. If no documentation file was provided, GenePattern displays a message indicating that no information is available.

  To supply a help file, click *Attach Documentation* and select a (previously created) documentation file. GenePattern uploads the documentation file to the server and displays its name in the pipeline properties panel. Only one help file can be attached to the pipeline. To select a different file, click *Attach Documentation*.

- **License:** To supply a text file containing an end user license agreement to the pipeline, click *Attach License*. Anyone using the pipeline will be

prompted to accept or decline the license terms when they first run the pipeline, and their acceptance will be tracked in a database maintained at the Broad Institute. They will not have to accept the terms again unless and until a new version of the pipeline is created. Only one license file can be attached to the pipeline. To select a different license file, click *Attach License*. This is an optional property; as such it will often be empty.

- **Prompt When Run**: Most modules in a pipeline have one or more parameters. You can supply values for these parameters or mark them as prompt-when-run. When the pipeline runs, GenePattern prompts the user for all prompt-when-run parameters. By default, GenePattern prompts the user for a parameter by displaying its name and description. Optionally, you can supply alternate text for the prompt. Click *Display Prompt When Runs* to view all prompt-when-run parameters in the pipeline and, optionally, modify the prompt text. For more information about setting parameter values, see Editing Module Properties and Setting File Input Parameters.

Back to top

Adding Modules/Files

To add a module to the pipeline:

1. Select the module from the list on the left. The designer adds the module to the diagram.
   Tip: You can select a module in one of two ways:
   - Locate the module in the list and click on it.
   - Start typing the module name in the search box. GenePattern displays module names that include the typed characters. Click the desired module and then click *Add*.
2. Drag the module to the desired location.
   Tip: The diagram is read from left to right and from top to bottom, as you would read a book written in English.
3. Click the module to view and edit its properties in the editing panel.
4. Click *Save* or *Save & Run* to save your changes.

   Note: If you add a module with an end-user license to your pipeline, users who have not accepted that module's license terms will be presented with a license-acceptance window before the pipeline will run.

To specify a file as input to a module in the pipeline, you must first add the file to the pipeline diagram. To add a file to the diagram:

1. Click *Attach File*. GenePattern prompts you for the file.
2. Browse for a file. GenePattern adds the file to the diagram.
3. Drag the file to the desired location. You can now use this file as an input file for one or more modules.
4. Click *Save* or *Save & Run* to save your changes.

The pipeline diagram uses color to distinguish between files, modules, and pipelines. Connections between objects show the flow of data through the pipeline. The following diagram shows a file, a module and a pipeline. The file (all_aml_train.gct) is used as an input file parameter (input.file) for the module. To delete an object and all of its connection, click its ☒ delete icon.
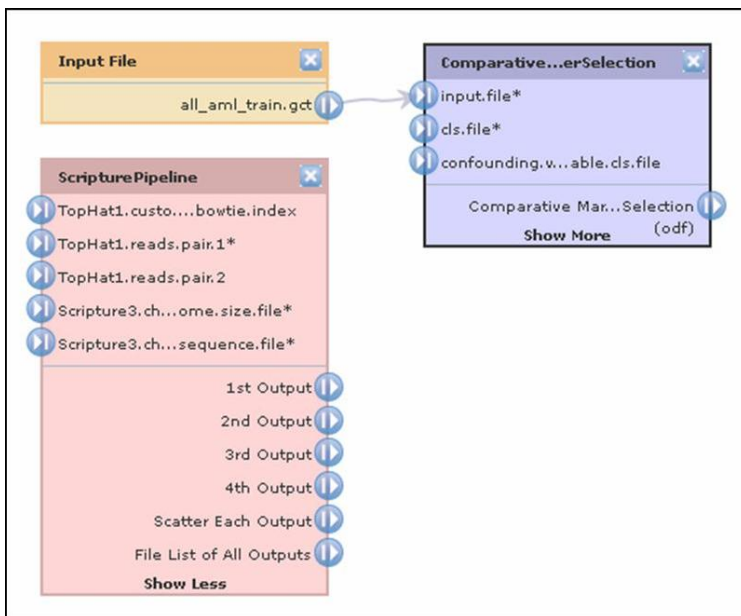
Back to top

Editing Module Properties

To edit a module's properties:

1. Click the module in the pipeline diagram. GenePattern displays its properties in the editing panel.
2. Edit the properties, as shown below.
3. Click *Save* or *Save & Run* to save your changes.

① By default, a pipeline runs the most recent version of a module. The drop-down list shows all versions of the module that are installed on the GenePattern server. To have the pipeline run a different version of the module, select it from the list.

② The Life Science Identifier (LSID) for this module. You cannot create or edit LSIDs. The GenePattern server automatically assigns an LSID to each version of a module.

③ Click the *Documentation* button to display the module documentation.

④ Warnings shown here must be addressed before you can save the pipeline. GenePattern highlights all parameters affected by the warnings.

⑤ All module parameters are listed here. **Input file parameters** are critical and are generally listed first. They control the flow of data through the pipeline. For more information, see Setting Input File Parameters.

⑥ Click the check box next to a parameter to mark it **prompt-when-run**. When the pipeline runs, GenePattern prompts the user for all prompt-when-run parameters in the pipeline. By default, GenePattern prompts the user for a parameter by displaying its name and description. Optionally, click *Set Prompt When Run Display Settings* to supply alternate text for the prompt.

⑦ For most parameters, you enter a value, select a value from a drop-down list, or use the default value supplied by GenePattern.

Back to top

Setting Input File Parameters

In the pipeline diagram, the connections between modules show the flow of data through the pipeline. You modify the flow of data by modifying the input file parameters. The connections in the diagram are a graphical representation of the input file parameter settings. When you click on the module, the editing panel provides a textual representation of the same input file parameter settings.

You can supply the file for an input file parameter in one of three ways:

- Use the output file generated from another module as the input file.
- Prompt the user to provide an input file.
- Specify a file as the input file.

The following pipeline diagram uses the ComparativeMarkerSelection module to illustrate the different ways of supplying input file parameters:
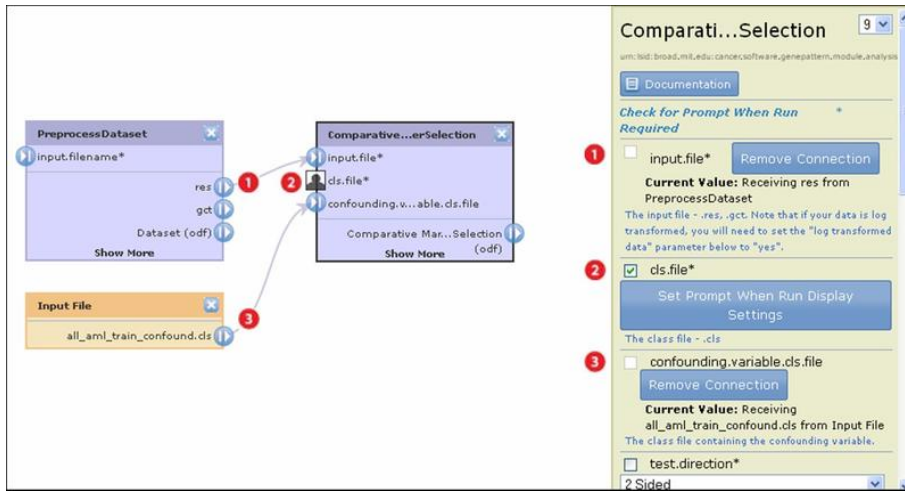
For the input.file parameter, use the output file generated by the PreprocessDataset module.

①
- Diagram: Displays a connection from the res output of PreprocessDataset to the input.file parameter of ComparativeMarkerSelection.
- Editing panel: Displays the current value of the input.file parameter as: Receiving res from PreprocessDataset.

For the cls.file parameter, prompt the user for input.

②
- Diagram: Displays the prompt-when-run icon to indicate that the cls.file parameter has been marked prompt-when-run.
- Editing panel: Displays the prompt-when-run check box for the cls.file parameter as selected.

For the confounding.variable.cls.file, specify the file all_aml_train_confound.cls.

③
- Diagram: Displays a connection from the uploaded file, all_aml_train_confound.cls, to the confounding.variable.cls.file parameter of ComparativeMarkerSelection.
- Editing panel: Displays the current value of the confounding.variable.cls.file as: Receiving all_aml_train_confound.cls from Input File.

The method you use to set the input file parameter depends on how you plan to supply the file:

- To use the output file generated from another module as the input file, click and drag from the generated file (res output of PreprocessDataset) to the input file parameter (input.file parameter of ComparativeMarkerSelection). **Tip**: When drawing a connection, click and drag from the output icon to the input icon .
- To prompt the user to provide the input file, use either the diagram or editing panel to mark the input file parameter (cls.file) as a prompt-when-run parameter. If the input file parameter already has a connection (that is, you have already specified a file for this parameter), you must delete the connection before you can mark it as a prompt-when-run parameter.
  - Diagram: Click the input icon of the input file parameter. When GenePattern displays the Choose Action window, click *Prompt When Run*.

  - Editing panel: Click the check box next to the parameter to mark it as prompt-when-run.
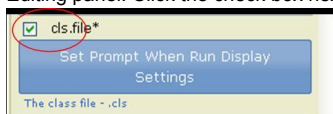
- To specify a file as the input file, you must first attach the file to the pipeline diagram:

- If you have already attached the file to the pipeline diagram (see Adding Modules/Files), click and drag from the attached file (all_aml_train_confound.cls) to the input file parameter (all_aml_train_confound.cls parameter of ComparativeMarkerSelection).
- If you have not yet attached the file the pipeline diagram, click the input icon of the input file parameter. When GenePattern displays the Choose Action window, click *Attach File*. GenePattern prompts you for a file and you browse for the file. GenePattern then adds the file to the diagram and draws the connection from the attached file (all_aml_train_confound.cls) to the input file parameter (all_aml_train_confound.cls parameter of ComparativeMarkerSelection).

Click *Save* or *Save & Run* to save your changes.

Reusing a user-supplied file: Occasionally, a pipeline requires that the same input file be specified for multiple parameters. For example, consider a pipeline with two modules:

1. ComparativeMarkerSelection, with parameters input.file (an expression dataset file) and cls.file (a class file).
2. ExtractComparativeMarkerResults, with parameters comparative.marker.selection.filename (ComparativeMarkerSelection result file) and dataset.filename (the expression dataset file used for ComparativeMarkerSelection).

You want to use the same input file for both the ComparativeMarkerSelection.input.file parameter and the ExtractComparativeMarkerResults.dataset.filename parameter. If the input file that you want to use is either the output file generated from another module (perhaps an expression dataset generated by the PreprocessDataset module) or an uploaded file, this is not a problem. You can connect the file that you want to use to both the ComparativeMarkerSelection.input.file parameter and the ExtractComparativeMarkerResults.dataset.filename parameter.

However, what happens if you want to prompt the user for the expression dataset file? If you mark the ComparativeMarkerSelection.input.file parameter as a prompt-when-run parameter, you still need an input file for ExtractComparativeMarkerResults.dataset.filename parameter. If you mark both parameters as prompt-when-run, you have to rely on your user to submit the same expression dataset file for both parameters. The workaround is to add the ConvertLineEndings module to your pipeline:

1. Add ConvertLineEndings to your pipeline.
2. Mark the ConvertLineEndings.input.filename parameter as prompt-when-run. ConvertLineEndings generates an output file almost identical to the input file; it simply converts the line endings in the file to those used by perl on the host operating system.
3. Use the output file generated by ConvertLineEndings as the input file for both the ComparativeMarkerSelection.input.file parameter and the ExtractComparativeMarkerResults.dataset.filename parameter.

Back to top

Reordering Modules

GenePattern orders modules in a pipeline based on their position in the diagram. The diagram is read from left to right and top to bottom, as you would read a book written in English.

You can reorder the modules in a pipeline by repositioning the modules in the diagram. Similarly, you can insert a module into a pipeline simply by adding it to the diagram and dragging it to the appropriate location. However, in either case, you are changing the flow of the data and, therefore, must delete and recreate any affected connections.

To reposition a module in the pipeline:

1. Remove its connections to other modules.
2. Drag the module to its new position.
3. Recreate connections as needed to reflect the flow of data through the modified pipeline.

Back to top

Working with Suites

GenePattern includes close to 200 analysis and visualization modules, and the list is constantly growing. With so many modules, it's helpful to have shortcuts for finding your favorites. Suites can help you organize modules in much the same way playlists help you organize music. Each suite is a themed collection of user-selected modules. You can create as many suites as you like. The GenePattern team has created several suites, which are included on the GenePattern public server. For example, the **ClusteringSuite** contains the HierachicalClustering, KMeansClustering, and SOMClustering modules. If you are unfamiliar with GenePattern suites, see Concepts.

- 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: 1. Click *Suites>Manage*. 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. • To export a suite to a zip file, click *Suites>Manage*, click the icon next to a suite to display the suite menu, and select *Export Including dependents* or *Export excluding dependents.* • To install a suite from a zip file: click *Suites>Install from zip*. For more information, see Exporting and Installing Suites Using Zip Files.
Install suites from the repository	The Broad Institute maintains a repository of modules, pipelines, and suites. To install suites from the Broad repository, click *Suites>Install from Repository*. For more information, see Installing Suites from the Repository.
Create suites	To create a suite, click *Suites>New*. 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 *Suites>Manage*. For more information, see Managing Suites.

Displaying Suite Definitions

To display the definition of a suite:

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

suite definition

- *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 of this GenePattern server.
- *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 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.

suite definition

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 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 Concepts.

- 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. Most of these modules and pipelines are available on the GenePattern public server. To install these modules and pipelines on your local 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 (see Managing Modules & Pipelines). 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*.
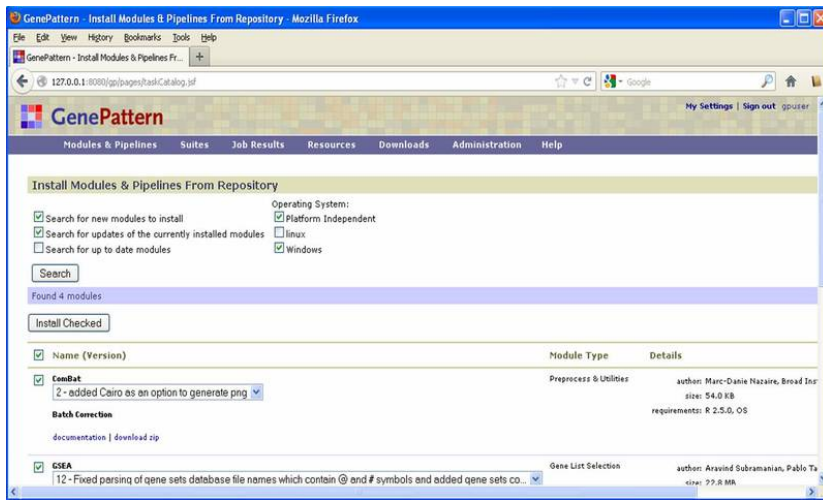
Use the top section of the form to find the modules to install. To update the list of modules/pipelines, select the modules/pipelines to search for and click *Search*:

- *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:
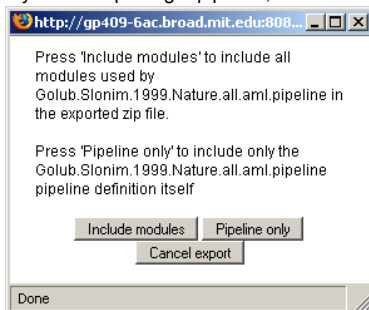
- Click *Include modules* to create a zip file that includes all modules in the pipeline.
- Click *Pipeline only* to create a zip file that contains only the pipeline definition. (If you install a pipeline without installing the modules that it uses, when you run the pipeline, GenePattern lists the missing 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:
   - To select a zip file from a directory, click *Browse*.
   - 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:
   - Your user name: Choose this option for a private installation. Only you (or an administrator) can view and run the module/pipeline.
   - 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*.
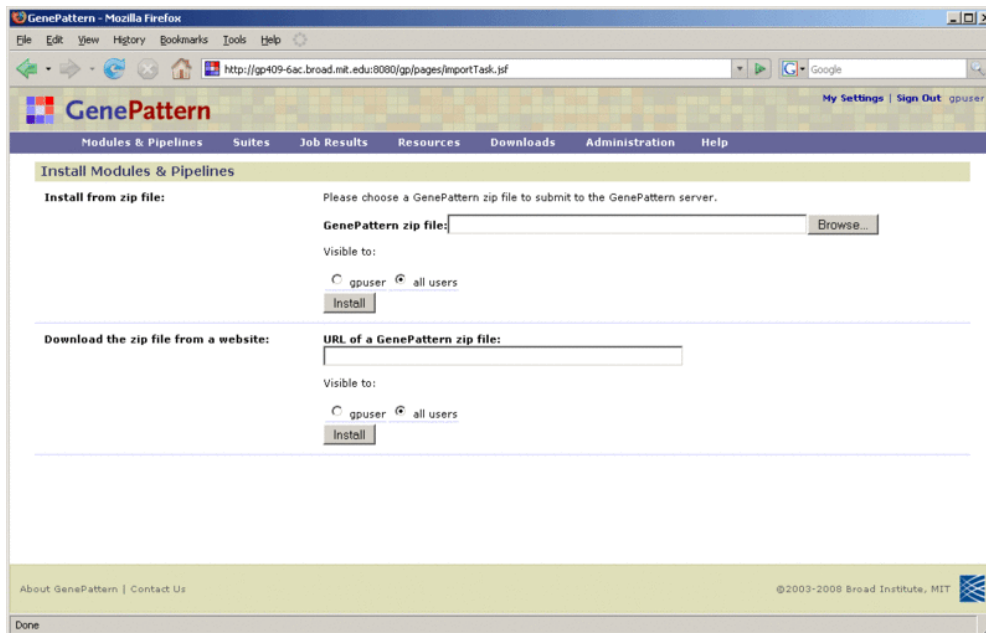
Managing Modules & Pipelines

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.
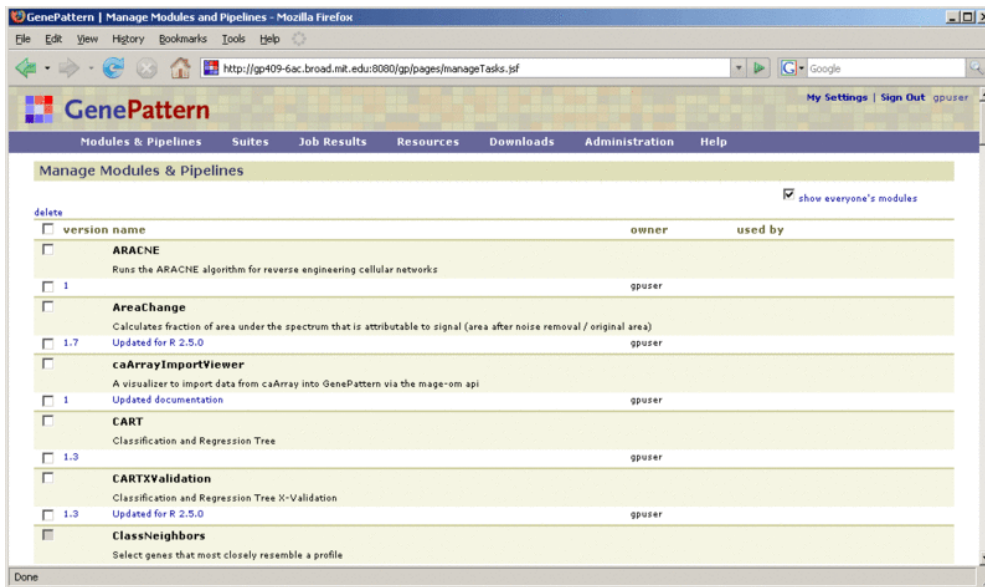
Installing Suites from the Repository

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:
   - Your user name: Choose this option for a private installation. Only you (or an administrator) can view and run the module/pipeline.
   - All users (default): Choose this option for public installation. Any user connected to this server can view or run the module/pipeline.
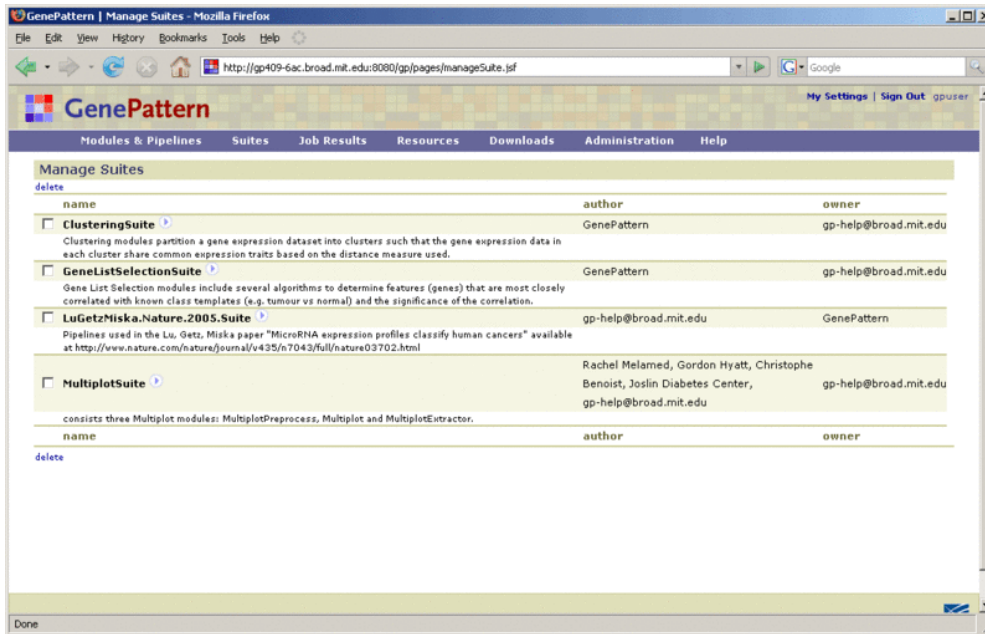4. Click *Install*.
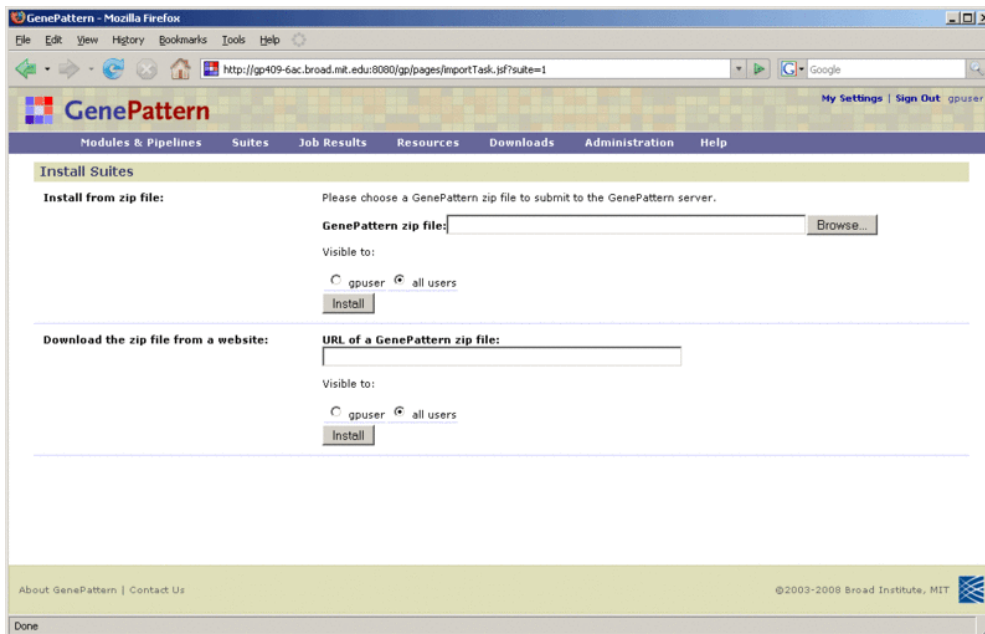
Managing Suites

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.