

PEPPeR GenePattern Quick Start Guide

This guide assumes a basic understanding of GenePattern and familiarity with loading new modules.

For more information about GenePattern please see the documentation:

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

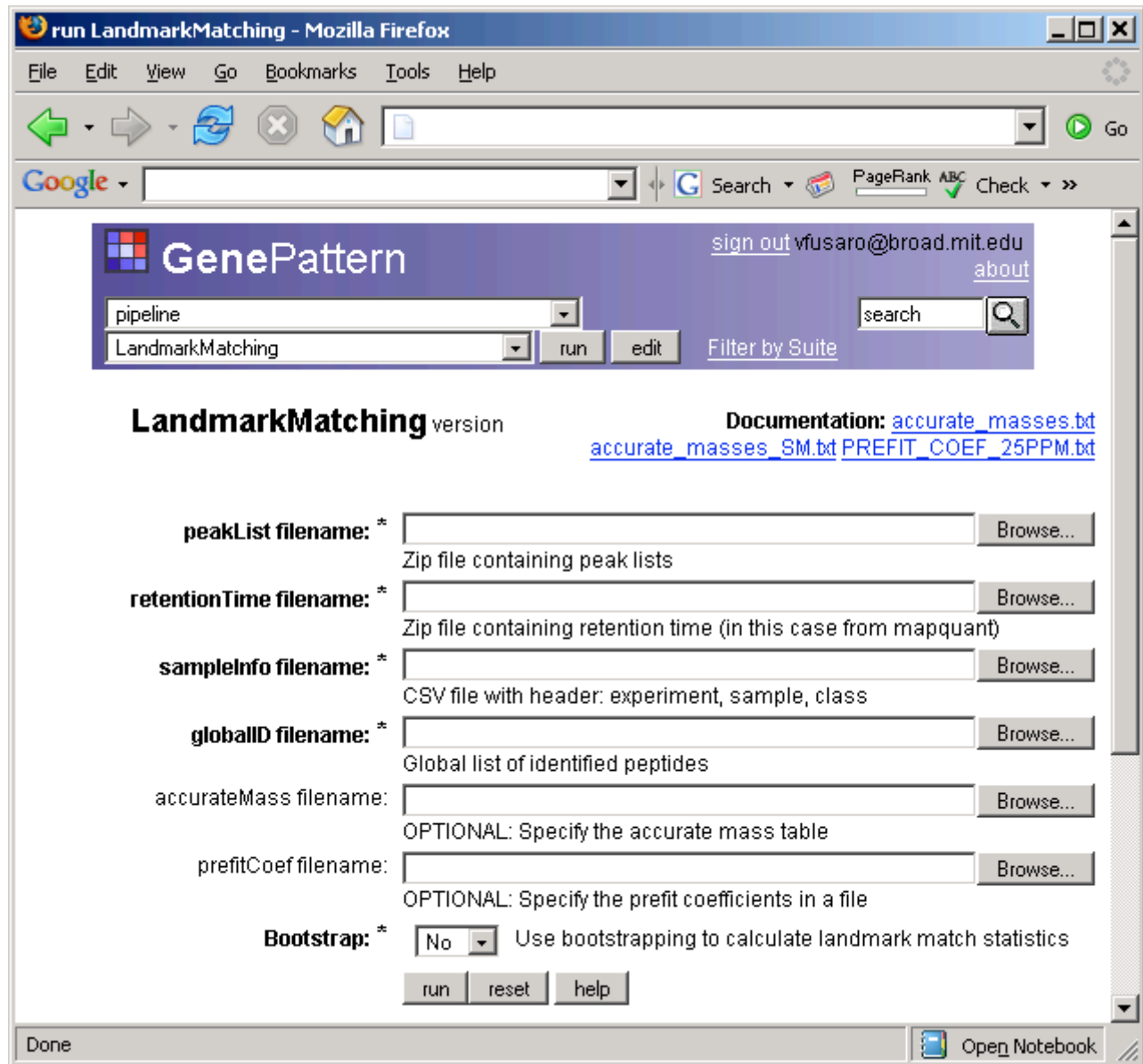
Requirements

PEPPeR is designed to leverage multiple processors in a cluster environment. It will function on a single processor desktop computer but the computation time will increase substantially depending on the number of experiments analyzed.

Example

This guide will quickly walk through using LandmarkMatching and PeakMatching using the provided example file. Please download the *PepperExample.zip* file and extract it to a known location.

LandmarkMatching Step 1: Select the task



The screenshot shows a Mozilla Firefox browser window displaying the GenePattern web interface. The page title is "run LandmarkMatching - Mozilla Firefox". The browser's address bar is empty, and the search bar contains "Google". The GenePattern logo is visible in the top left, and the user is logged in as "vfusaro@broad.mit.edu". The main content area shows the "LandmarkMatching" task configuration page. The task is selected in a dropdown menu, and the "run" button is highlighted. The configuration page includes several input fields with "Browse..." buttons and a "Bootstrap" dropdown menu. The "Bootstrap" dropdown is set to "No". The "run" button is highlighted in blue.

LandmarkMatching version Documentation: [accurate_masses.bt](#)
[accurate_masses_SM.bt](#) [PREFIT_COEF_25PPM.bt](#)

peakList filename: * Browse...
Zip file containing peak lists

retentionTime filename: * Browse...
Zip file containing retention time (in this case from mapquant)

sampleInfo filename: * Browse...
CSV file with header: experiment, sample, class

globalID filename: * Browse...
Global list of identified peptides

accurateMass filename: Browse...
OPTIONAL: Specify the accurate mass table

prefitCoef filename: Browse...
OPTIONAL: Specify the prefit coefficients in a file

Bootstrap: * Use bootstrapping to calculate landmark match statistics

- Select the LandmarkMatching task
- The **bold** options are required inputs

LandmarkMatching Step 2: Fill out the form

The screenshot shows the GenePattern web interface in a Mozilla Firefox browser window. The page title is "run LandmarkMatching - Mozilla Firefox". The browser's address bar is empty. The GenePattern logo is in the top left, and the user is logged in as "vfusaro@broad.mit.edu". The main content area is titled "LandmarkMatching" and includes a "pipeline" dropdown menu set to "LandmarkMatching", a "search" input field, and "run" and "edit" buttons. Below this, there is a section for "LandmarkMatching version" with documentation links: "accurate_masses.bt", "accurate_masses_SM.bt", and "PREFIT_COEF_25PPM.bt". The form contains several input fields with "Browse..." buttons: "peakList filename:" (C:\PepperExample\peakData.zip), "retentionTime filename:" (C:\PepperExample\rtData.zip), "sampleInfo filename:" (C:\PepperExample\smallSample.csv), "globalID filename:" (C:\PepperExample\warmixGlobalID.txt), "accurateMass filename:" (empty), and "profitCoef filename:" (empty). A "Bootstrap:" dropdown is set to "No". At the bottom of the form, there are "run", "reset", and "help" buttons. The "run" button is circled in red. The browser's status bar at the bottom shows "Done" and an "Open Notebook" button.

- Select the proper file for each option. The files are contained in the PepperExample.zip file.
- Note: there are two sampleInfo files – small and large. The small file will process 2 experiments while the large sample will take longer and process 10 experiments.
- After filling out the form select **run**.

LandmarkMatching Step 3: Processing...

The screenshot shows the GenePattern web interface in a Mozilla Firefox browser window. The browser title is "Running Task - Mozilla Firefox". The address bar is empty. The page header includes the GenePattern logo, a "sign out vfusaro@broad.mit.edu" link, and an "about" link. Below the header, there are dropdown menus for "pipeline" and "task", a "search" button, and a "Filter by Suite" link. A "stop..." button is visible. The main content area displays the status of a running task: "Running [LandmarkMatching](#) as job # [164](#) on Wed Oct 18 15:03:59 EDT 2006". Below this, the task parameters are listed: "LandmarkMatching (peakList filename = [peakData.zip](#) , retentionTime filename = [rtData.zip](#) , sampleInfo filename = [smallSample.csv](#) , globalID filename = [varmixGlobalID.txt](#) , accurateMass filename = , prefitCoef filename = , Bootstrap = No)". At the bottom of the browser window, a status bar shows "Transferring data from cp995-9d5.broad.mit.edu..." and an "Open Notebook" button.

- After pressing **run** GenePattern will begin to process the data. For the small sample experiment (2 experiments) it will typically take about 10 minutes to finish.
- The more experiments the longer the computational time required to process the data.

LandmarkMatching Step 4: Results and output

The screenshot shows a Mozilla Firefox browser window titled "Running Task - Mozilla Firefox". The address bar is empty. The page content includes the GenePattern logo and navigation links for "sign out vfusaro@broad.mit.edu" and "about". There are input fields for "pipeline" (containing "pipeline") and "task" (containing "task"), along with "run" and "edit" buttons. A search bar is also present. Below the input fields is a "stop..." button and an "email notification to:" field. The main content area displays the following text:

Running [LandmarkMatching](#) as job # [164](#) on Wed Oct 18 15:03:59 EDT 2006

LandmarkMatching (peakList filename = [peakData.zip](#) , retentionTime filename = [rtData.zip](#) , sampleInfo filename = [smallSample.csv](#) , globalID filename = [varmixGlobalID.txt](#) , accurateMass filename = , prefitCoef filename = , Bootstrap = No)

- [MATCH_DATA](#)
- [PEAK_LIST](#)
- [RT_DATA](#)
- [match_que.pl](#)
- [DUMP_from_mass_calibrator.txt](#)
- [LMOutput.zip](#) ●
- [stdout.txt](#)
- [stderr.txt](#) ●
- [gp_task_execution_log.txt](#)

Annotations on the right side of the screenshot:

- A line connects the [LMOutput.zip](#) checkbox to the text: "The main output is a zipped directory of all the processed files. This can be used as input into the PeakMatch module."
- A line connects the [stderr.txt](#) checkbox to the text: "It is a good idea to check the error log to make sure that everything was processed correctly."

At the bottom of the page, there are buttons for "download selected results", "check all", "uncheck all", and "delete selected results". The status bar at the bottom left shows "Done" and the bottom right shows "Open Notebook".

PeakMatch Step 1: Select the task

The screenshot shows the GenePattern web interface in a Mozilla Firefox browser window. The page title is "run PeakMatch - Mozilla Firefox". The browser's address bar is empty, and the search bar contains "Google". The GenePattern logo is visible in the top left, and the user is logged in as "v fusaro@broad.mit.edu". The main content area displays the "PeakMatch" task configuration page. The page has a purple header with the GenePattern logo and user information. Below the header, there is a search bar and a "PeakMatch" dropdown menu. The main content area is titled "PeakMatch version 1" and contains several input fields and buttons. The input fields are: "peakList filename:" (required), "sampleInfo filename:" (required), "LandmarkMatchOutput filename:", "MZ tolerance:", "RT tolerance:", "outputName:" (required), and "numberProcesses:". Each input field has a "Browse..." button next to it. The "MZ tolerance:" field has a description: "m/z tolerance (ppm). Used if landmark match output is not supplied. Defaults to 10 ppm." The "RT tolerance:" field has a description: "retention time tolerance (min). Used if landmark match is not supplied. Defaults to 2 min." The "outputName:" field has a description: "File name prefix for the output files". The "numberProcesses:" field has a description: "Number of processes when running in parallel. Defaults to 1 (sequential)". At the bottom of the form, there are three buttons: "run", "reset", and "help". The browser's status bar at the bottom shows "Done" and "Open Notebook".

peakList filename: * Browse...
Zip file containing peak lists

sampleInfo filename: * Browse...
CSV file with header: experiment, sample, class

LandmarkMatchOutput filename: Browse...
Zip file containing the output from landmark matching.

MZ tolerance: m/z tolerance (ppm). Used if landmark match output is not supplied. Defaults to 10 ppm.

RT tolerance: retention time tolerance (min). Used if landmark match is not supplied. Defaults to 2 min.

outputName: * File name prefix for the output files

numberProcesses: Number of processes when running in parallel. Defaults to 1 (sequential).

- Select the PeakMatch task
- The **bold** options are required inputs

PeakMatch Step 2: Fill out the form

The screenshot shows the PeakMatch web interface in Mozilla Firefox. The browser title is "run PeakMatch - Mozilla Firefox". The interface includes a menu bar (File, Edit, View, Go, Bookmarks, Tools, Help), a search bar, and a navigation bar with "GenePattern" and "sign out vfusaro@broad.mit.edu about". The main content area displays the "PeakMatch" pipeline configuration form. The form includes the following fields and options:

- pipeline:** PeakMatch (selected from a dropdown)
- peakList filename: *** C:\PepperExample\peakData.zip (with a "Browse..." button). Description: Zip file containing peak lists.
- sampleInfo filename: *** C:\PepperExample\smallSample.csv (with a "Browse..." button). Description: CSV file with header: experiment, sample, class.
- LandmarkMatchOutput filename:** C:\PepperExample\Output\LandmarkMatch\LMOutput.zip (with a "Browse..." button). Description: Zip file containing the output from landmark matching.
- MZ tolerance:** [] m/z tolerance (ppm). Used if landmark match output is not supplied. Defaults to 10 ppm.
- RT tolerance:** [] retention time tolerance (min). Used if landmark match is not supplied. Defaults to 2 min.
- outputName: *** PeakMatchOutput. Description: File name prefix for the output files.
- numberProcesses:** [] Number of processes when running in parallel. Defaults to 1 (sequential).

At the bottom of the form, there are three buttons: "run" (circled in red), "reset", and "help". The status bar at the bottom of the browser shows "Done" and "Open Notebook".

- Select the proper file for each option. The files are contained in the PepperExample.zip file.
- Note: PeakMatch is designed to run independently from LandmarkMatching and does not require the LandmarkMatching output. If it is included MZ and RT tolerances will automatically be calculated from the data.
- After filling out the form select **run**.

PeakMatch Step 3: Processing...

Running Task - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

Go

Google Search PageRank ABC Check AutoLink >>

GenePattern [sign out vfusaro@broad.mit.edu](#) [about](#)

pipeline task run edit Filter by Suite search

stop...

email notification to:

Running [PeakMatch](#) as job # [165](#) on Wed Oct 18 15:24:27 EDT 2006

PeakMatch (peakList filename = [peakData.zip](#) , sampleInfo filename = [smallSample.csv](#) ,
LandmarkMatchOutput filename = [LMOutput.zip](#) , MZ tolerance = , RT tolerance = , outputName =
PeakMatchOutput , numberProcesses =)

Transferring data from cp995-9d5.broad.mit.edu... Open Notebook

- After pressing **run** GenePattern will begin to process the data. For the small sample experiment (2 experiments) it will typically take about 10 minutes to finish.
- The more experiments the longer the computational time required to process the data.

PeakMatch Step 4: Results and output



The screenshot shows the GenePattern web interface in a Mozilla Firefox browser window. The page title is "GenePattern" and the user is logged in as "v fusaro@broad.mit.edu". The interface includes a search bar, a "pipeline" dropdown menu, and a "task" dropdown menu. The "task" dropdown is set to "PeakMatch". The "run" button is highlighted, and the "stop..." button is visible below it.

email notification to:

Running [PeakMatch](#) as job # [165](#) on Wed Oct 18 15:24:27 EDT 2006

PeakMatch (peakList filename = [peakData.zip](#) , sampleInfo filename = [smallSample.csv](#) , LandmarkMatchOutput filename = [LMOutput.zip](#) , MZ tolerance = , RT tolerance = , outputName = PeakMatchOutput , numberProcesses =)

- [PEAK_LIST](#)
- [run-peak-match.r](#)
- [PeakMatchOutput-matched.gct](#) ●
- [MATCH_DATA](#)
- [PeakMatchOutput-matched.cls](#) ●
- [run-peak-match.r.Rout](#)
- [stdout.txt](#) ●
- [gp_task_execution_log.txt](#)

The main output is a gct file that can be used in other GenePattern modules for further analysis.

The corresponding class file is also output. This corresponds to the information that was listed in the sampleInfo file.

It is a good idea to check the error log to make sure that everything was processed correctly.

[check all](#) [uncheck all](#)

Done

Troubleshooting

- For more information about each module please see the PDF documentation included with each module. The documentation will describe the necessary file formats.
- We are currently only supporting LSF clusters. If there is interest in getting PEPPeR to work with a different cluster we can help you modify the code.