

GenePattern Frequently Asked Questions

General

What is the latest version of GenePattern?

The latest versions of the GenePattern components are:

Component	Version	Build
server	2.0.0	2.0.0.358
client	2.0.0	2.0.0.358

What platforms does GenePattern support?

GenePattern installers are available for Windows, Mac OS X, and Linux. GenePattern 2.0 should work with any flavor of Linux that has a Java 1.4.2 virtual machine installed. GenePattern is known to work on the following OS platforms:

Windows XP, 2000
Mac OS X 10.3, 10.4
Linux SuSE, Red Hat, Mandrake
Other Unix Solaris, AIX, Tru64

Users are also running GenePattern on the Debian, Fedora, and Gentoo distributions of Linux.

What are the hardware requirements for GenePattern?

GenePattern's hardware requirements are found on almost all currently available machines:

- 256 MB RAM
- 500 MHz Pentium 3 or equivalent
- Hard drive space:
 - Server: 434 MB
 - Client: 70 MB

How can I get help with GenePattern or provide feedback?

Send email to gp-help@broad.mit.edu

How do I cite GenePattern?

Please cite this URL regarding the code:

<http://www.broad.mit.edu/genepattern/>

The individual algorithm modules have citations in the documentation. Please refer to the specific documentation for the appropriate references.

If I am a member of the press, how can I get more information?

If you are a member of the press and need additional information about GenePattern, please contact our Public Relations department by calling 617.252.1064.

Installation

How can I get GenePattern?

You can get GenePattern by signing a [license agreement](#) and following the instructions there.

Are there release notes available?

Click [here](#) for the latest release notes.

I already have R/Perl/Java on my machine.

- Will the versions of R/Perl/Java that GenePattern installs interfere with these?

No. The R, Perl, and Java installations that come with the GenePattern are installed within the GenePattern directory and do not affect any other versions that you may currently have.

- Do I need to install yours, or can I configure GenePattern to work with the versions already on my machine?

The versions of R, Perl, and Java that are bundled with GenePattern are the ones that have been fully tested. We cannot guarantee that other versions will work.

I am behind a web proxy/firewall and my GenePattern server says it cannot connect to the module repository to load the modules. What do I do?

First login to your GenePattern server web client (<http://localhost:8080/gp> is usually the URL). Click on the "modify settings" link in the administration section. In the "Proxy Settings" section, enter the hostname and port of your web proxy server. If you do not know them, contact your IT help desk to get the values. If you need to login to the proxy

server, also enter your username and password (these will NOT be saved to a file and will need to be reentered following a server restart next time you want to connect). Save (submit) the values and then click on the "Install/Update tasks" link to go back to the task catalog and load the modules.

I still cannot connect to the repository. How do I install modules?

The modules are all available on the Broad Institute's public FTP server (<ftp://ftp.broad.mit.edu/pub/genepattern/modules>). You can connect to that server and download the zip files individually and then load them into your server via the "import" button in the tasks section. You can also email us at gp-help@broad.mit.edu and we can help you get the modules more quickly via a manual process.

I want to install GenePattern into our corporate/departmental/other Web server and not have GenePattern run in its own Web server. How do I install it?

You need to use the war file installation. Instructions are available at http://www.broad.mit.edu/cancer/software/genepattern/download/gp_server_install_warfile.html.

When should I choose to install the GenePattern server on a different port than the default 8080?

If you already have a server such as Tomcat running on this port, you need to install the GenePattern server on a different port to avoid conflicts. If you install the server on a non-default Tomcat port (not 8080), the GenePattern Client installer will not automatically select this port for you to connect to. You must change it manually during installation. You can choose which port to install in the "GenePattern server configuration" window as the "webserver port" parameter.

Configuration

How do I increase the memory allocated to the GenePattern server or client?

Mac OS X

- Control-click on the file GenePattern/Tomcat/StartGenePatternServer (server) or the GenePatterClient/GenePattern Client icon (client).
- Select 'Show Package Contents' from the pop-up menu. The Contents directory should open in the finder.
- In the Contents directory, double-click the Info.plist file. This should open the Property List Editor program.
- Add the child 'VMOptions' under the 'Java' node.

- Change the Class of the added VMOptions node to 'Array'
- Add the child with Class 'String' with the value -Xmx512M. You can replace the value 512 with the maximum amount of memory in MB that you want the GenePattern Client to use.

Windows and Linux

- Edit the file in the GenePattern directory called StartGenePatternServer.lax (server) or in the GenePatternClient directory called GenePattern Client.lax (client).
- In either of these files, look for the entries noted below and increase these values by doubling up to the max memory size of the machine you are using. (Note: Windows limits the total space available to a process to 2 GB. Some of that is used for overhead, so slightly less is really available to the JRE.)
 - lax.nl.java.option.java.heap.size.initial
 - lax.nl.java.option.java.heap.size.max

How do I increase the memory allocated to a task?

If the task is written in Java, you can do the following:

- **Analytic (non-visualizer) modules:** Edit the file GenePattern/resources/genepattern.properties and look for the entry java_flags. You can change this value up to the maximum memory size of the machine you are using.
- **Visualizers:** In your home directory, open the directory gp. Inside the gp directory edit the file genepattern.properties and add the text java_flags=-Xmx512M on a new line. You can replace the value 512 with anything up to the maximum memory size, in MB, of the machine you are using.

For modules written in other languages, you may or may not be able to change the amount of memory that is allocated to the module.

Can I run more than one instance of the GenePattern server on a machine?

Yes. If you are running more than one installation of GenePattern on the same machine, you must make sure that the port numbers for the GenePattern server and the HSQL server are unique to each installation. The Tomcat server listens on two ports, 8080 (requests) and 8005 (shutdown) by default, and the HSQL server listens on port 9001. All 3 ports need to be modified on the second copy of Tomcat. For example, you can set the GenePattern server port to 8080 and 8005 on one install and 8081 and 8086 on the other, and set the HSQL port to 9001 on one and 9002 on the other. You can configure these port numbers when you are installing the server.

How do I configure GenePattern to use a grid engine?

The [Server Administration](#) section of the GenePattern tutorial discusses how to integrate GenePattern with queuing systems such as the Load Sharing Facility (LSF) and the Sun Grid Engine (SGE).

How do I change the colors that are used to display module names?

You can change the colors used in displaying module names in both the GenePattern Graphical Client and the GenePattern Web Client:

- To change module colors in the GenePattern Graphical Client, open the file `gp.properties` in a text editor. The file is located in the folder `gp` in your home directory. The properties named `authority.foreign.color`, `authority.mine.color`, and `authority.broad.color` control the module colors. Edit the RGB values of these properties and restart the GPGE to change the module colors.
- To change module colors in the GenePattern Web Client, open the file `GenePatternServer/Tomat/webapps/gp/skin/styleSheet.css` in a text editor. The elements `option.tasks-mine`, `option.tasks-broad`, `option.tasks-foreign` control the color of tasks in drop-down menus while the elements `.tasks-mine`, `.tasks-broad`, `.tasks-foreign` controls the colors of task names that appear outside of drop-down menus.

Data Formats

Does GenePattern support cDNA and other 2-channel microarray data?

Yes, you can use most of the GenePattern analyses as easily on 2-channel or ratio-based data as on single channel or absolute value data. To run 2-channel data in GenePattern, do the following:

- If your data is not already in a GenePattern format (RES, GCT), convert it to one of those formats. More information on how to do this in the [File Formats](#) section of the GenePattern tutorial.
- If your data has missing values, convert it using the `ImputeMissingValues.KNN` module, documented on the [Modules](#) page.

Your data can now be analyzed in most of GenePattern's modules. If you want to use non-negative matrix factorization (NMF) and your data contains negative values, see the NMF note in the Analysis Modules and Visualizers section below.

Where can I find information about file formats used by GenePattern?

Information on file formats supported by the modules currently in GenePattern is available in the [File Formats](#) section of the GenePattern tutorial.

How can I convert between RES, GCT, and ODF formats?

Run your file through PreprocessDataset. Select the desired output format for your file. If you only want to convert the file type without filtering, select "no filter" as the choice for the "filter flag" parameter.

How do I convert a file to GenePattern format?

The final exercise in the [File Formats](#) section of the GenePattern tutorial describes how to convert files to GenePattern format.

How can I use CEL and MAGE-ML files in GenePattern?

The ExpressionFileCreator module converts a set of individual CEL files into an expression data set that is usable by GenePattern modules. The MAGEMLImportViewer module imports data in MAGE-ML format into GenePattern.

Graphical Client

How do I remove (close) a project node in the object browser?

Right-click the project node. You will see two choices: Refresh and Remove. Select Remove. The node will be removed from the browser.

When I try to save a file, I see a message that says "An item with that name already exists. Replace/Cancel." But I have not saved a file with that name before. When I click Replace and try to view the file, it isn't there. What is the problem?

On servers running on Windows, the following input filenames will not work, regardless of the extension they are given: con, prn, aux, nul, com1, com2, com3, com4, lpt1, lpt2, lpt3. If your filename consists of one of these with any extension, you will not be able to save it.

Pipelines

Which browsers support the GenePattern pipeline environment?

Pipeline environment functionality has been tested on the following browsers:

Platform	Browsers
Windows	Netscape Navigator, MS Internet Explorer, Mozilla Firefox
Mac	Netscape Navigator, Safari, Mozilla Firefox
Linux	Netscape Navigator

How do I keep my old/existing modules and pipelines when upgrading GenePattern to a new version?

Simply install the new version of GenePattern into the same directory as your previous version and the modules will be found by the new version. Do **not** uninstall first. It is unnecessary and will delete your existing pipelines and modules.

Task Creation and Integration

How can I indicate that I want a user to enter a file as a parameter?

Parameters that are files MUST contain 'filename' in their name. For example, the parameter 'input.filename' is a valid name for an input file, but 'input.file' is not.

How can I retrieve external database information from GenePattern?

The GenePattern server itself does not connect to any database, but modules can and have been written to connect to databases and retrieve data from them including caArray (caArrayImportViewer) and Gene Expression Omnibus (GEOImporter). To connect to any database of your choice, write a simple command-line program to connect to the database and retrieve data into a file format and install this program as a task into GenePattern (see [Adding Modules](#)).

GenePattern will not launch the task that I created. What should I do?

Typically, this is because your task does not have parameters. GenePattern is currently unable to launch modules and visualizers that have zero (0) parameters. To work around this problem, add a dummy parameter with a default value to your module.

Analysis Modules and Visualizers

I have no modules installed, why? How do I get them?

The most likely reason is that when the installer finished, you missed the step to install the modules that came up in a browser window after the installer closed. To install the modules, first login to your GenePattern server web client (<http://localhost:8080/gp> is usually the URL). Click on the link to "Install/Update tasks". The taskCatalog page will come up with all new and/or updated tasks for your system checked. Click the "Install Checked" button and they will be installed onto your GenePattern server.

How can I run non-negative matrix factorization NMF on data that contains negative values, such as log-ratio or unthresholded Affymetrix data?

To run NMF on data that contains negative values, you must do the following (using the method of Kim, P. M. & Tidor, B. (2003) *Genome Res.* **13**, 1706-1718):

- Create one dataset with all negative numbers zeroed
- Create another dataset with all positive numbers zeroed and the signs of all negative numbers removed
- Merge the two (eg. by concatenation), resulting in a dataset twice as large as the original, but with positive values only and zeros, hence appropriate for NMF.

To do this in MATLAB, you can execute the following:

```
anew=[max(a,0);-min(a,0)];
```

where a is the original data.

We are currently developing a GenePattern module to perform this operation as well.

What are the pipelines whose names start with Lu.Getz.Miska.Nature.June.2005?

These pipelines are the actual analyses performed in the Nature paper, "MicroRNA Expression Profiles Classify Human Cancers" by Lu, Getz, Miska, et. al. Running these pipelines will exactly reproduce their results. You can also tweak the parameters used to see for yourself how their results change if they had performed any step in their analyses differently. Additional information about the paper is available at <http://www.broad.mit.edu/cancer/pub/miGCM>.

When I do a Hierarchical Clustering analysis, two files are produced, but the Hierarchical Cluster Viewer (JavaTreeView) looks like it needs three files. Do I need another one?

No, you can use the two files that are created and leave the remaining input box blank. HierarchicalClustering creates a cdt file and one or two additional files: an atr file if you clustered by samples (columns), a gtr file if you clustered by genes (rows), or both atr and gtr files if you clustered by both samples and genes (columns and rows). The JavaTreeView module accepts the two or three files created by HierarchicalClustering.

How can I export a Heat Map image with gene annotations?

The HeatMapView module currently does not include gene annotations with the saved image. Use the HeatMapImage module to include gene annotations.

Why do the scores from ComparativeMarkerSelection and ClassNeighbors differ?

When computing the t-test or signal to noise ratio, ClassNeighbors thresholds the standard deviation to ensure that it is at least twenty percent of the mean. Additionally, if the standard deviation is zero, ClassNeighbors sets it to 0.1.

Programming Language Environments

Where can I find out more about how to launch GenePattern modules from other programming languages?

A reference guide for accessing GenePattern modules from Java, MATLAB, and R is available in the [Programming Language Environment](#) section of the GenePattern tutorial.

Other

If you haven't found what you are looking for, please send an email to gp-help@broad.mit.edu.