GenePattern
Web Client Guide

Software Copyright

The Broad Institute
SOFTWARE COPYRIGHT NOTICE AGREEMENT

Table of Contents

Getting Started

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

GenePattern has three components:

● The server, which is the GenePattern engine.

● The Java Client, which is the primary GenePattern interface for most users. The GenePattern Java Client Guide describes the Java Client and how to use it.

● The Web Client (also called the GenePattern home page or GenePattern server home page), which is primarily used for administrative operations, such as adding and deleting analysis modules. This guide describes the Web Client and how to use it.

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

● Installing GenePattern

● Starting the GenePattern Server

● Starting the GenePattern Web Client

● Exiting from GenePattern

● Terminology

● Getting Help

Installing GenePattern

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

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

Starting the GenePattern Server

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

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

StartGenePatt
ernServer

On Windows, while the server is starting, a console window displays messages. The server is ready when the console displays a message similar to the following (version and build numbers depend on the version of the software that you have installed):

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

Starting the GenePattern Web Client

You must start the GenePattern server before you can start the Web Client.

To start the GenePattern Web Client:

1. Double-click the GenePattern Web Client icon, shown below. By default, the GenePattern installation places this icon on your desktop.

GenePatternH
ome.html

The GenePattern login page appears:
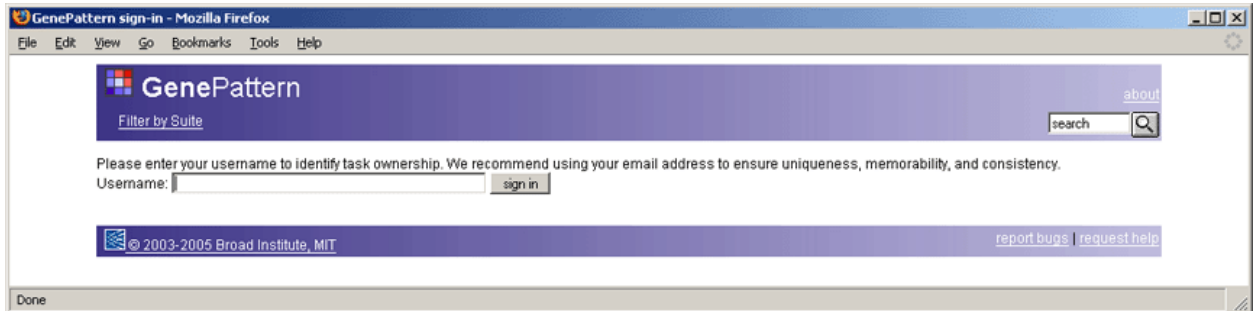
2. Enter your user name, usually your e-mail address, and click *sign in*. The Web Client home page appears, as shown below:
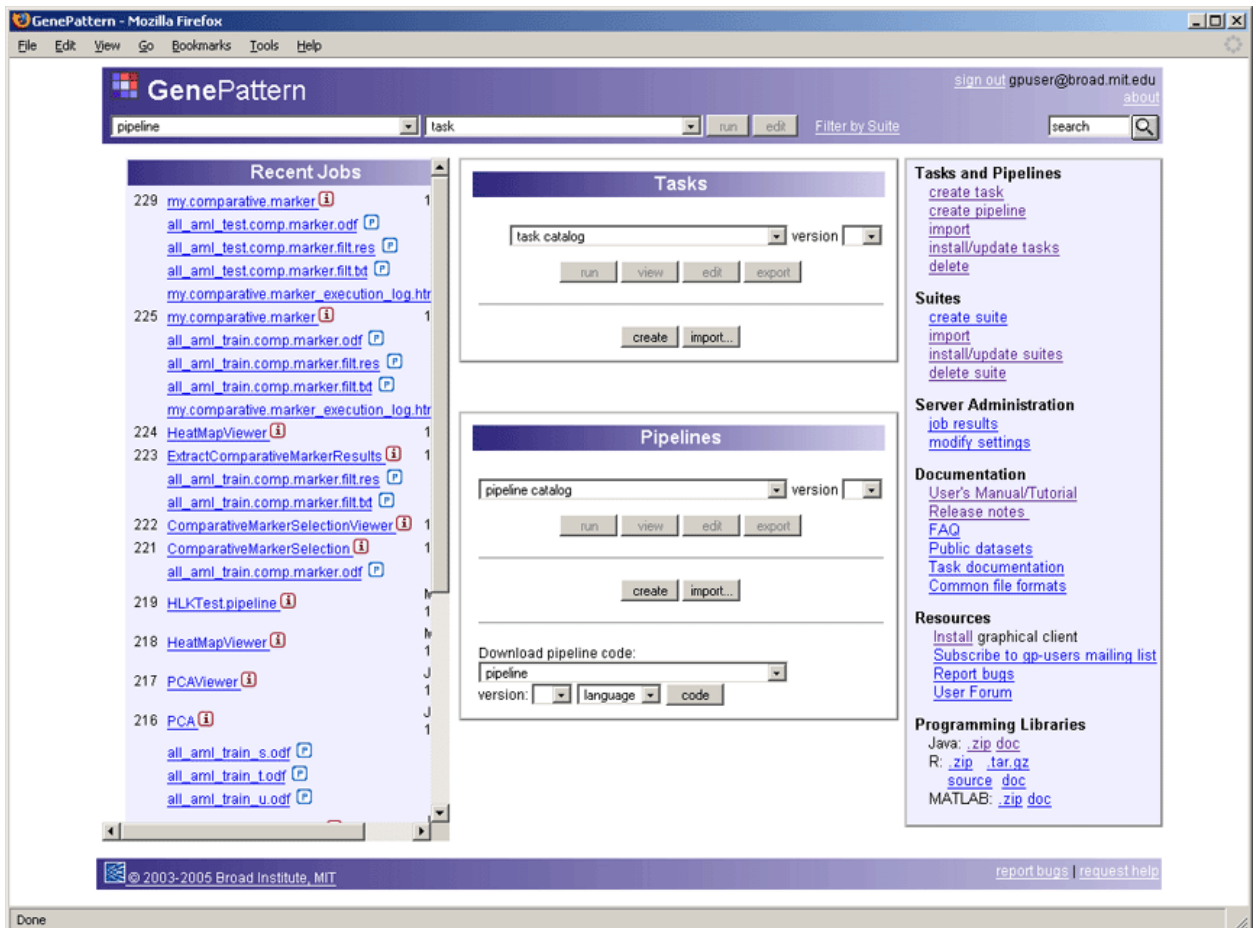
Exiting from GenePattern

To exit from GenePattern:

1. Exit from the Java and Web clients.

- To exit from the GenePattern Java Client, if it is running, select *File>Quit*.

- ▪ To exit from the GenePattern Web Client, click *sign out* in the top right corner of the title bar.

2. Shutdown the server by double-clicking the Stop GenePattern Server icon, shown below. Alternative: if you are using Windows, you can close the console window.

StopGenePatt
ernServer

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

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

Terminology

This section defines a few terms used throughout the Web Client interface and documentation:

- **Task** refers to an analysis or visualization module. If you are unfamiliar with GenePattern analysis and visualization modules, see About Analysis and Visualization Modules in the *GenePattern Java Client Guide*.

- **Pipelines** define sets of tasks to be run. Tasks run from a pipeline work exactly the same as those run directly from the Web Client. If you are unfamiliar with GenePattern pipelines, see About Pipelines in the *GenePattern Java Client Guide*.

- **Suites** group tasks and pipelines with similar functionality into packages; for example, the ClusteringSuite contains the ConsensusClustering, HierarchicalClustering, HierarchicalClustering.MATLAB, NMFClustering, and SOMClustering tasks. You use suites to work with groups of tasks and pipelines; for example, you can install all of the tasks and pipelines in a suite or filter the Web Client menus to show only the tasks and pipelines in a suite.

- **Module** refers to an analysis module, visualization module, pipeline, or suite.

Getting Help

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

- The GenePattern Tutorial provides a brief tour of common operations.

- The GenePattern Java Client Guide describes the Java Client and how to use it.

- This guide, the GenePattern Web Client Guide, describes the Web Client and how to use it.

- The GenePattern Programmer's Guide describes how to access GenePattern from various programming environments and provides instructions for writing GenePattern modules.

- The Modules page lists all GenePattern modules, with links to their documentation. Each analysis, visualization, and pipeline module has its own documentation.

- The File Formats page describes the GenePattern file formats.

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

Web Client Overview

Most GenePattern operations are available from both the Java Client and the Web Client. A few administrative operations are available only from the Web Client. The following figure shows the Web Client home page with these administrative operations circled.

The Java Client offers a friendly, easy-to-use interface; therefore, most GenePattern users prefer to use the Java Client for common operations, such as running analyses and reviewing analysis results. The Web Client tends to be used primarily for administrative operations.
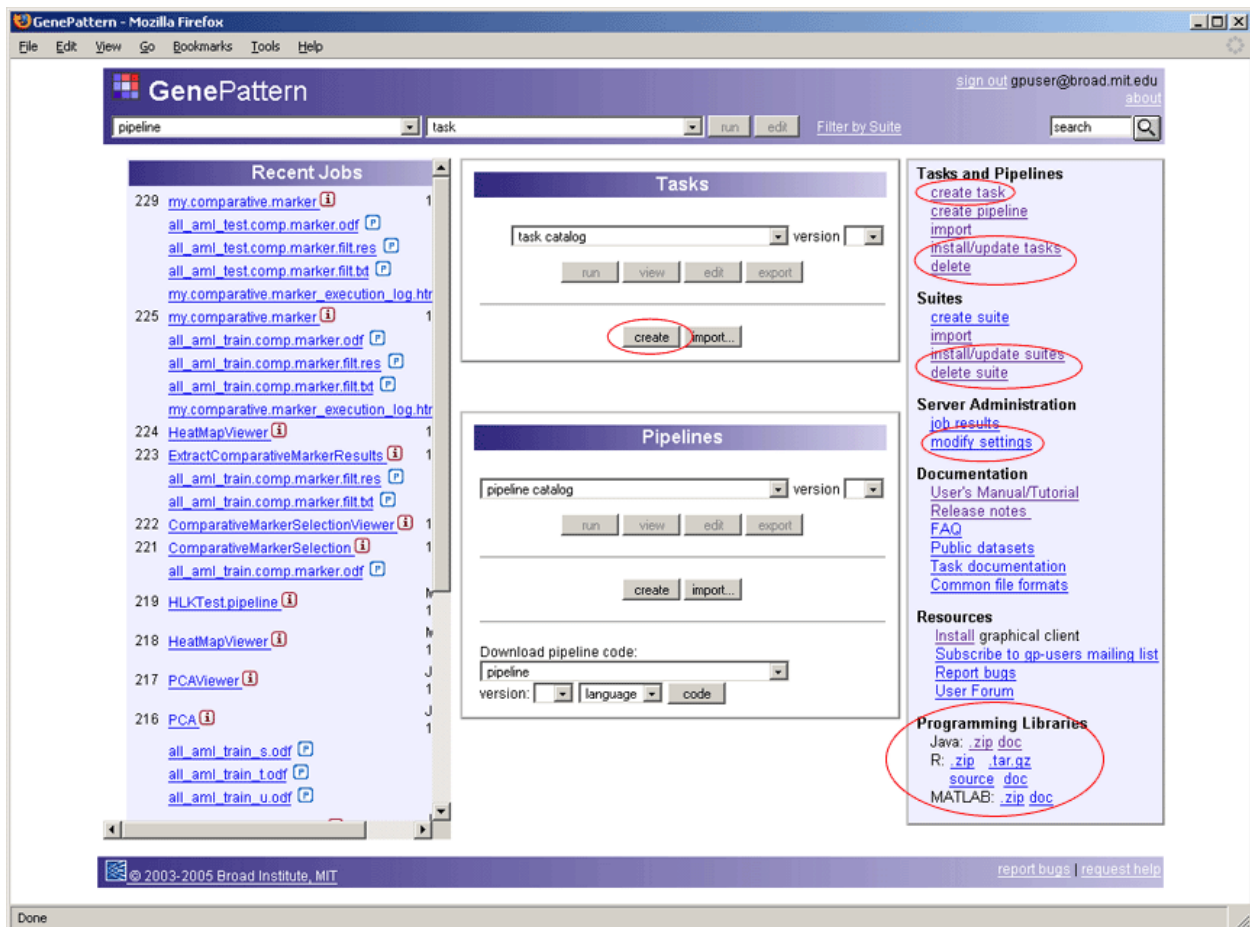
Following are brief descriptions of the key elements on the Web Client home page:

● In the title bar:

  ▪ Click the GenePattern icon at any time to return to the Web Client home page.

  ▪ Use the drop-down lists to run tasks and pipelines. For more information, see Running Tasks and Pipelines.

  ▪ Use *Filter by Suite* to have the Web Client drop-down lists show only a subset of the available tasks and pipelines. For more information, see Managing Suites.

● The Recent Jobs column lists the jobs recently run on the GenePattern server. Both the Java Client and the Web Client show jobs that you have run. The Web Client shows all jobs that have been run; jobs that you have run and those run by other GenePattern users. For more information, see Managing Jobs.

● The Tasks pane in the middle column and the Tasks and Pipelines links in the right column provide full access to the analysis and visualization modules on the GenePattern server. Both the Java Client and the Web Client allow you to run, view, edit, import, and export analysis and visualization modules. The Web Client also allows you to create, install, and delete analysis and visualization modules. For more information, see Managing Tasks.

● The Pipelines pane in the middle column and the Tasks and Pipelines links in the right column provide full access to the pipeline modules on the GenePattern server. Both the Java Client and the Web Client allow you to run, view, edit, import, export, and create pipelines. The Web Client also allows you to install and delete pipelines. For more information, see Managing Pipelines.

● The Suites links in the right column provide access to the suites on the GenePattern server. Both the Java Client and the Web Client allow you to view, edit, create, import, and export suites, and to apply filters based on those suites. The Web Client also allows you to install and delete suites. For more information, see Managing Suites.

● The Server Administration links provide access to two administrative pages:

  ▪ The *job results* link displays the Job Results page, which allows you to save and delete analysis results files. Both the Java Client and the Web Client allow you to save and delete analysis results files for jobs that you have run. The Web Client allows you to save analysis results files for any job that has been run; jobs that you have run or those run by other GenePattern users. For more information, see Managing Jobs.

8

- ▪ The *modify settings* link displays the Server Administration page, which allows you to modify the GenePattern server configuration. Only the Web Client allows you to modify server configuration settings. For more information, see Modifying Server Configuration Options.

- ● The Documentation links provide quick access to GenePattern documentation.

- ● The Resources links provide quick access to support resources for GenePattern. In addition, note the two links in the lower right corner of the window, which you can use to request help or report bugs.

- ● The Programming Libraries links provide access to the software and documentation required to access GenePattern from your programming environment. For more information, see the GenePattern Programmer's Guide.

Running Tasks and Pipelines

The Web Client generally uses the term **task** to refer to analysis modules and visualization modules. If you are unfamiliar with GenePattern analysis and visualization modules, see About Analysis and Visualization Modules in the *GenePattern Java Client Guide*.

GenePattern pipelines define sets of tasks to be run. Tasks run from a pipeline work exactly the same as those run directly from the Web Client. If you are unfamiliar with GenePattern pipelines, see About Pipelines in the *GenePattern Java Client Guide*.

To run a task/pipeline:

1. Display the Web Client home page.

2. Select the task/pipeline from the Task/Pipeline drop-down list in the title bar and click the *Run* button in the title bar. Alternatively, select the task/pipeline from the drop-down list in the Task/Pipelines pane and click the *Run* button in that pane.

   Most tasks have parameters; most pipelines do not. If necessary, the Web Client prompts you for the parameter values. The following figure shows the parameters for the Comparative Marker Selection task.
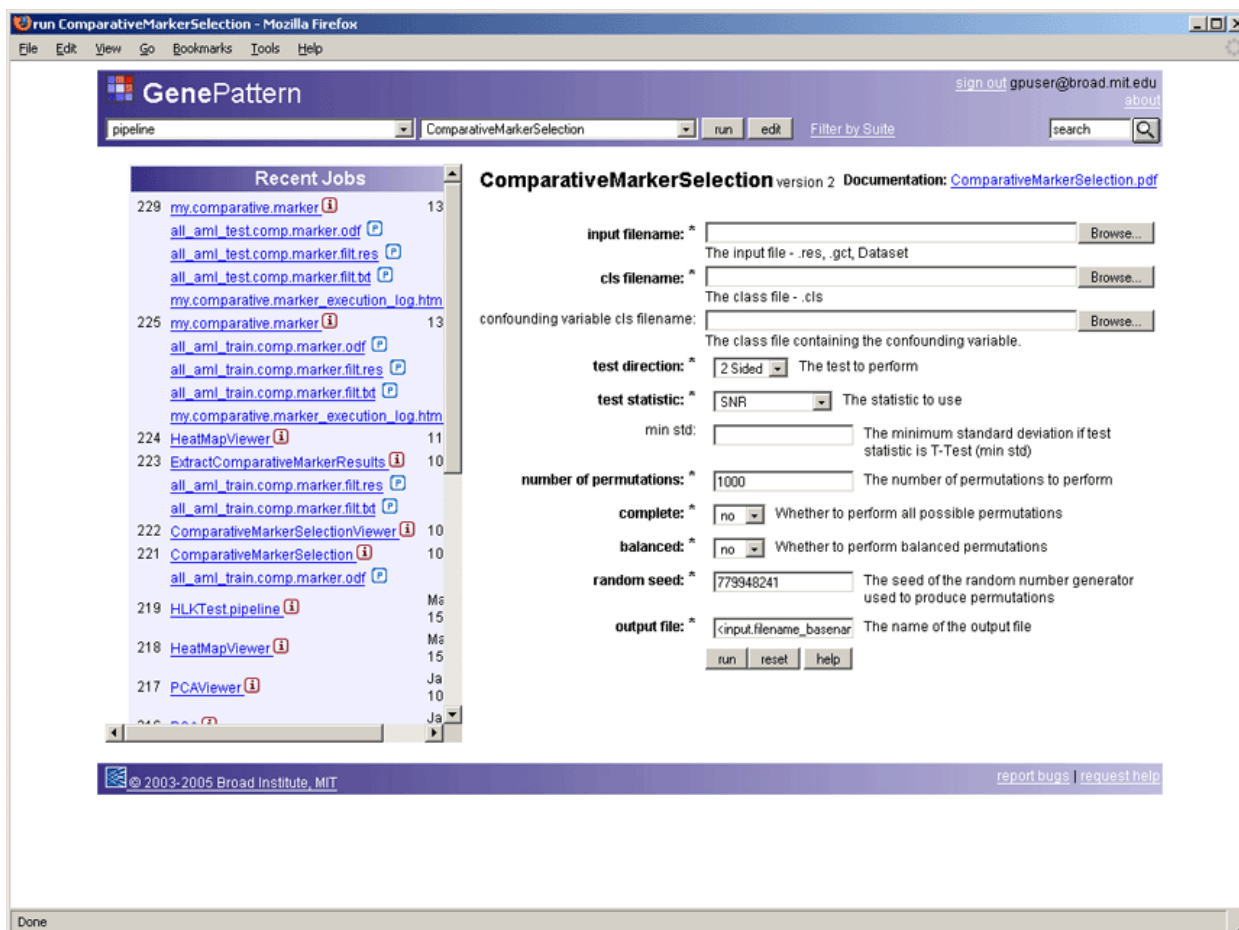
3. Enter values for each parameter.

The form lists each parameter followed by an entry box for the parameter value. Required parameters are indicated by an asterisk (*) and listed in **bold**; other parameters are optional. When a parameter requires a file name, you can enter it in many ways:

▪ Click *Browse* (next to the file parameter box) to display the File Upload window. Navigate to the desired file, select it, and click *Open*.

▪ Right-click a file listed in the Recent Jobs pane and select *Copy Link Location*. The web browser copies the file name to the clipboard. Paste the file name into the file parameter box.

▪ Enter the full path name of a file in the file parameter box. This method is not recommended because file names can too easily be mistyped.

4. If necessary, click *Reset* to restore the parameters to their default values or *Help* to display documentation for the task/pipeline.

5. Click *Run* to start the job. The Web Client sends the job to the GenePattern server for processing and displays the execution log page. When the task completes, or as each task in the pipeline completes, the Web Client adds the analysis results files to the execution log page. The following figures show the execution log page for a pipeline (the first figure shows the top of the window and the second shows the window scrolled to the bottom):
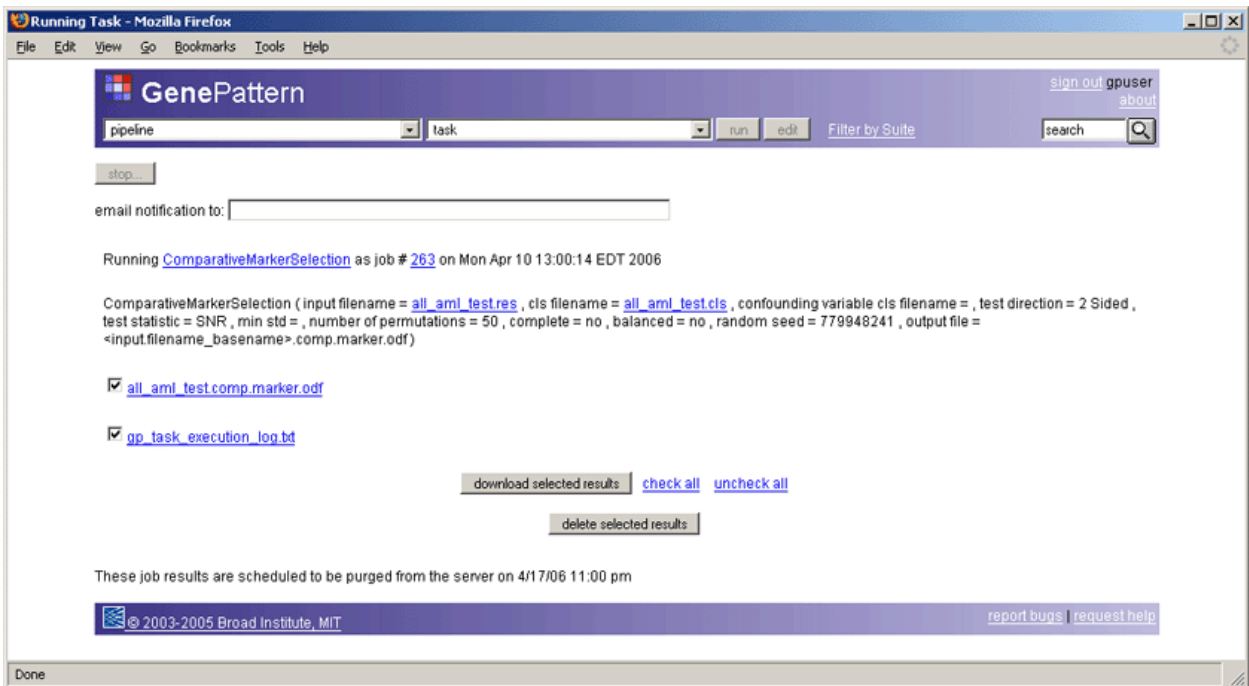
All analysis results files are stored on the GenePattern server. Server storage is temporary and results files are deleted after they have been on the server for a certain length of time (by default, one week). A note at the end of the log page indicates when the files will be deleted.

From the log page you can:

▪ Click *Stop* at the top of the page to cancel the current job.

▪ Click an analysis results file to display the file in a text viewer.

▪ Click *Download selected results* to save the selected files to a permanent location.

▪ Click *Delete selected results* to delete the selected files from the GenePattern server.

6. When you are finished, click the GenePattern icon in the title bar to display the Web Client home page. The Recent Jobs list includes the task/pipeline job that you just ran.

You can use the log page to display an analysis results file in a text viewer. Alternatively, you can display most analysis results files using the visualization modules (tasks) designed for that purpose. For example, you can use the Comparative Marker Selection Viewer task to display the analysis results file from the Comparative Marker Selection task. To display an analysis results file using a visualization module:

1. Run the visualization module (task), as described above.

2. When prompted for an input file, select the analysis results file as the input file for the visualization module.

Managing Jobs

Each time you run a task or pipeline from a GenePattern client, the client starts a job on the GenePattern server to execute that task or pipeline. For more information about server jobs, see About Analysis and Visualization Modules in the *GenePattern Java Client Guide*.

You can view and manage jobs from the Web Client home page or the Job Results page. This section describes both pages:

- Web Client Home Page
- Job Results Page

Web Client Home Page

To display the Web Client home page, start the Web Client or, if the Web Client is running, click the GenePattern icon in the title bar. The Recent Jobs column lists the jobs that you have recently run on the server and the result files from those jobs:
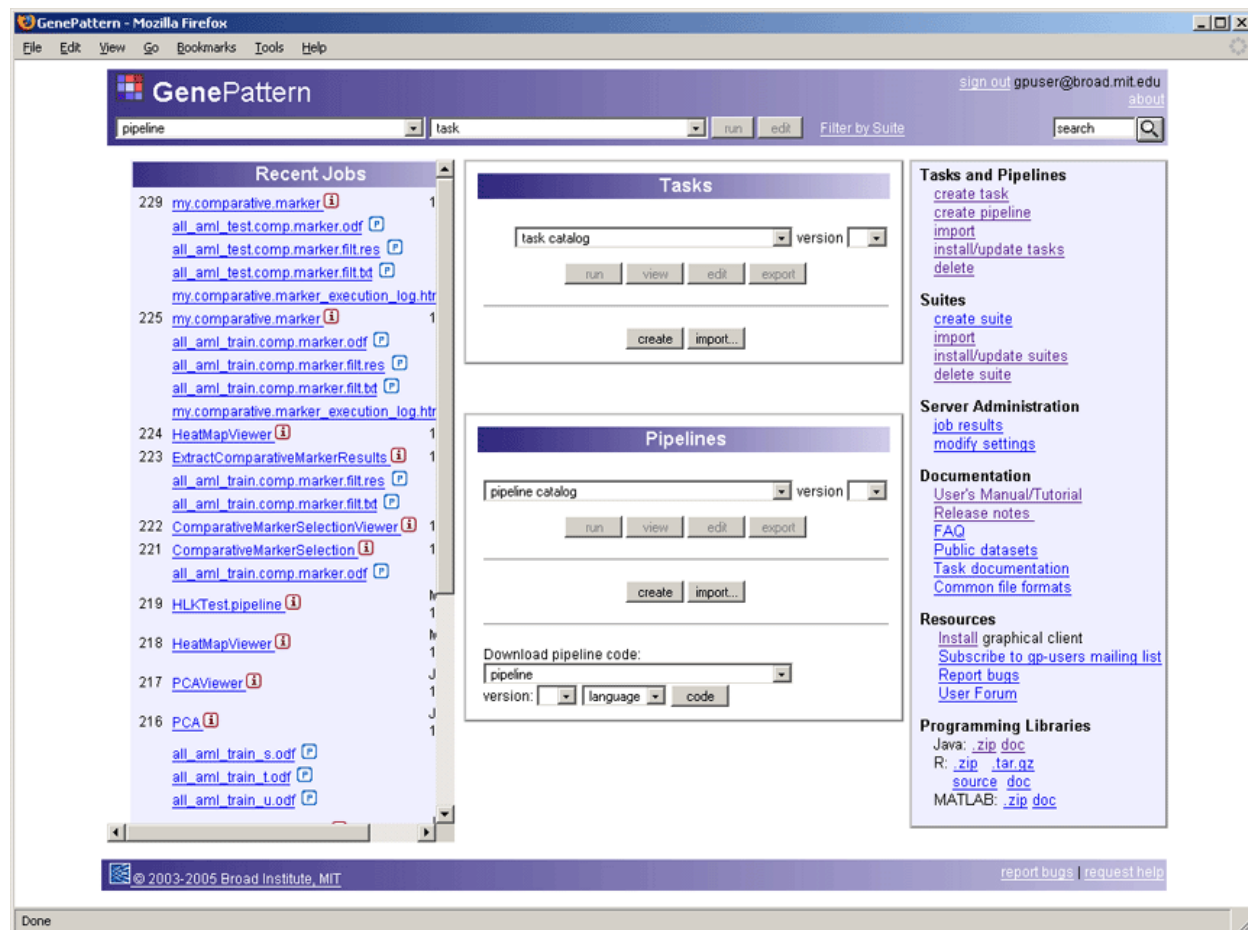
To manage jobs from this page, you can:

- Click a job name, or the information (i) next to the job name, to display the time the job completed, its output files, and its input parameters.
- Click a file name to display the file in a text viewer. You can use the viewer to save a file or to cut and copy text; you cannot use the viewer to edit the file.
- Click the pipeline (p) icon next to a file name to create a pipeline that contains the tasks used to generate that results file. For more information, see Creating Pipelines.
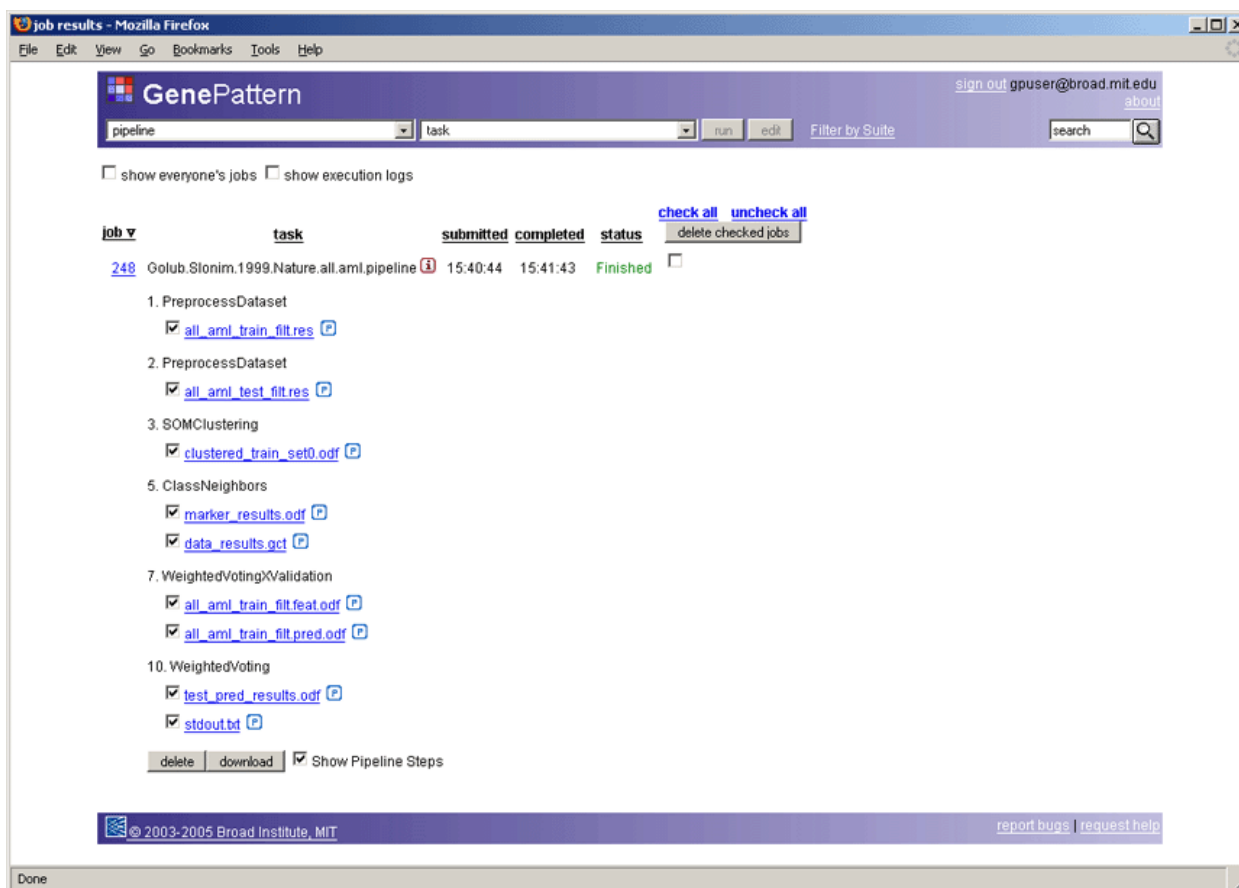
Job Results Page

To display the Job Results page:

1.   Display the Web Client home page.

2. Under Server Administration in the right column, click *job results.*

The Job Results page shows a detailed list of the jobs you have run on the server and the result files from those jobs:

From this page, you can:

● **Show your jobs or all jobs.** By default, the Job Results page shows jobs that you have run. Click the *show everyone's jobs* check box to display all jobs that have been run on the server. The Web Client displays the owner's name next to each results file.

● **Show execution logs.** By default, the Job Results page shows the results files for each job. Click the *show execution logs* check box to display the results files and execution log file for each job.

● **Display job details.** Click the information (i) icon to the right of the job name to display the time the job completed, its output files, and its input parameters.

● **View results files.** Click a file name to display the file in a text viewer. You can use the viewer to save a file or to cut and copy text; you cannot use the viewer to edit the file.

● **Save results files.** Analysis results files are stored on the GenePattern server for a limited amount of time (by default, one week). To save analysis results files, download them from the server to permanent storage. To download individual result files for a job, check the boxes next to those files and then click the d*ownload* button at the end of the job. To download all files for a job, click the job id link (located in the job column). In either case, the Web Client saves the files to a zip file whose name is the id of that job and prompts you to save the zip file.

● **Delete results files.** To delete individual files from a job, check the boxes next to those files and then click the *delete* button at the end of the job. To delete a job and all of its results files, check the box to the right of the job name and click *delete checked jobs* at the top of the window.

● **Show pipeline steps.** By default, for pipeline jobs, the Job Results page displays the pipeline tasks and the results files for each task. Clear the *Show Pipeline Steps* check box at the end of the pipeline job to show just the results files.

Managing Tasks

The Web Client generally uses the term **task** to refer to analysis modules and visualization modules. If you are unfamiliar with GenePattern analysis and visualization modules, see About Analysis and Visualization Modules in the *GenePattern Java Client Guide*.

This section describes how to manage tasks:

● Installing and Updating Tasks

● Exporting and Importing Tasks

● Viewing Tasks

● Creating Tasks

● Editing Tasks

● Deleting Tasks

You can export, import, edit, and view task code from either the Web Client or the Java Client. You can install/update, create, and delete tasks only from the Web Client.

Installing and Updating Tasks

The Broad Institute maintains a repository of tasks and pipelines. When you install tasks, you search the repository for tasks that you do not have on your GenePattern server and, optionally, add them to your server. When you update tasks, you search the repository for newer versions of tasks that you have installed on your GenePattern server and, optionally, add them to your server.

To install and update tasks:

1. Display the Web Client home page.

2. Under Tasks and Pipelines in the right column, click *install/update tasks*.

   The Web Client retrieves the list of tasks from the repository, compares that list to the tasks on your GenePattern server, and displays the install/update page. The top portion of the page shows the filter criteria used to search the repository; the bottom portion of the page lists new and updated tasks found in the repository.
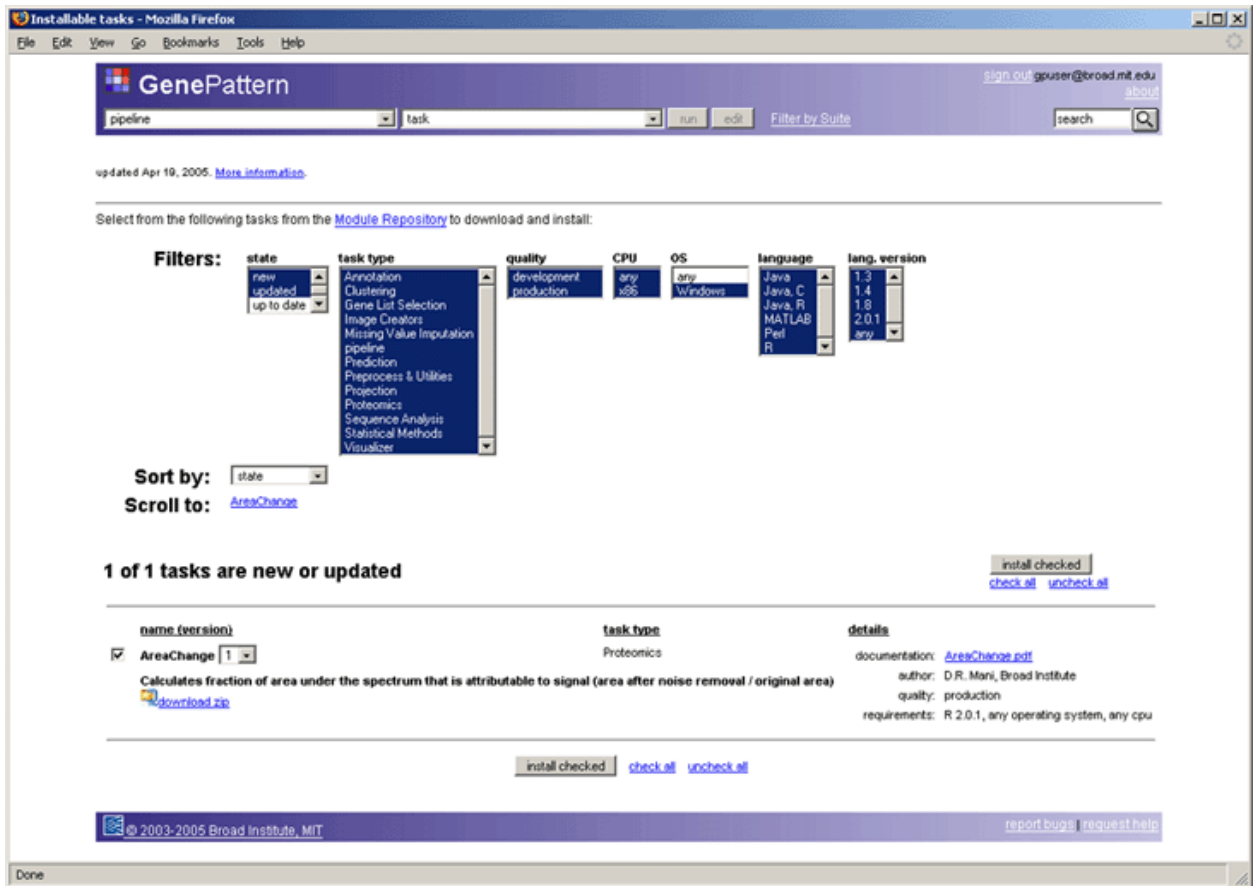
3. Click *Install Checked* to add all new and updated modules to your GenePattern server. The Web Client displays a log page listing the modules as they are installed.

4. Click the GenePattern icon in the title bar to display the Web Client home page.

Optionally, you can change the filtering options and select the modules that you want to install. Generally, however, you want to install all new and updated modules.

Exporting and Importing Tasks

GenePattern allows you to share tasks with other GenePattern users by exporting and importing the tasks. Exporting a task copies the task from your GenePattern server to a zip file; importing a task copies the task from a zip file to your GenePattern server.
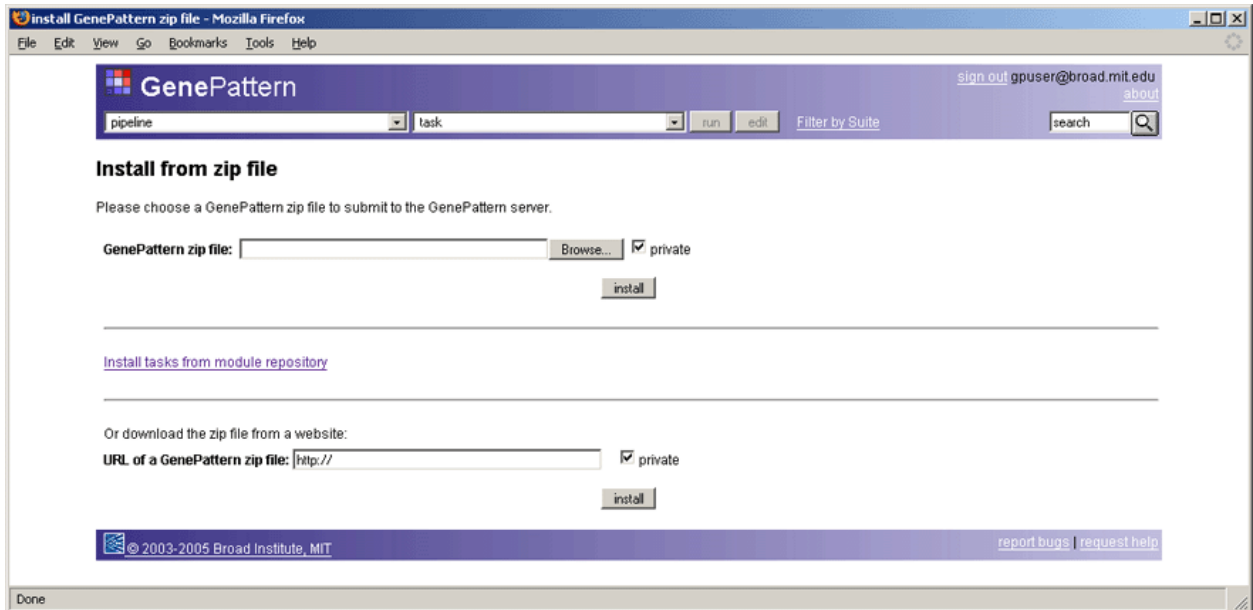
To export a task:

1. Display the Web Client home page.

2. In the Tasks pane, select the task that you want to export and click *Export*. The Web Client creates a zip file and displays a window similar to the following (the exact window that appears depends on your operating system):
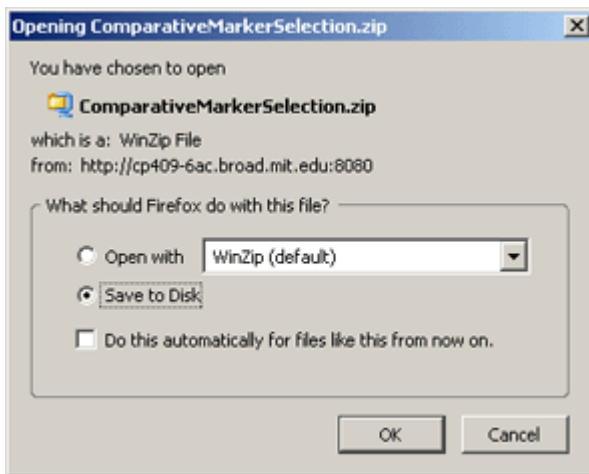
3. Open the zip file and then save it; or save the zip file to disk.

To import a task:

1. Display the Web Client home page.

2. Click the *Import* button in the Tasks pane or, under Tasks and Pipelines in the right column, click *import*. The Web Client displays the Import page:


3. Select the zip file and click *install*. Use the top portion of the page to browse for the zip file or the bottom portion of the page to enter a URL for the zip file. The Web Client imports the task and then runs it, which displays the task and its parameters.

4. Enter parameters and run the task, as described in Running Tasks and Pipelines, or click the GenePattern icon in the title bar to display the Web Client home page.

Viewing Tasks

Viewing a task shows you its definition, including its author, the command line used to invoke it, and the programs used to execute it.

To view a task:

1. Display the Web Client home page.

2. In the Tasks pane, select the task that you want to view and click *View*. The Web Client displays the task definition:
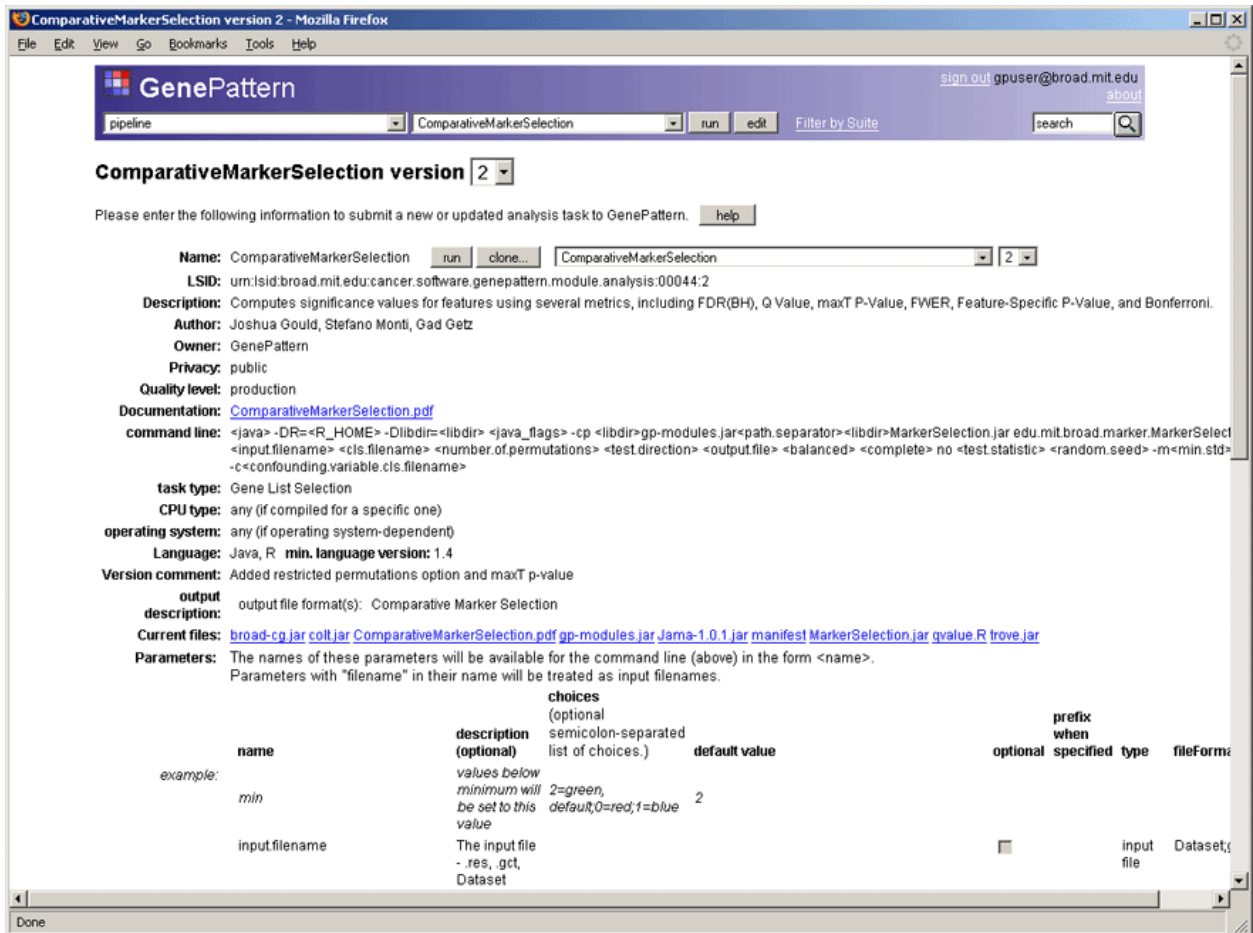
From this window, you can:

- Click *Help* for a description of the task definition form.
- Click *Run* to run the task.
- Click *Clone* to create a new task identical to this one.
- In the *Current files* field, click the program files to display the code used to execute the task.

Creating Tasks

Creating a GenePattern task is a two-step process:

1. Write a program that executes the desired function. You can write the program in the language of your choice; for example, you can use a compiled language, such as C, to create an executable or use a scripting language, such as Perl, to create a script that is run by an interpreter. The GenePattern Programmer's Guide provides guidelines for writing programs that will be run as GenePattern tasks.

2. Use the GenePattern Web Client to create a GenePattern task that invokes the program that you have written. This section describes how to create a GenePattern task and then provides a brief tutorial that walks you through the process.

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

1. Display the Web Client home page.

2. Open the task definition form in one of the following ways:

   ▪ Click *Create* in the Tasks pane.

   ▪ Click *create task* under Tasks and Pipelines in the right column.

   ▪ View or edit a task and use the *Clone* button to create an identical task.

   The Web Client displays the task definition form:

16

3. Click *Help* for instructions on how to complete the task definition form.

4. Complete the task definition form.

5. Click *Save* to create the task.

Note: By adding a task, a user can execute arbitrary code on the GenePattern server. Because arbitrary code may include malicious code, consider taking precautions to protect your server: for example, employing virus scanner software and restricting access to appropriately privileged (non-root) users. For more information about securing your server, see Securing the Server.

Tutorial: Creating a Task

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

To create the `log_transform` task:

1. Display the Web Client home page.

2. In the Tasks pane, click *Create*. The Web Client displays the task definition form. (The task definition form, with the tutorial data filled in, appears at the end of the procedure.)

3. Enter the following information in the first few fields:

   ▪ **Name**: LogTransform

   ▪ **Description**: Log transform a data (gct) file.

   ▪ **Author**: Your name and affiliation.

   ▪ **Privacy**: Select private. This means that only you (users logged in using the user name in the *Owner* field) can view and run the task. Public allows all users connected to this server to view and run the task.

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

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

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

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

▪ <perl> represents the full path to the Perl installation used by GenePattern.

  Similar variables are available for <java> and <R>. If your program was run by another interpreter, such a Python or LISP, you could use an absolute pathname (for example, /usr/bin/python) or create a new GenePattern variable for the interpreter pathname (for example, python=/usr/bin/python) in the `GenePatternServer/resources/genepattern.properties` file. If your program was an executable file, you would not need an interpreter, so would not need this variable.

▪ <libdir> represents the full path to the directory that contains the files for this task, including the program file.

▪ The Perl script, `log_transform.pl`, expects two parameters, an input file and an output file name: `-F <input.filename> -o <output.file>`. When your program has parameters, you include them in the command line and also define them in the *Parameters* field, which appears below.

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

For example, if the LogTransform task read its input from the standard input stream and wrote its output to the standard output stream, the command line would be:

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

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

▪ **task type**: Preprocess & Utilities

▪ **CPU**: Any

▪ **Operating system**: Any

▪ **Language**: Perl

▪ **Version comment**: A brief note of what is special or interesting about this version.

▪ **output description**: Select gct from the list of formats.

  The *output description* field defines all file formats used by your module. To select multiple formats, use Shift+click (Option+click or Apple+click). Include both input and output file formats in this list.

6. Use the *Support files* field to upload your program, any other files needed to execute the task, and documentation for the task (a file with a text extension, such as .pdf, .doc, or .txt):

  1. Click the *Browse* button at the end of the first line in the *Support files* field. The Web Client displays the File Upload window.

  2. Navigate to the `gp_tutorial_files/log_transform` directory, select the file `log_transform.pl`, and click *Open*. This is the script that implements the task.

  3. Click the *Browse* button at the end of the second line in the *Support files* field. The Web Client displays the File Upload window.

  4. Select the file `LogTransform.pdf` and click *Open*. This is the documentation for the task. When a GenePattern user displays your task and clicks the *Help* button, GenePattern displays the support file that has a text extension (such as .pdf .doc, or .txt).

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

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

▪ **name**: input.filename

▪ **description**: The dataset to be transformed (gct format).

▪ **type**: choose input file

▪ **fileFormat**: choose gct

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

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

8. Click *save*. The Web Client displays a message informing you that the task has been saved.

9. Run the task to confirm that it has been added to the GenePattern server correctly.

Editing Tasks

To edit a task:
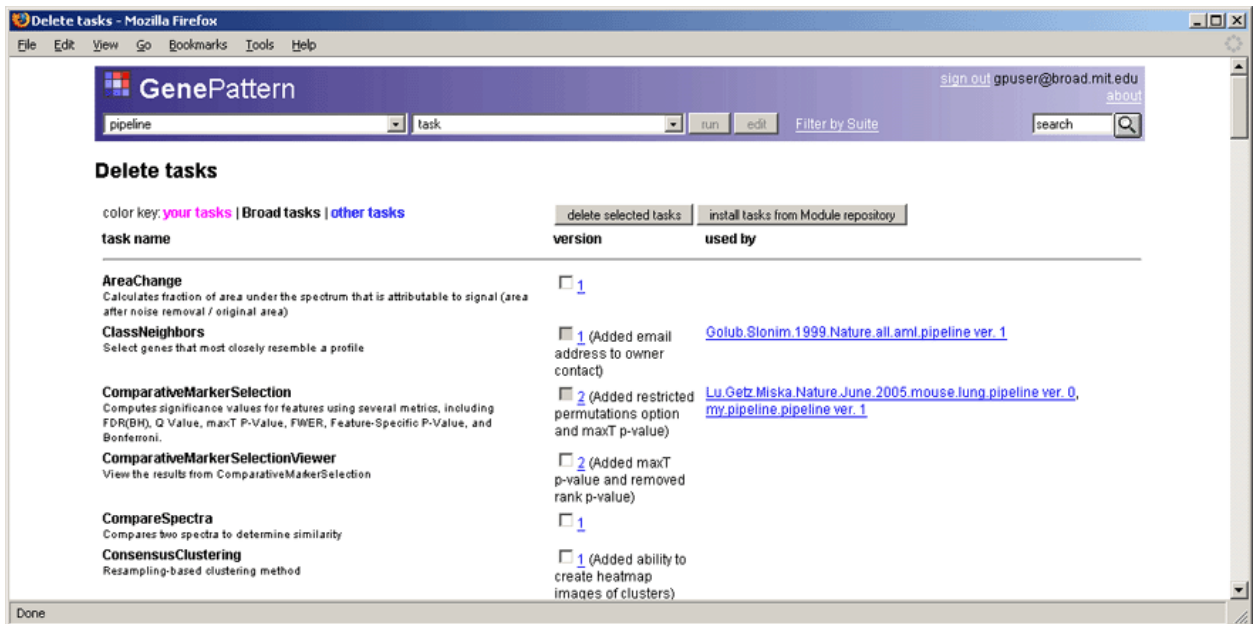
1. Display the Web Client home page.

2. Display a task definition form in one of the following ways:

   - In the Tasks pane, select a task that you created and click *Edit*.

   - In the Tasks pane, select a task created by another user and click *View*. From that window, you can then click *Clone*, which creates a copy of the task. You created the copy, so you can edit it.

3. Edit the task definition form. Click *Help* for descriptions of the fields and instructions.

4. Click *Save* to create a new version of the task.

Deleting Tasks

If a task is included in a pipeline, you must delete the pipeline (or remove the task from the pipeline), before you can delete the task. This prevents you from inadvertently breaking a pipeline by deleting any of its tasks.

To delete tasks from your GenePattern server:

1. Display the Web Client home page.

2. Under Tasks and Pipelines in the right column, click *delete*. The Web Client displays the Delete Tasks page, which lists the tasks and pipelines installed on your server. The color key at the top of the page defines the color-coding of the names, which indicates the source of the tasks and pipelines.

3. Select the check boxes next to each version of each module that you want to delete.

4. Click the *delete selected tasks* button at the top or bottom of the page. The Web Client displays a confirmation prompt:
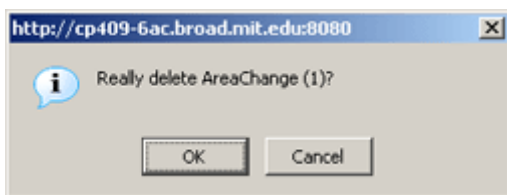
5. Click OK to delete the selected versions of the selected tasks and pipelines. The Web Client deletes the selected versions and redisplays the Delete Tasks page.

6. Click the GenePattern icon in the title bar to display the Web Client home page.

Alternative: You can also delete tasks from the task definition form: edit any task (see Editing Tasks) and click the *delete* button next to the *Task name* field.

Managing Pipelines

GenePattern pipelines define sets of tasks to be run. Tasks run from a pipeline work exactly the same as those run directly from the Web Client. If you are unfamiliar with GenePattern pipelines, see About Pipelines in the *GenePattern Java Client Guide*.

This section describes how to manage pipelines:

● Installing Pipelines

● Exporting and Importing Pipelines

● Viewing Pipelines

● Creating Pipelines

● Editing Pipelines

● Deleting Pipelines

You can export, import, view, create, and edit pipelines from either the Web Client or the Java Client. You can install/update and delete pipelines only from the Web Client.

Installing Pipelines

Installing a pipeline is the same as installing a task. For more information, see Installing and Updating Tasks.

Exporting and Importing Pipelines

You can share pipelines with other GenePattern users by exporting and importing the pipelines. Exporting a pipeline copies the pipeline from your GenePattern server to a zip file; importing a pipeline copies it from a zip file to your GenePattern server.

To export a pipeline:

1. Display the Web Client home page.

2. In the Pipelines pane, select the pipeline that you want to export and click *Export*. The Web Client displays the following prompt:
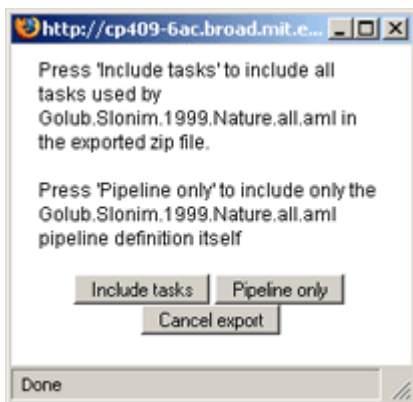
3. Choose whether to export the pipeline or the pipeline and all of its tasks:

   ▪ Click *Pipeline Only* to create a zip that contains a programmatic description of the pipeline and the pipeline documentation.

   ▪ Click *Include Tasks* to create a zip file that contains multiple zip files: the pipeline zip file (as described above) and a zip file for each task. The zip file for each task contains the program files that implement the task and the task documentation.

   The Web Client creates the zip file and displays a window similar to the following (the exact window that appears depends on your operating system):
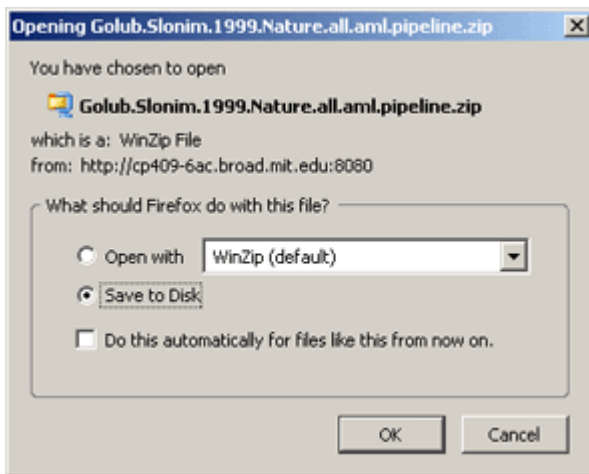
4. Open the zip file and then save it; or save the zip file to disk.

To import a pipeline:

1. Display the Web Client home page.

2. Click the *Import* button in the Pipelines pane or, under Tasks and Pipelines in the right column, click *import*. The Web Client displays the Import page:
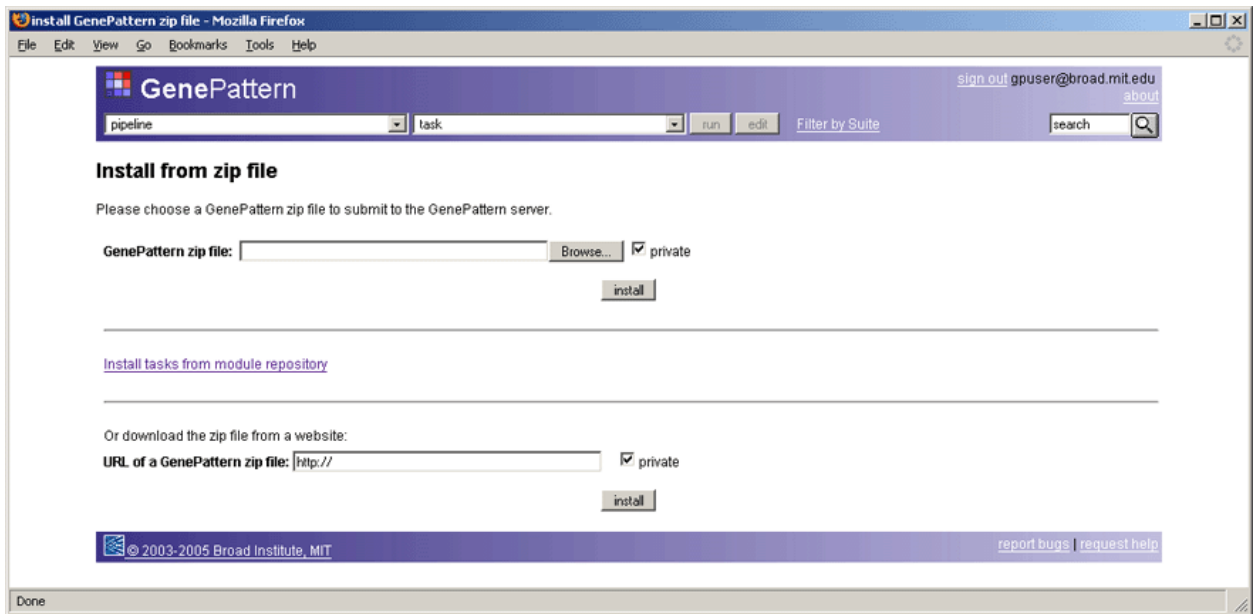
3. Select the zip file and click *install*. Use the top portion of the page to browse for the zip file or the bottom portion of the page to enter a URL for the zip file. The Web Client imports the pipeline and displays a confirmation message.

4. Click the GenePattern icon in the title bar to display the Web Client home page.

Viewing Pipelines

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

To view a pipeline:

1. Display the Web Client home page.

2. In the Pipelines pane, select the pipeline that you want to view and click *View*. The Web Client displays the tasks run by the pipeline:
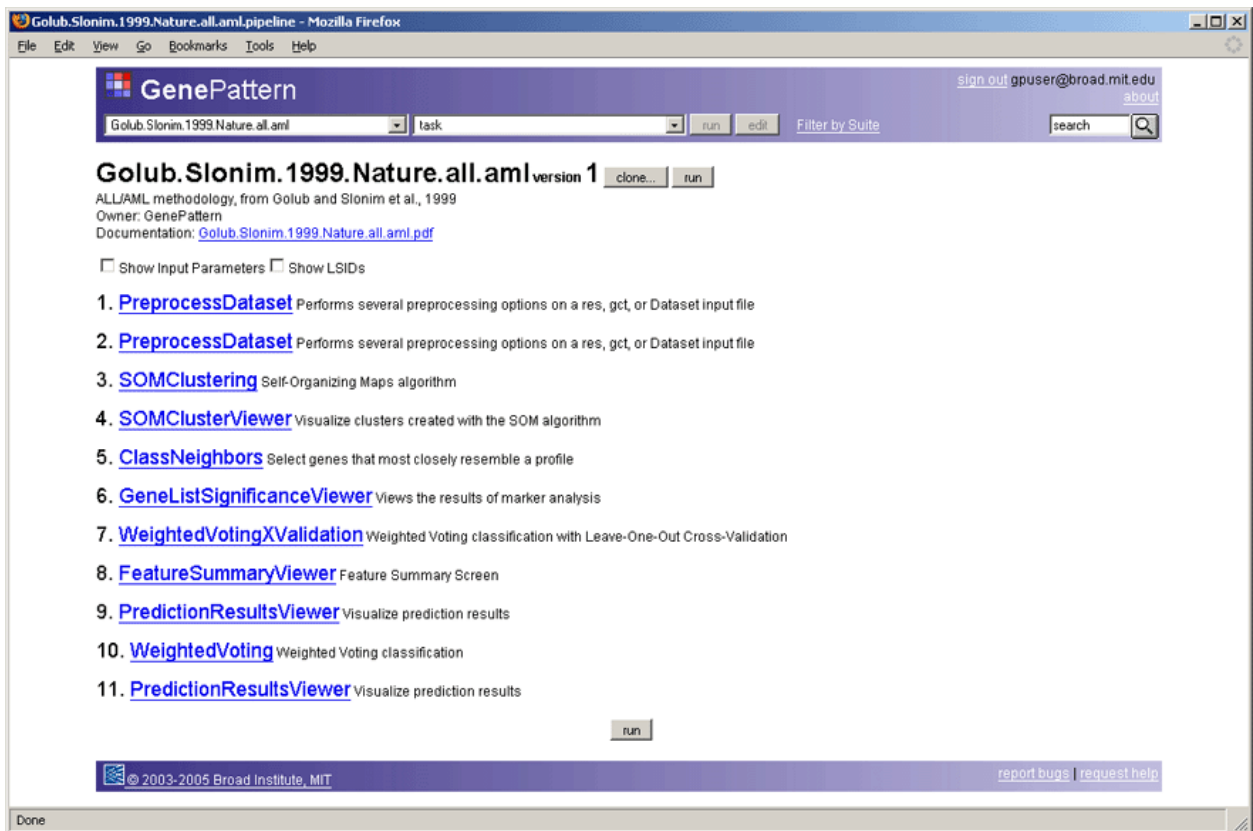
From this window, you can:

● Click the *Show Input Parameters* check box to show the input parameters for each task.

● Click the *Show LSIDs* check box to show the Life Science Identifiers (LSIDs) for the pipeline and each of its tasks. Each version of each GenePattern object has a unique LSID. For more information, see Understanding Version Numbers in the *GenePattern Java Client Guide*.

● Click a task name to view the task definition, which includes the program that executes the task and the command line used to invoke it.

● Click *Run* to run the pipeline.

● Click *Clone* to create a new pipeline identical to this one.

Note: To view the code used to implement a pipeline, use the *Download pipeline code* fields of the Pipelines pane on the Web Client home page.

Creating Pipelines

To create a pipeline:

1. Display the Web Client home page.

2. Display a pipeline definition form in one of the following ways:

   ▪ Click the *Create* button in the Pipelines pane or, under Tasks and Pipelines in the right column, click *create pipeline* to create an empty pipeline. The Web Client displays an empty pipeline definition form, as shown in the first figure below.

   ▪ Click the pipeline (P) icon next to a results file in the Recent Jobs column to create a pipeline that contains the tasks used to generate that results file. The Web Client prompts you for a pipeline name and then displays a populated pipeline definition form, as shown in the second figure below.

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

- ▪ View or edit a pipeline and click the *Clone* button to create an identical pipeline. The Web Client prompts you for a pipeline name and then displays a populated pipeline definition form, as shown in the second figure below.

3. Complete the pipeline definition by editing the pipeline definition form, as described in Editing Pipelines.

4. Click *Save* to create the pipeline.

When you create an empty pipeline, the Web Client displays an empty pipeline definition form:

When you create a pipeline from a results file or by cloning, the Web Client prompts you for a pipeline name and then displays a populated pipeline definition form:
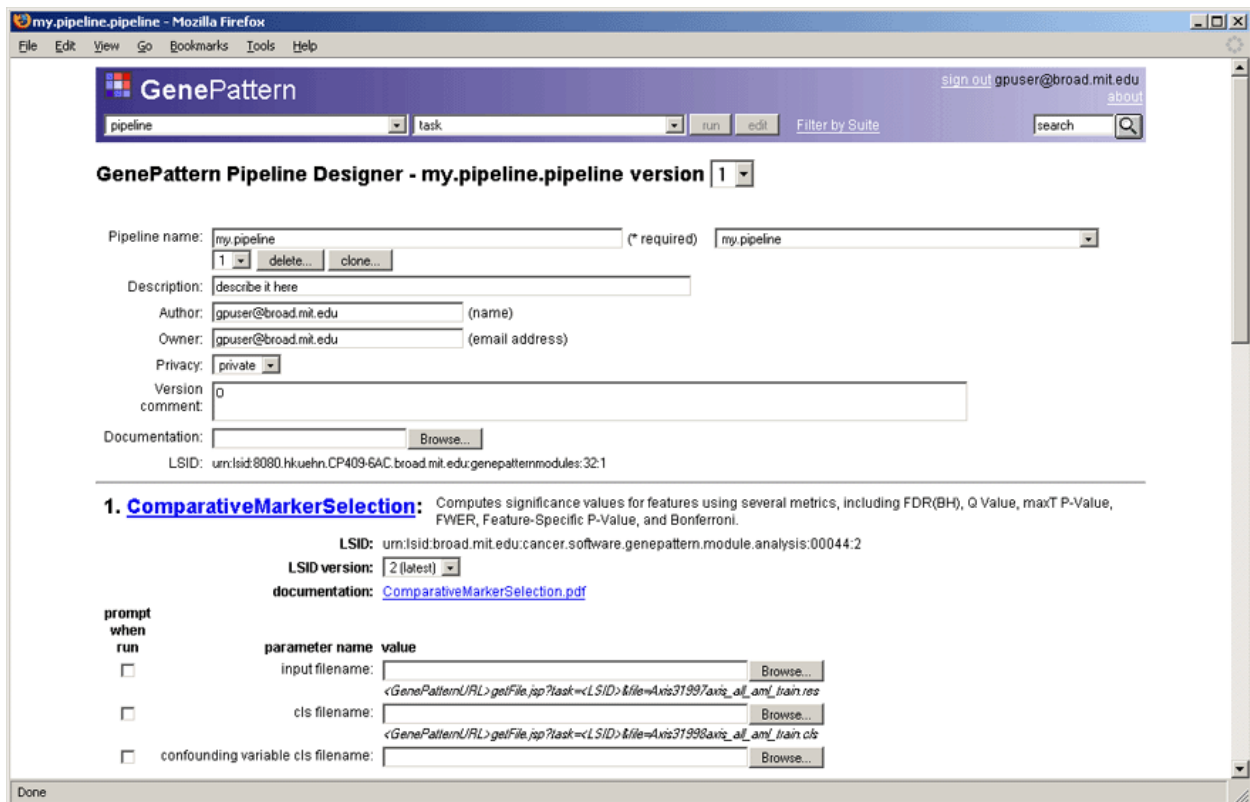
Editing Pipelines
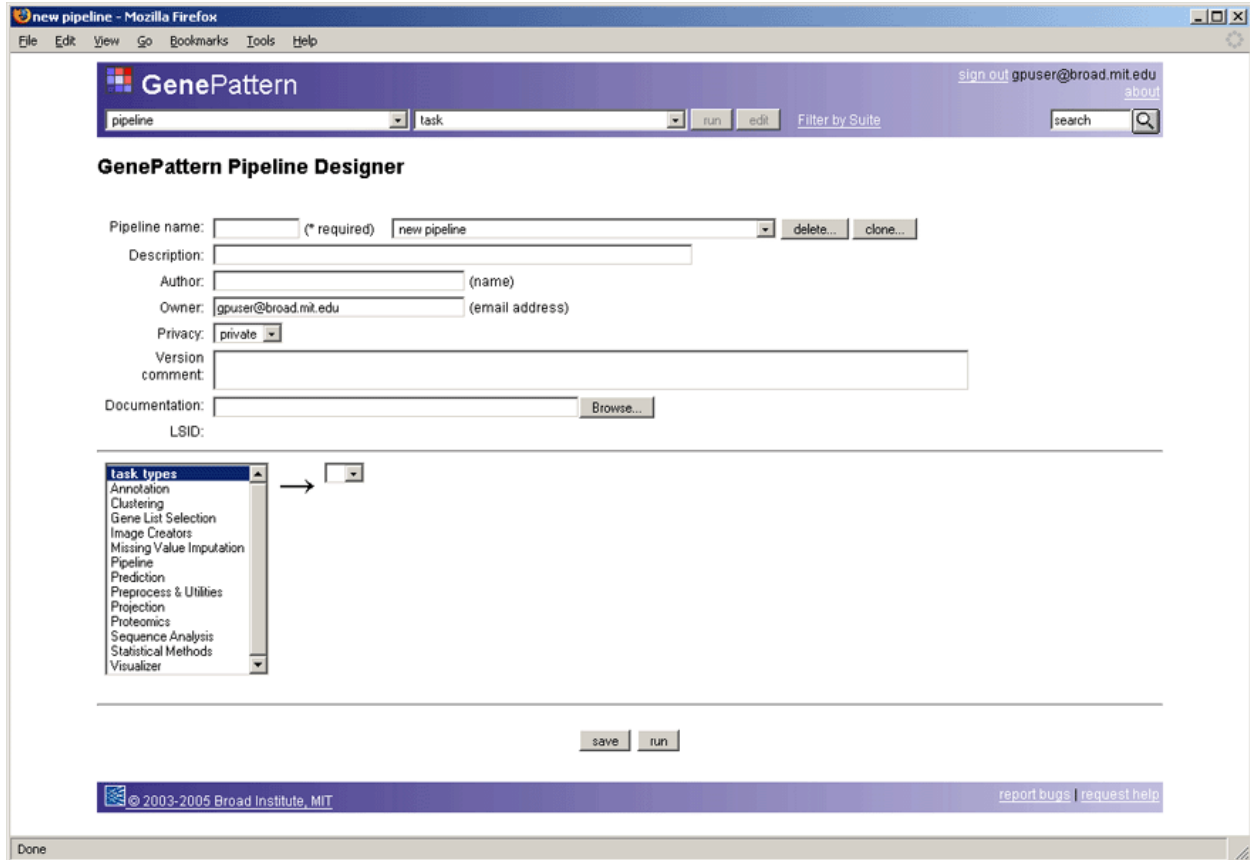
To edit a pipeline:

1. Display the Web Client home page.

2. Display a pipeline definition form in one of the following ways:

   ▪ Create a pipeline, as described in Creating Pipelines.

   ▪ In the Pipelines pane, select a pipeline that you created and click *Edit*.

   ▪ In the Pipelines pane, select a pipeline created by another user and click *View*. From that window, you can then click *Clone*, which creates a copy of the pipeline. You created the copy, so you can edit it.

3. Edit the pipeline definition, as described below.

4. Click *Save* to create a new version of the pipeline.

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

● Editing the Pipeline Description

● Adding Tasks

● Removing Tasks

● Reordering Tasks

Editing the Pipeline Description

The fields at the top of the pipeline definition form describe the pipeline. For descriptions of these fields, see Editing the Pipeline Description in the *GenePattern Java Client Guide*.
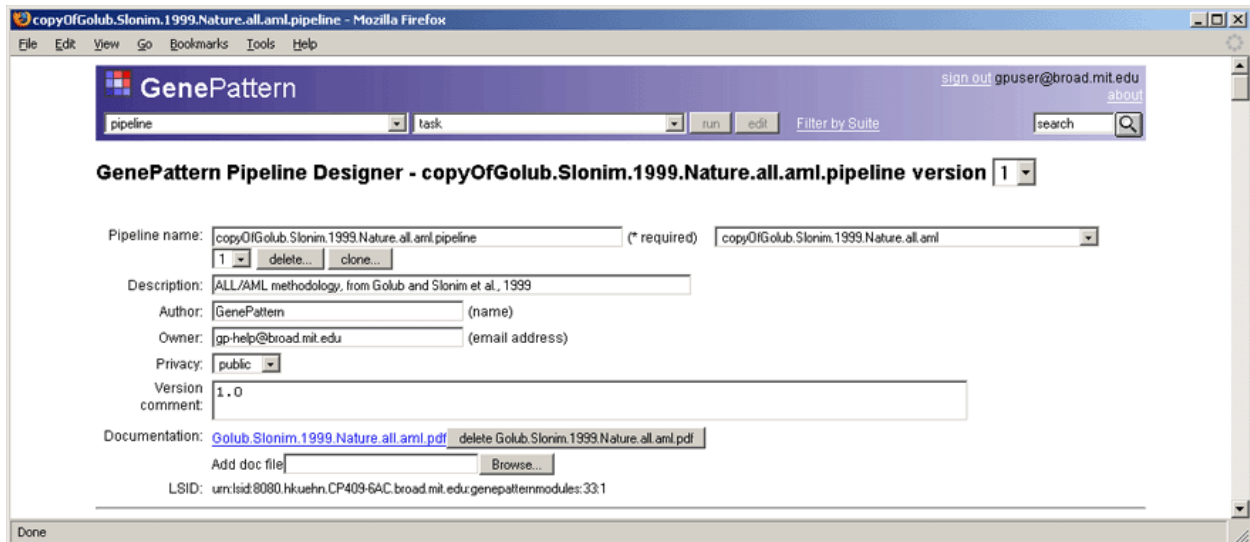
Adding Tasks

Each task in a pipeline has its own definition; for example:
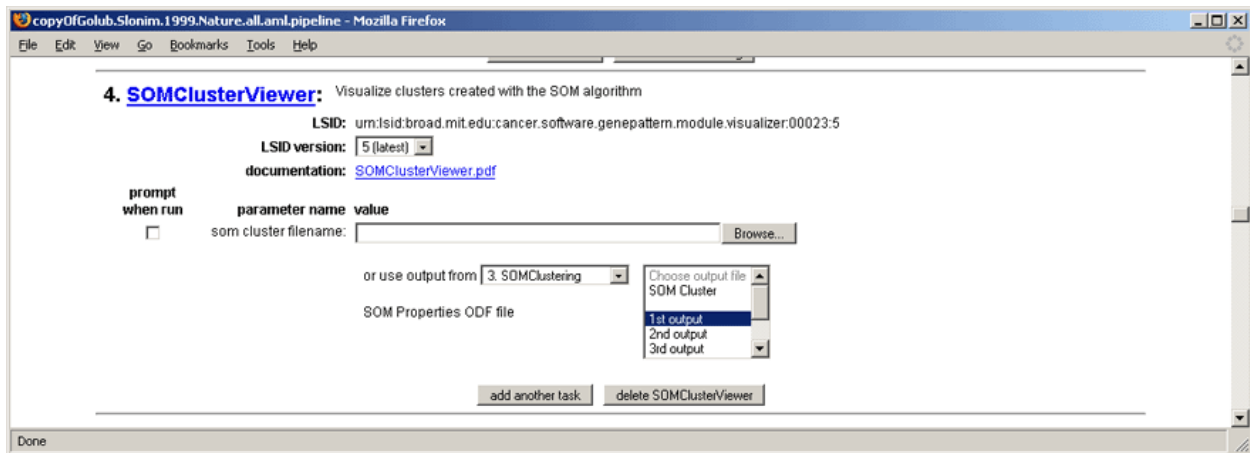
To add a task:

1. Select the *Add Another Task* button in an existing task definition. The Web Client starts a new task definition **after** the current task:
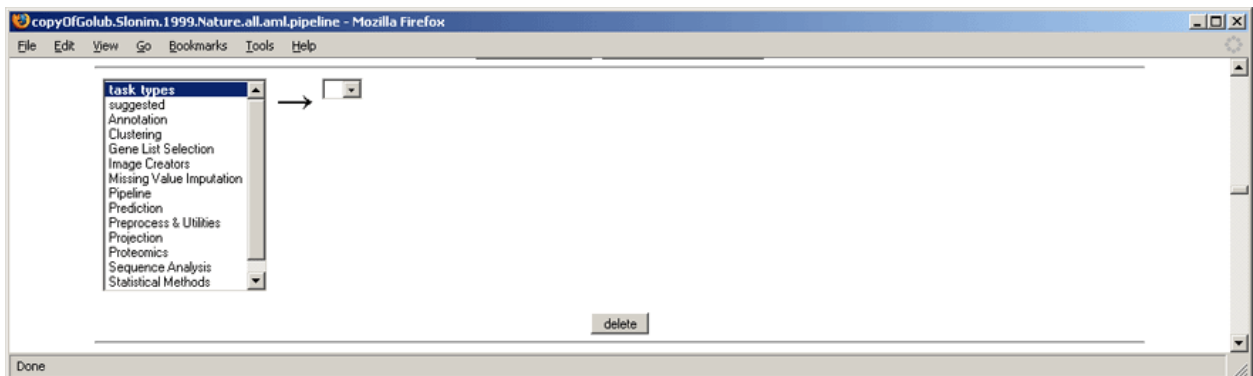
    Note: The only task in a new empty pipeline definition is the start of the first task.

2. Select a category of tasks from the task types list. The Web Client displays a drop-down list of tasks in that category:
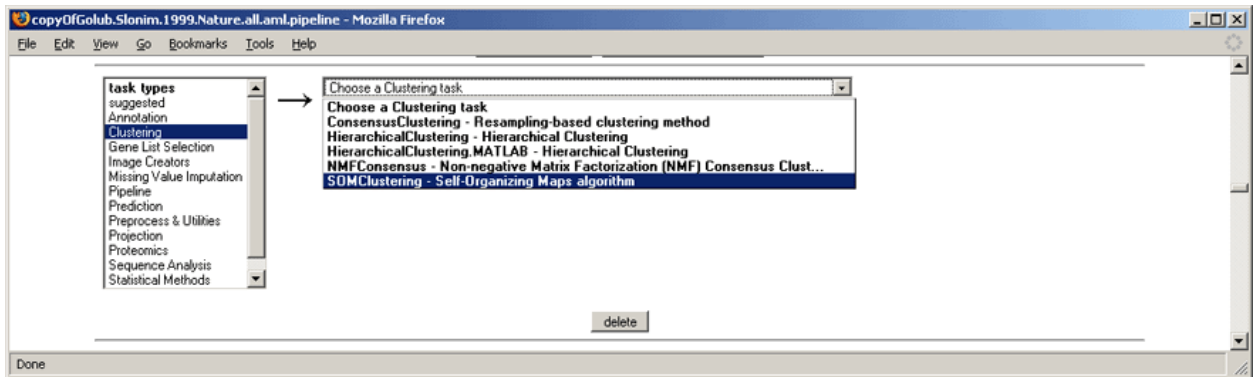
3. Select a task from the drop-down list of tasks. The Web Client displays the definition form for that task:
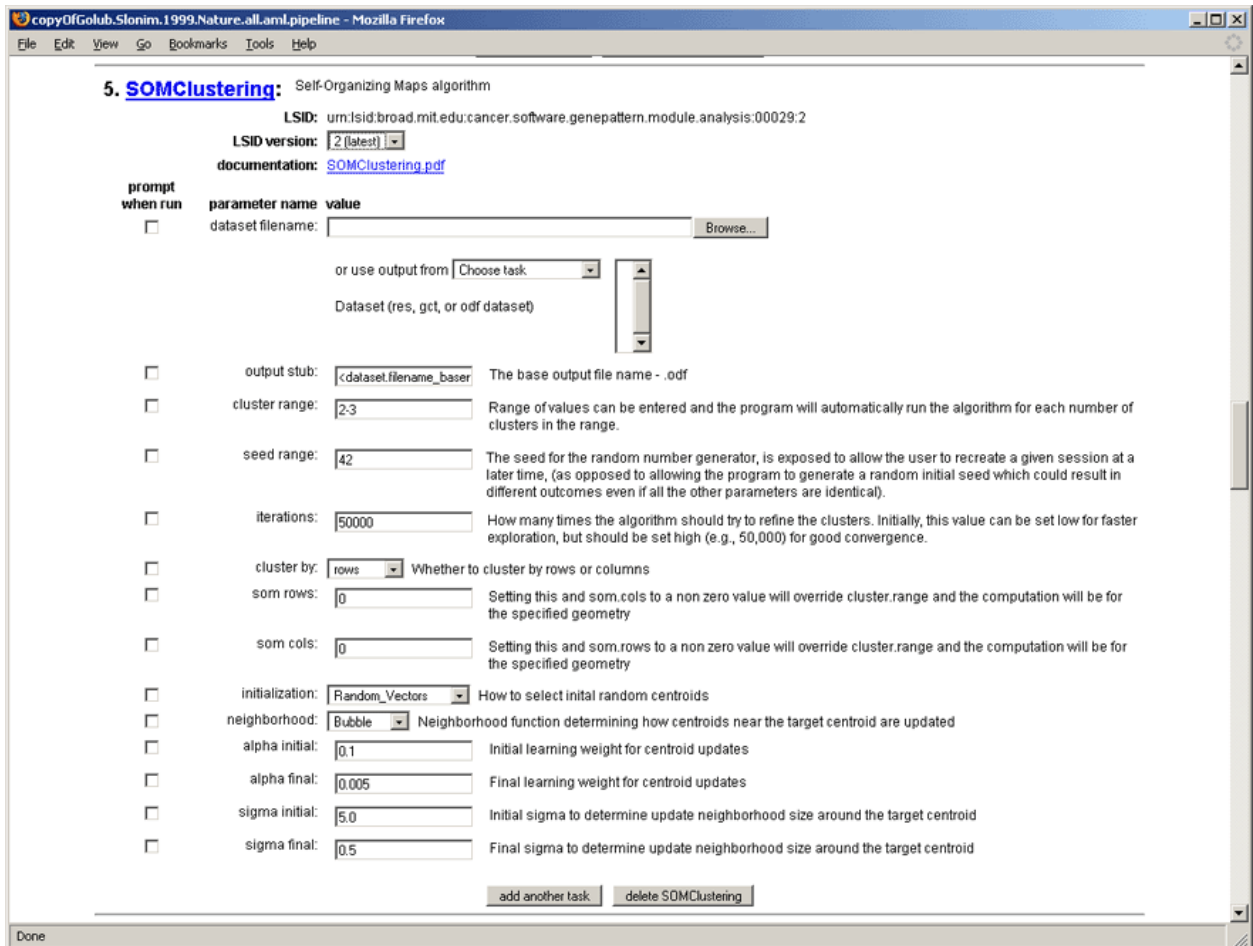
4. At the top of the task definition form is the task name, followed by a drop-down list of versions. By default, you are using the most recent version of the task. To have the pipeline run a different version of the task, select that version from the drop-down list.

5. Enter the parameter values for the task. For most parameters, you enter a value, select a value from a drop-down list, or use the default value supplied by GenePattern.

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

   For input file parameters, select a file or use an output file from a previous task:

   ▪ To select a file, click the *Browse* button to the right of the parameter field. The Web Client displays the Upload File window. Select a file and click *Open*. The selected file becomes the value for that parameter field.

27

- To have the input file be an output file generated by a previous task, choose a previous task from the *use output from* drop-down list. If the selected task generates multiple output files, the Web Client displays a drop-down list of the generated output files. Select the output file that you want to use.

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

Removing Tasks

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

Reordering Tasks

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

Deleting Pipelines

Deleting a pipeline is the same as deleting a task. For more information, see Deleting Tasks.

Managing Suites

Use suites to group modules (tasks and pipelines) into packages that have similar functionality; for example, the ClusteringSuite contains the Consensu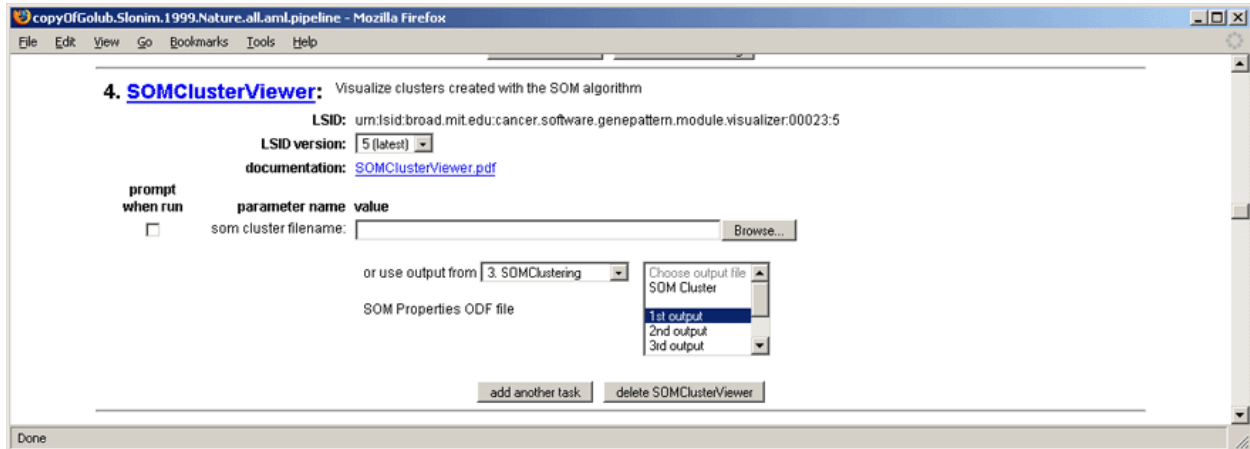sClustering, HierarchicalClustering, HierarchicalClustering.MATLAB, NMFClustering, and SOMClustering tasks. You can work with the modules in a suite as a group; for example, you can install all of the modules in a suite or filter the Web Client menus so that they list only the modules in a selected suite.

This section describes how to manage suites:

- Installing Suites
- Exporting and Importing Suites
- Viewing and Editing Suites
- Creating Suites
- Using Filters
- Deleting Suites

You can export, import, view, edit, create, and filter by suites from either the Web Client or the Java Client. You can install/update and delete suites only from the Web Client.

Installing Suites

The Broad Institute maintains a repository of suites. When you install suites, you search the repository for suites that you do not have on your GenePattern server and, optionally, add them to your server. When you update suites, you search the repository for newer versions of suites that you have installed on your GenePattern server and, optionally, add them to your server.

When you install/update a suite, you install/update the suite and its definition; you also install/update the tasks and pipelines contained in that suite. For example, installing the ClusteringSuite, adds the ClusteringSuite to your GenePattern server; it also installs/updates the ConsensusClustering, HierarchicalClustering, HierarchicalClustering.MATLAB, NMFClustering, and SOMClustering tasks, which are contained in the ClusteringSuite.

To install and update suites:

1. Display the Web Client home page.

2. Under Suites in the right column, click *install/update suites.*

    The Web Client retrieves the list of suites from the repository, compares that list to the suites on your GenePattern server, and displays the Manage Suites page. The top portion of the page lists the suites that you have not yet installed; the bottom portion lists suites that are on your GenePattern server.
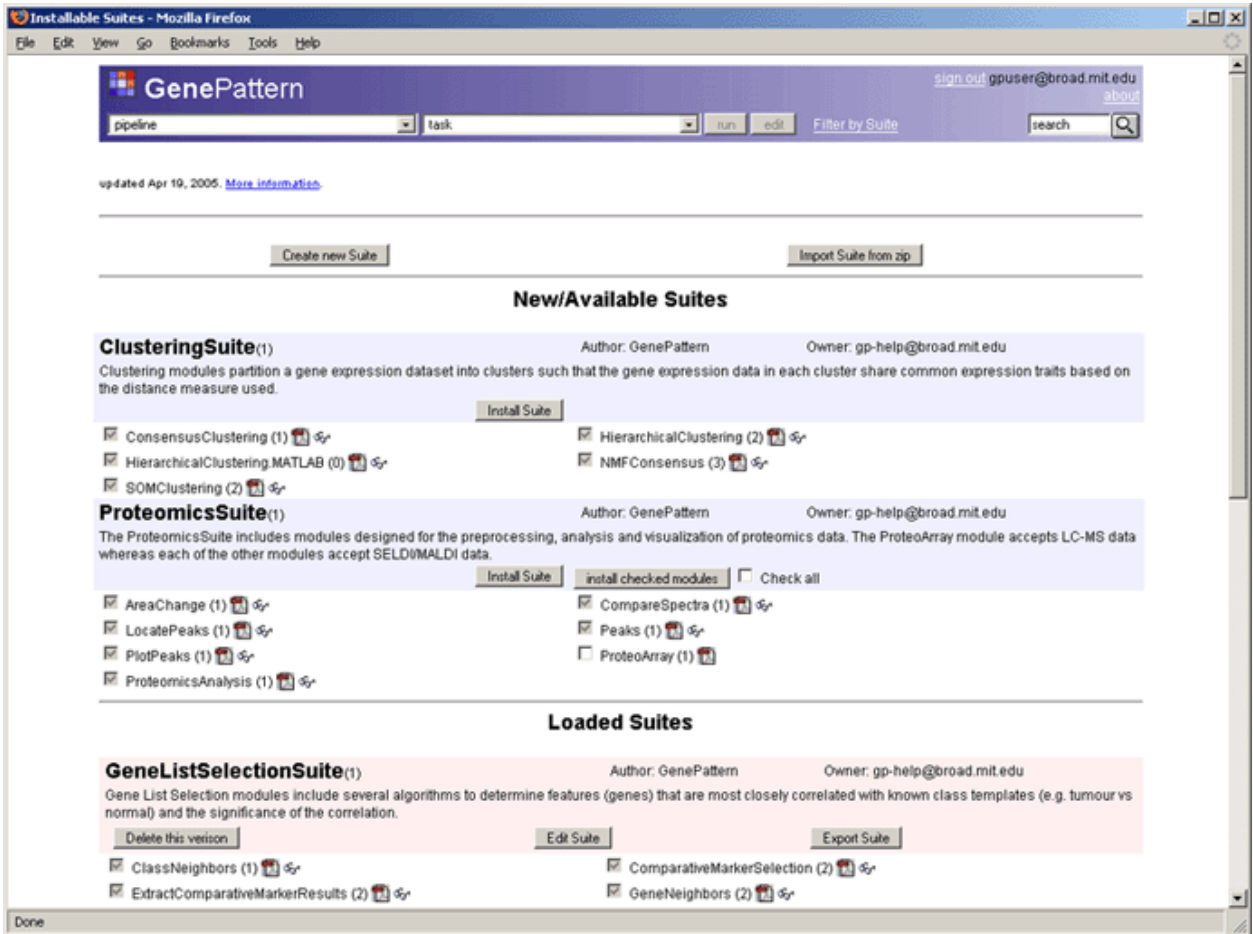
3. Choose a suite to install and click its *Install Suite* button. The Web Client displays a log page listing the modules in that suite as they are installed:

Exporting and Importing Suites

GenePattern allows you to share suites with other GenePattern users by exporting and importing the suites. Exporting a suite copies the suite definition from your GenePattern server to a zip file; importing a suite copies the suite from a zip file to your GenePattern server.

To export a suite:

1. Display the Web Client home page.

2. Under Suites in the right column, click *install/update suites*.

   The Web Client displays the Manage Suites page. The bottom portion of the page lists the suites that are on your GenePattern server.

3. Choose a suite to export and click its *Export Suite* button. The Web Client creates a zip file and displays a window similar to the following (the exact window that appears depends on your operating system):
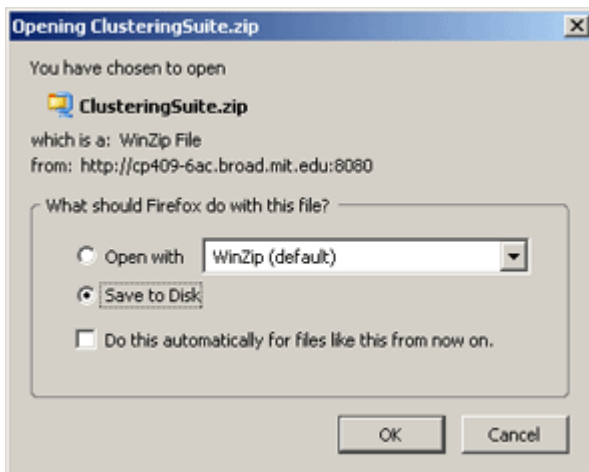
4. Open the zip file and then save it; or save the zip file to disk.

To import a task:

1. Display the Web Client home page.

2. Under Suites in the right column, click *import*. The Web Client displays the Import page:
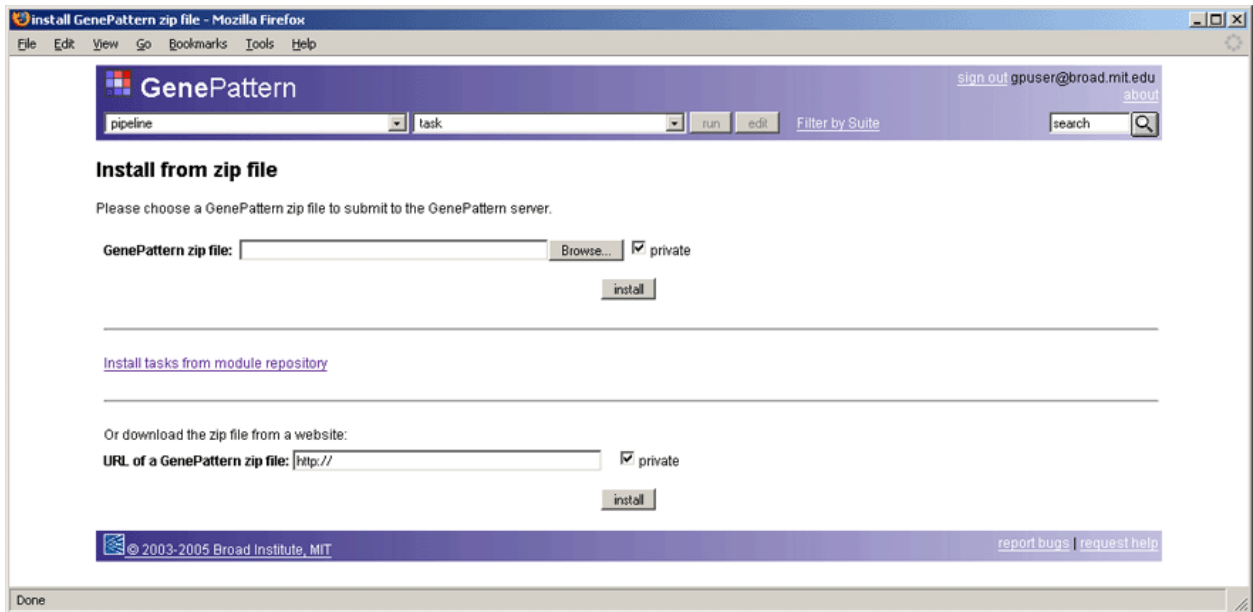
3. Select the zip file and click *install*. Use the top portion of the page to browse for the zip file or the bottom portion of the page to enter a URL for the zip file. The Web Client imports the suite.
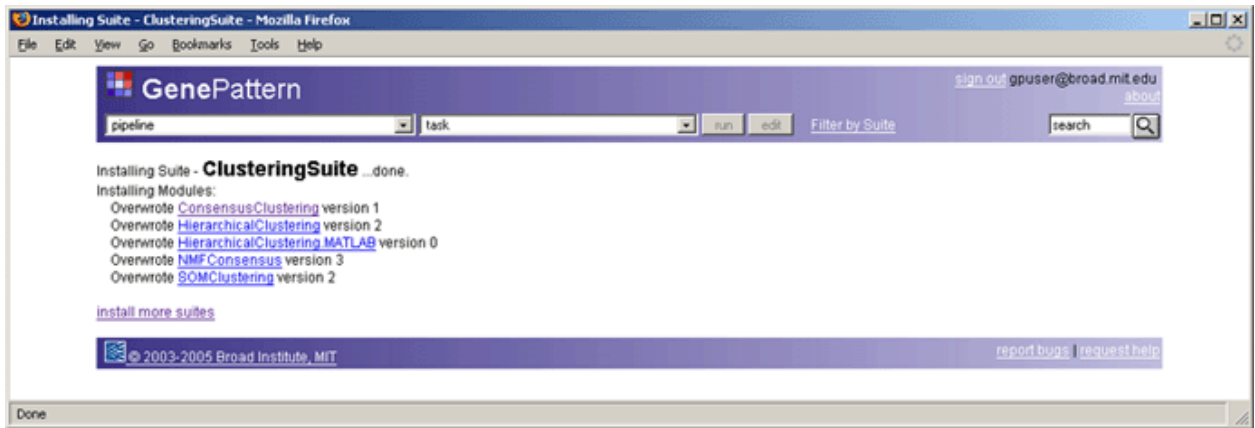
Viewing and Editing Suites

To view or edit a suite:

1. Display the Web Client home page.

2. Under Suites in the right column, click *install/update suites*.

   The Web Client displays the Manage Suites page. The bottom portion of the page lists each suite on your GenePattern server and its definition.

   Choose a suite to edit and click its *Edit Suite* button. The Web Client displays the Suite Definition form, as shown below. The definition form lists all of the modules (tasks and pipelines) that are on your GenePattern server. A check mark next to a module indicates that it is in this suite.

3. To edit the suite definition, edit the fields at the top of the form or toggle the check mark next to a module by clicking the check box.

4. Click *Save*. GenePattern creates a new version of the suite.
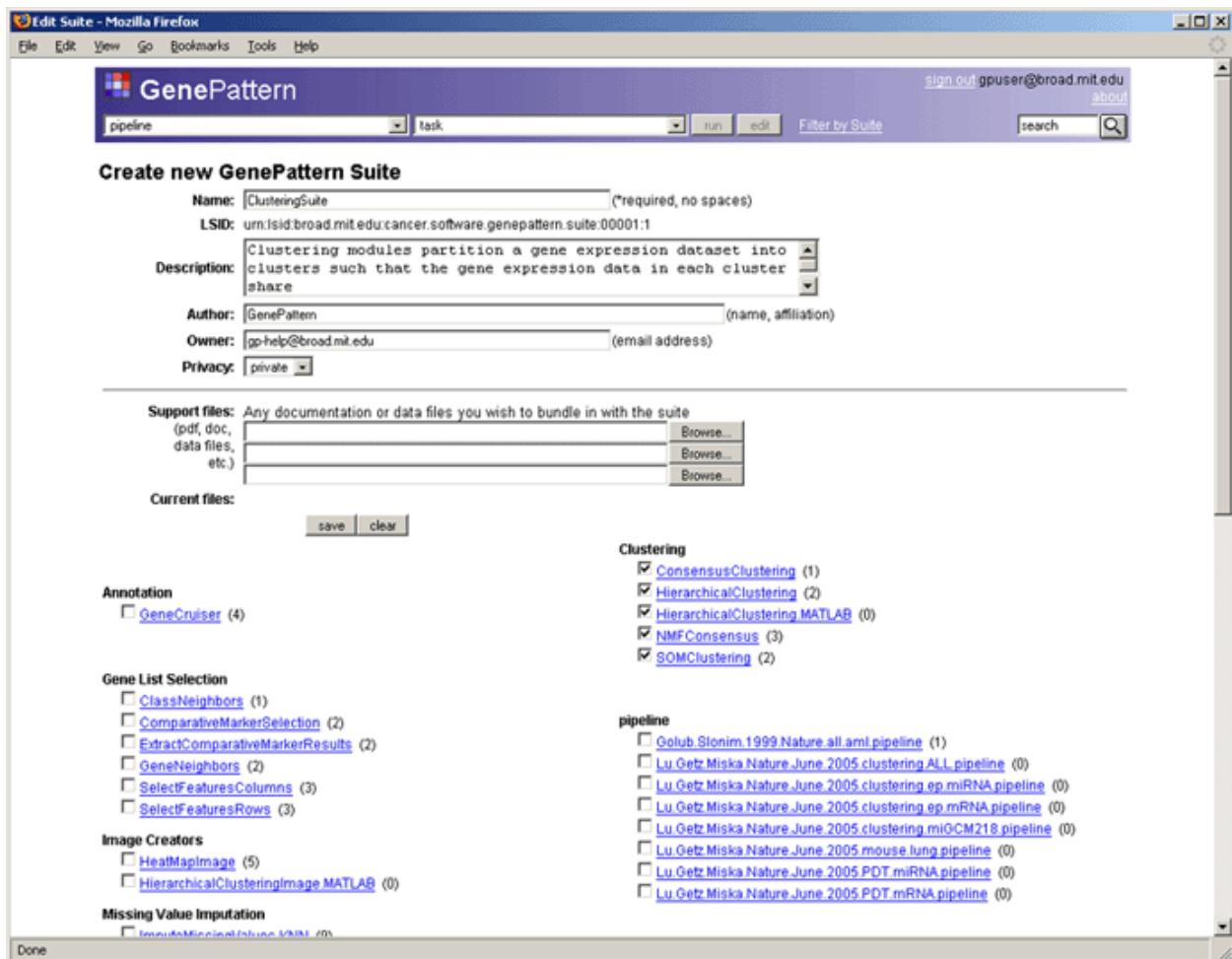
All suites use a similar definition form:

- *Name*: The name of the suite.

- *LSID*: The Life Science Identifiers (LSIDs) for this suite. GenePattern maintains this field; you cannot edit it. For more information, see Understanding Version Numbers in the *GenePattern Java Client Guide*.

- *Description*: A brief description of the suite.

- *Author*: The name and organization of the suite's author. Typically, you are the author.

- *Owner*: The user name of the person responsible for maintaining the suite. By default, this is your user name (typically, your e-mail address).

- *Privacy*: Select Private (default) or Public. A private suite can only be seen and used by a user logged in using the user name specified in the *Owner* field of this form; a public suite can be seen and used by all users.

- *Support files*: Any files that you want to include with the suite. Documentation is strongly encouraged for public suites.

- *Modules*: The remainder of the definition form lists all of the analysis modules, visualization modules, and pipelines that are on your GenePattern server. A check mark next to a module or pipeline indicates that it is in this suite.

Creating Suites

To create a suite:

1. Display the Web Client home page.

2. Under Suites in the right column, click *create suite*. The Web Client displays an empty suite definition form.

3. Define the suite by entering values for each field, as described in Viewing and Editing Suites.

4. Click *Save*. GenePattern creates the suite.

Using Filters

You can use suites to filter the Web Client menus so that they list only the modules (tasks and pipelines) in selected suites.

To apply a filter:

1. Click *Filter by suite* in the title bar of the Web Client. The Web Client displays the Filter window:
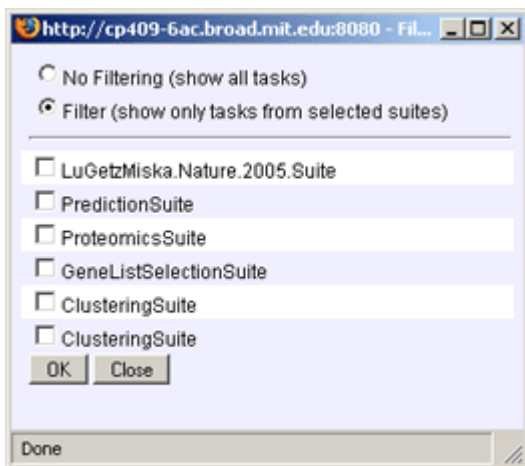
2. Select the *Filter* radio button.

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

4. Click *OK*. The Web Client updates its menus to show only the modules in the selected suites.

To remove the filter from the menus:

1. If necessary, click *Filter by suite* in the title bar of the Web Client to display the Filter window.

2. Select the *No Filtering* radio button.

3. Click *OK*. The Web Client updates its menus to show all of the modules on your GenePattern server.

Deleting Suites

To delete a suite:

1. Display the Web Client home page.

2. Under Suites in the right column, click *delete suite*.

    The Web Client displays the install/update page. The bottom portion of the page lists the suites that are on your GenePattern server.

3. Choose a suite to delete and click its *Delete this version* button. The Web Client deletes the selected suite.

Managing the GenePattern Server

GenePattern can be run standalone on a small machine or separated into its client and server components to take advantage of a more powerful compute server. When you installed your GenePattern server, you set the basic server configuration options for your site. This section describes how to modify GenePattern server configuration options and how to manage other aspects of the GenePattern server environment:

● Modifying Server Configuration Options

● Using a Queuing System

● Securing the Server

Modifying Server Configuration Options

You can modify GenePattern server configuration options in two ways:

● Use the Server Administration page of the Web Client to modify the configuration options that most commonly need to be changed, as described in this section.

- Use the `GenePatternServer/resources/genepattern.properties` configuration file to modify additional configuration options, which rarely need to be changed. The next section, Using a Queuing System, describes a situation in which you modify the configuration file.

To modify commonly changed server configuration options:

1. Display the Web Client home page.

2. Under Server Administration in the right column, click *modify settings*. The Web Client displays the Server Administration page:
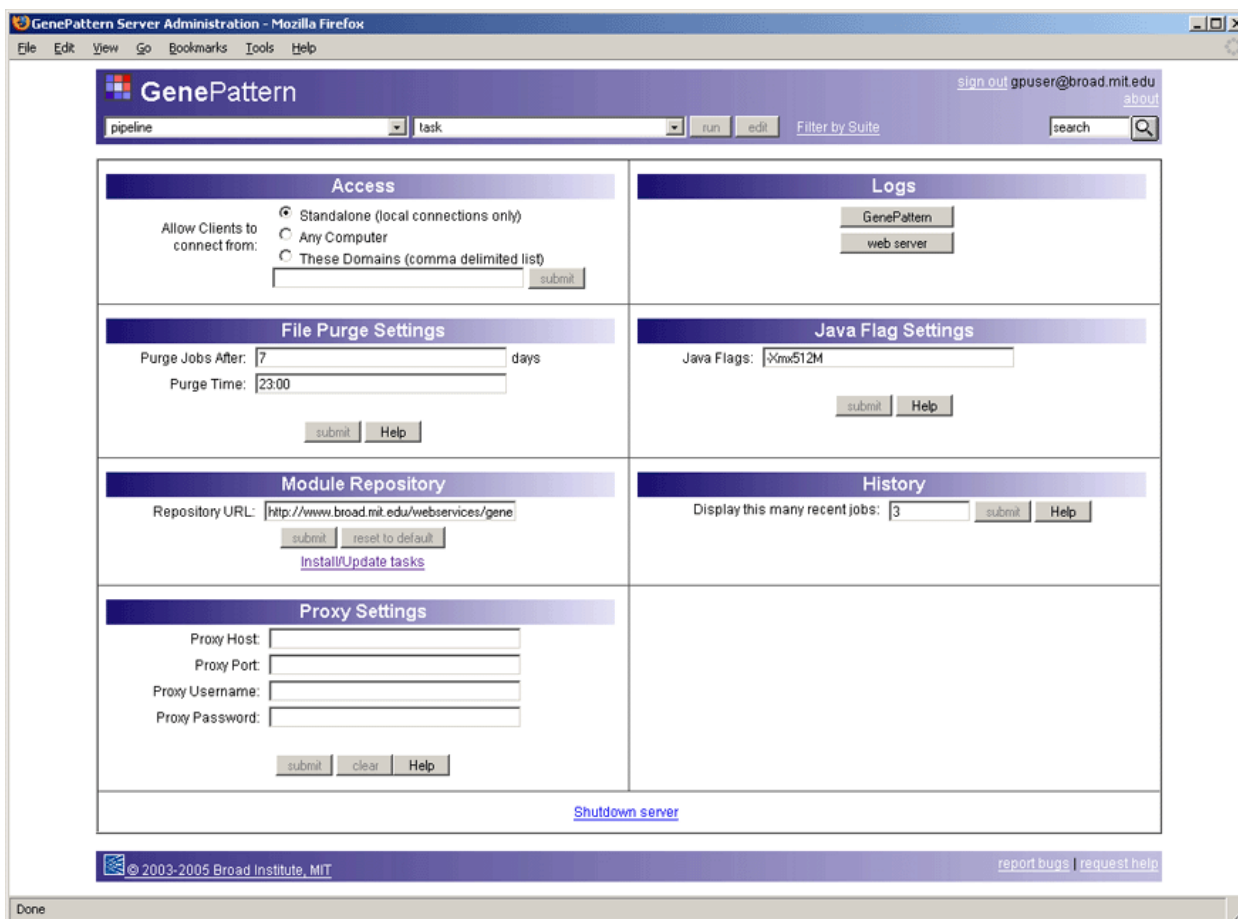
From this page, you can modify configuration options and view the server logs:

- **Access.** Use these options to define which clients have access to the server:

  - Click *Standalone* to allow only local clients to connect to the server; that is, you can access this GenePattern server only from the computer that it is running on.

  - Click *Any Computer* (default) to allow any client to connect to the server.

  - Click *These Domains* to allow only clients from specific domains to connect to the server. Enter a comma-separated list of domains or IP addresses in the text box, for example: `broad.mit.edu,dfci.harvard.edu,mit.edu`.

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

  For more information about restricting access to your server, see Securing the Server.

- **File Purge Settings.** Use these options to specify when analysis results files are deleted from the server:

  - Use *Purge Jobs After* to specify the number of days the server keeps the analysis results files. To prevent the server from automatically deleting the files, set this value to -1.

  - Use *Purge Time* to specify what time of day (24-hour format) the server deletes the files.

34

Note: To save your analysis results files, use the Job Results page to download them from the server to permanent storage.

- **Module Repository.** When you install/update modules (tasks, pipelines, or suites), GenePattern searches the repository for new/updated modules and, optionally, adds them to your server. Use this option to change the location of the repository that GenePattern uses to install/update modules. By default, this option points to the module repository maintained by the Broad Institute. If you would like to implement and maintain a module repository at your site, contact the GenePattern help desk (gp-help@broad.mit.edu).

- **HTTP Proxy Settings.** If your server is behind a firewall, use these options to specify the HTTP Proxy information. Without the proxy information, the server cannot download modules (tasks, pipelines, or suites) from the repository maintained by the Broad Institute. If you do not know the proxy information, contact your systems administrator.

- **Logs.** Use these options to view the server log files:

  - Click *GenePattern* to view warnings and messages generated by the GenePattern server.

  - Click *web server* to view messages generated by the web server that GenePattern uses.

- **Java Flag Settings.** Use this option to specify any Java command line flags that are to be used by all Java-based analysis modules. By default, this option is used to specify the amount of memory to allocate when running Java-based analysis modules (-Xmx512M).

- **History.** Use this option to specify the number of recent analyses GenePattern displays at the top of the pipeline and tasks drop-down lists in the title bar of the Web Client home page.

You can shutdown the GenePattern server by clicking the *Shutdown server* link. For easier ways of shutting down the server, see Exiting from GenePattern.

Using a Queuing System

This section describes how you can integrate the GenePattern server with your queuing system by adding a property to the `GenePatternServer/resources/genepattern.properties` configuration file. If you need information about other properties in the configuration file or help with integrating the GenePattern server into your environment, contact the GenePattern help desk (gp-help@broad.mit.edu).

Queuing systems such as the Load Sharing Facility (LSF) and the Sun Grid Engine (SGE) allow computational resources to be used effectively. If you have such a queuing system, you typically want the GenePattern server to use it. To configure your GenePattern server to execute jobs using LSF or SGE, add the commandPrefix property to the `GenePatternServer/resources/genepattern.properties` configuration file. The server prepends the value of this property to the command line each time it executes a task.

For example, to configure GenePattern to run jobs using LSF:

1. Add the following commandPrefix property to the `genepattern.properties` file:

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

   - The –K flag instructs the bsub command to wait for the job to complete before returning.

   - The –o flag specifies the file to which the job writes standard output and standard error messages.

2. Optionally, set the environment variables BSUB_QUIET and BSUB_QUIET2 to prevent bsub from printing common job messages to standard out:

   - Setting BSUB_QUIET prevents bsub from printing the messages <<Job is submitted to default queue <normal>>> and <<Waiting for dispatch>>.

   - Setting BSUB_QUIET2 prevents bsub from printing the message <<Job is finished>>.

Securing the Server

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

This section describes several ways to secure your GenePattern server:

- Access Filtering
- User Authentication

- User Permissions
- Secure Sockets Layer (SSL) Support

Access Filtering

Access filtering prevents users from connecting to your GenePattern server unless they come from a known computer. This is the simplest way to secure your GenePattern server because you control access filtering from the Server Administration page of the Web Client.

To use access filtering (as described in Modifying Server Configuration Options):

1. Display the Web Client home page.

2. Under Server Administration in the right column, click *modify settings*. The Web Client displays the Server Administration page.

3. Use the Access pane to determine which clients have access to your GenePattern server:

- Click *Standalone* to allow only local clients to connect to the server; that is, you can access this GenePattern server only from the computer that it is running on.

- Click *Any Computer* (default) to allow any client to connect to the server.

- Click *These Domains* to allow only clients from specific domains to connect to the server. Enter a comma-separated list of domains or IP addresses in the text box, for example:
  `broad.mit.edu,dfci.harvard.edu,mit.edu`.

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

User Authentication

By default, your GenePattern server does not require a password to authenticate that a user is who they claim to be. However, stronger authentication can easily be added by any capable Java programmer at your site. The remainder of this section is written for the Java programmer.

GenePattern user authentication is performed by a servlet filter installed in front of the GenePattern web application in its `web.xml` file. To implement stronger authentication, you need to implement a new servlet filter that performs the desired authentication (for example, validating login against a database) and a new login page that requests a user name and password from the user. The new servlet filter must, after authenticating a user, set a session attribute in the `java.servlet.http.HttpSession` object called `userID`. All pages in the GenePattern server and all database interactions will use this as the user's user name.

To see the source code for the default `AuthenticationFilter.java` and the `web.xml` files, click the following links: AuthenticationFilter.java, web.xml.

Once you have written and compiled the new ServletFilter:

1. Place the jar file containing the new ServletFilter into the following directory:

   `*/GenePatternServer/Tomcat/webapps/gp/WEB-INF/lib`

2. Modify the GenePattern server's `web.xml` document.

   Note: It is important to maintain the existing order of the servlet filters in the `web.xml` document as they are used in the order they are defined in the document. The Authentication filter must come before the Authorization filter for the Authorization filter to work.

3. Change the definition of the AuthenticationFilter to use the class that you have provided.

4. Add any necessary configuration elements that it requires.

5. Restart the GenePattern server for the changes to take effect.

   Note: If you look at the code for the default Authentication Filter (AuthenticationFilter.java), you will see that it allows requests through that have a parameter called `jsp_precompile` that have come from the `localhost`. If you do not allow these requests through unauthenticated, you will see a series of errors when you start the GenePattern server as it attempts to precompile the JSP pages. These are **not** fatal errors, but they slow down server response for users the first time that pages are accessed following a server restart.

User Permissions

As of GenePattern 2.0, each GenePattern server has a built-in authorization mechanism based on three configuration files in the `GenePatternServer/resources` directory:

- `userGroups.xml` defines user groups

- `permissionMap.xml` defines which user groups have which permissions

- `actionPermissionMap.xml` defines which permissions control which server actions

User authorization is based on the user's identity as provided at GenePattern login (for more information, see User Authentication). Once a user has been authenticated, the GenePattern server uses these three configuration files to determine which permissions the user has and which actions those permissions allow the user to perform. By default, all users can perform all actions.

This section describes the default permissions loaded into GenePattern and the three configuration files that define those permissions: userGroups.xml, permissionMap.xml, and actionPermissionMap.xml.

Note: GenePattern authorization, like GenePattern user authentication, is performed by a servlet filter installed in front of the GenePattern web application in its web.xml file. To implement stronger authorization, you need to implement a new servlet filter that performs the desired authorization. It is important to maintain the existing order of the servlet filters in the web.xml document as they are used in the order they are defined in the document. The Authentication filter must come before the Authorization filter for the Authorization filter to work. To see the source code for the default `AuthorizationFilter.java` and the `web.xml` files, click the following links: AuthorizationFilter.java, web.xml.

Default Permissions

The default configuration files installed with GenePattern define the following permissions (all users have all permissions):

- **createTask** allows a user to create new analysis tasks on the GenePattern server by

  - creating or editing a task

  - installing/updating a task

  - installing/updating a pipeline that includes tasks

  - installing/updating a suite that includes tasks

  - importing a zip file that contains a task

  Note: This is a **critical permission** to control. Users with createTask permission can upload any executable to be run on the server including worms, viruses and other malicious software.

- **createPipeline** allows a user to create new pipelines by

  - creating or editing a pipeline

  - importing a zip file that contains a pipeline (if a pipeline zip file includes task zip files, the tasks will not be imported unless the user also has the createTask permission)

- **createSuite** allows a user to create new suites by

  - creating or editing suite

  - importing a zip file that contains a suite (if a suite zip file includes task zip files, the tasks will not be imported unless the user also has the createTask permission)

- **administrateServer** allows the user to change server settings on the administrateServer page.

  Note: This is a **critical permission** to control. Users with administrateServer permissions have the ability to modify connection settings thus preventing other users from logging into and using the server.

- **deleteTask** allows the user to delete tasks from this GenePattern server.

- **deleteJob** allows the user to delete jobs from this GenePattern server.

userGroups.xml

To see the default `userGroups.xml` file, click here: userGroups.xml. If you are defining permissions, apply file permissions to ensure that this file is not modified unintentionally.

The `userGroups.xml` file is used to define groups of users to whom permissions may be assigned. Any number of groups may be created as long as their names are unique. If the user named * is included in a group, as shown in the default settings, then any and all users are considered to be a part of that group. A user may belong to any number of groups.

permissionMap.xml

To see the default `permissionMap.xml` file, click here: permissionMap.xml. If you are defining permissions, apply file permissions to ensure that this file is not modified unintentionally.

The `permissionMap.xml` file is used to map permission names to user groups; that is, you use this file to state that members of a particular group have a particular permission.

Any number of new permissions may be created as long as their names remain unique. Any number of groups may be assigned a permission by adding additional `<group name="thename"/>` elements to a permission. The presence of a group named * means that any and all groups have that permission, which means any and all users have that permission.

To create a permission that no one is allowed to have, omit its group elements; for example:

`<permission name="NoOneHasThisPermission"></permission>`

actionPermissionMap.xml

To see the default `actionPermissionMap.xml` file, click here: actionPermissionMap.xml. If you are defining permissions, apply file permissions to ensure that this file is not modified unintentionally.

The `actionPermissionMap.xml` file is used to define what permissions apply to what server actions. In this context, server actions correspond to pages in the web application and methods in the SOAP interfaces. Essentially, this file links jsp pages and SOAP methods with their required permission names.

Note: Many of the existing GenePatternServer web pages have been written so that if a user does not have permission to execute a link, the link is not displayed. This logic is implemented in the jsp pages, not in the configuration files. If you limit permissions to other additional pages, users may still see the links but will be redirected to an error page when they attempt to execute them.

Secure Sockets Layer (SSL) Support

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

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

Step 1. Configure Tomcat for SSL support

Follow the instructions available at http://tomcat.apache.org/tomcat-4.1-doc/ssl-howto.html to configure the Tomcat instance for using SSL. In doing so, you will modify the Tomcat configuration file, which is located in the `GenePatternServer/Tomcat/conf` directory.

Step 2.Configure GenePattern for SSL

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

1. Add a new key, `java.net.ssl.trustStore`=<path to keystore>.

This should point to the keystore you created when configuring Tomcat (above) or some keystore that GenePattern can use to establish SSL connections.

2. Modify the value for the key GENEPATTERN_PORT to use the https port you selected when configuring Tomcat (above).

3. Modify the value for the key GenePatternURL to use the https protocol and the https port you selected, for example:

   `http://localhost:8080/gp` becomes `https://localhost:8443/gp`

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