

# Integrative Genomic Analysis with



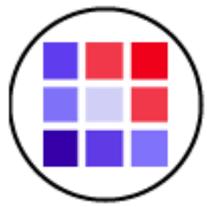
# GenePattern

**January 23, 2018**

**Edwin Juarez   Ted Liefeld   Michael Reich**

# Agenda

- Introduction and Overview
- GenePattern Notebook Features
- Machine Learning in GenePattern Notebook:
  1. Differential Gene Expression Analysis
  2. Clustering
  3. Classification/Prediction
- Other GenePattern features
- Open Q&A



# GenePattern Overview

# Tools for Bioinformatics



Best-Practices Documentation Blog Forum Events Download

Search

[Back to Tool Docs Index](#)

## MuTect2

Call somatic SNPs and indels via local re-assembly of haplotypes

## HISAT2

graph-based alignment of next generation sequencing reads to a population of genomes

## Samtools

Samtools is a suite of programs for interacting with high-throughput sequencing data. It consists of three separate repositories:

Samtools	Reading/writing/indexing/viewing SAM/BAM/CRAM format
BCFtools	Reading/writing BCF2/VCF/gVCF files and calling/filtering/summarising SNP and short indel sequence variants
HTSlib	A C library for reading/writing high-throughput sequencing data

Samtools and BCFtools both use HTSlib internally, but these source packages contain their own copies of htseq so they can be built independently.



**Bowtie 2**  
Fast and sensitive read alignment



Home Installation Documentation Examples

## 1.4. Support Vector Machines

## Principal Component Analysis

## Picard

build passing

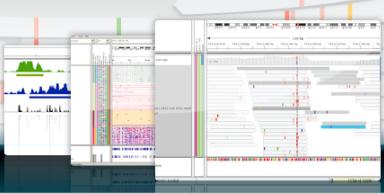
A set of command line tools (in Java) for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF.

## Hierarchical Clustering / Dendrograms

## Cufflinks

Transcriptome assembly and differential expression analysis for RNA-Seq.

## Integrative Genomics Viewer



## Burrows-Wheeler Aligner

### Introduction

BWA is a software package for mapping low-divergent sequences against a large reference genome, such as the human genome. It consists of three algorithms:

## NMF: Non-negative Matrix Factorization

### What is HAPSEG?

HAPSEG is a probabilistic method to interpret bi-allelic marker data in cancer samples.

## MAGECK

Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout

### What is RNA-SeQC?

RNA-SeQC is a java program which computes a series of quality control metrics for RNA-seq data.



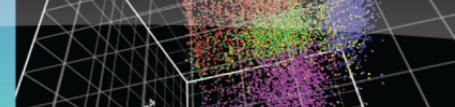
Network Data Integration, Analysis, and Visualization in a Box

## Trimmomatic: A flexible read trimming tool for Illumina NGS data



**MSigDB**  
Molecular Signatures Database

**FLAME**  
Flow analysis with Automated Multivariate Estimation



# Problems with bioinformatics tool use and interoperability

- Tools are built using different languages and with different architectural assumptions.
- Each tool has its own installation and operational requirements.
- Tools require (sometimes extensive) Unix knowledge.
- Tools are not designed to communicate with each other.

```
bowtie -a --best --strata -S -m 100 -X 400 --chunkmb 256 --fullref -p 4
Dmel.BDGP5-transcripts \ -1 SRR031714_1.fastq -2 SRR031714_2.fastq |
samtools view -F 0xC -bs - | \ samtools sort -n - ~/Desktop/untreated3-
transcriptome
```

# Solution features: Wrapping tools

- “Wrap” tools in a web-based interface
- No installation/running requirements
- No programming required
- Fill out required parameters and provide input files

The screenshot shows a web browser window titled "GenePattern - Bowtie.aligner". The URL is <https://genepattern.broadinstitute.org/gp/pages/index.jsf?lsid=urn:lsid:broad.mit.edu:cancer.software.genepattern.module.analysis.Bowtie.aligner>. The page displays the "Bowtie.aligner" module, version 4. It provides a user interface for aligning short reads using Bowtie 2. The form includes fields for specifying a prebuilt or custom Bowtie index, input format (FASTA or FASTQ), reads pairs, quality value scale (Phred), integer quality values (no), and maximum reads to align. There are also "Batch" checkboxes and "Run" and "Reset" buttons.

**Bowtie.aligner** version 4

Bowtie2 (v. 2.1.0) is an ultrafast and memory-efficient short read aligner.

\* required field

**prebuilt bowtie index**

A prebuilt Bowtie 2 index. Either a prebuilt or a custom Bowtie 2 index must be specified.

**custom bowtie index**

Upload File... Add Path or URL... Drag Files Here

2GB file upload limit using the Upload File... button. For files > 2GB upload from the Files tab.

**input format\***

The format of the reads input files.

**reads pair 1\***

Unpaired reads file or first mate for paired reads. A file or zip of files containing reads in FASTA or FASTQ format (can be compressed - ie .gz).

**reads pair 2**

Second mate for paired reads. A file or zip of files in FASTA or FASTQ format (can be compressed - ie .gz).

**quality value scale**

Phred

**integer quality values**

no

Whether the quality values are represented as space-separated ASCII integers (i.e 40 40 30 40 ..).

**max reads to align**

Batch Documentation Run Reset

# Solution features: Tool Repository

- Collection of hundreds of wrapped tools
- Gene expression, sequence variation, proteomics, network analysis, machine learning, flow cytometry, etc.
- Searchable by name, keyword, etc.
- Widely-used community tools, lab-developed tools, utilities
- Users can contribute their own tools

AddNoiseToFCS	Add noise to specified parameters in an FCS data file. <a href="#">Flow Cytometry</a>		AffySTExpressionFileCreator [Beta Release** Contact gp-help with any issues. Check stdout.txt and stderr.txt for errors]... <a href="#">Preprocess &amp; Utilities</a>	
ApplyGatingML	Apply a Gating-ML file on an FCS data file (gate and/or transform list mode data) <a href="#">Flow Cytometry</a>		ARACNE Runs the ARACNE algorithm for reverse engineering cellular networks <a href="#">Pathway Analysis</a>	
AreaChange	Calculates fraction of area under the spectrum that is attributable to signal (area after noise)... <a href="#">Proteomics</a> , <a href="#">ProteomicsSuite</a>		Arff2Gct Convert an .arff file into a gene pattern .gct / .cls file pair <a href="#">Multi-label Protein Prediction Suite (MiPPS)</a> , <a href="#">Preprocess ...</a>	
ATARI	Runs ATARI on RNAi reagent-level data. <a href="#">RNAi</a>		AuDIT Automated Detection of Inaccurate and Imprecise Transitions in MRM Mass Spectrometry <a href="#">Proteomics</a>	
BedToGtf	Converts BED files to GFF or GTF format <a href="#">Data Format Conversion</a>		Beroukhim.Getz.2007.PNAS.Glioma.... pipeline	
BlastTrainTest	Sequence similarity classification using BLAST <a href="#">Multi-label Protein Prediction Suite (MiPPS)</a> , <a href="#">Prediction</a>		BlastXValidation Sequence similarity cross validation prediction using BLAST <a href="#">Multi-label Protein Prediction Suite (MiPPS)</a> , <a href="#">Prediction</a>	
Bowtie.aligner	Bowtie2 (v. 2.1.0) is an ultrafast and memory-efficient short read aligner. <a href="#">RNA-seq</a>		Bowtie.indexer Builds a Bowtie2 (v. 2.1.0) index from a set of DNA sequences <a href="#">RNA-seq</a>	
BWA.aln	[**Beta Release** Contact gp-help with any issues. Check stdout.txt and stderr.txt for errors]... <a href="#">RNA-seq</a>		BWA.bwasw [**Beta Release** Contact gp-help with any issues. Check stdout.txt and stderr.txt for errors]... <a href="#">RNA-seq</a>	
BWA.indexer	[**Beta Release** Contact gp-help with any issues. Check stdout.txt and stderr.txt for errors]... <a href="#">RNA-seq</a>		CaArray2ImportViewer Imports data files from CaArray 2.4.1 and creates gct or cls files <a href="#">Visualizer</a>	
CART	Classification and Regression Tree <a href="#">Prediction</a>		CARTXValidation Classification and Regression Tree Cross-Validation <a href="#">Prediction</a>	
CBS	Segments DNA copy number data into regions of estimated equal copy number using circular binary... <a href="#">SNP Analysis</a>		ChIPSeq.CreateHeatmap Generates a heatmap based on the ChIP-Seq signal extracted from a BAM file, according to the...	

# Solution features: Reproducibility

- Record and replay of all analyses
- Retain all versions of code – so results can be reproduced even if code changes
- Chain analyses into “pipelines” that can be shared and published

# What can you do with GenePattern

Function	Description
Differential Expression	Find the genes that distinguish between two conditions (i.e., tumor vs normal, relapse vs non-relapse, etc.)
Gene Set Enrichment Analysis	Find pathways of genes that are “enriched” between two conditions
Clustering	Find subsets of a dataset (e.g., genes, samples) that are similar to one another in structure or function
Classification	Create a model that will predict the class of an unknown sample (e.g., relapse, non-relapse, etc.)
Sequence Variation analysis	Find chromosomal regions of similar copy number, call genotypes, etc.
Disease-specific analyses	E.g. genomic identification of significant targets in cancer (GISTIC), identify loss of homozygosity, etc.
Dimension Reduction	Transform your data into a lower number of dimensions to facilitate analysis

# GenePattern vocabulary: Modules

Copy Number  
Divide  
by Normals

GSEA

Variation  
Filter

GISTIC

CBS

k-Nearest  
Neighbors

Classification  
and  
Regression Trees

Support  
Vector  
Machines

Hierarchical  
Clustering

GISTIC

Expression  
File  
Creator

Metagene  
Projection

# GenePattern vocabulary: Modules

Copy Number  
Divide  
by Normals

GSEA

Variation  
Filter

GISTIC

CBS

k-Nearest  
Neighbors

Classification  
and  
Regression Trees

Support  
Vector  
Machines

**Hierarchical  
Clustering**

GISTIC

Expression  
File  
Creator

Metagene  
Projection

## Hierarchical Clustering

### Files

HCL.jar  
cluster.exe  
ant.jar  
gp-modules.jar  
Jama-1.0.2.jar

### Documentation

HierarchicalClustering.pdf

### Parameter descriptions

```
-f <input.filename>
    <log.transform>
        <row.center>
        <row.normalize>
        <column.center>
        <column.normalize>
    -u <output.base.name>
    -e <column.distance.measure>
    -g <row.distance.measure>
    -m <clustering.method>
```

# >250 GenePattern Modules, 10/2017

The screenshot shows the GenePattern web application interface. At the top, there's a navigation bar with links for 'Modules & Pipelines', 'Suites', 'Job Results', 'Resources', 'Help', and 'GenomeSpace'. Below the navigation bar, there are tabs for 'Modules', 'Jobs', and 'Files', with 'Modules' currently selected. A search bar says 'Search Modules & Pipelines' and a button says 'Browse Modules >'. A section titled 'Favorite Modules' contains a dashed box labeled 'Drag Modules Here'. Another section titled 'Recent Modules' lists 'ComparativeMarkerSelectionViewer', 'FeatureSummaryViewer', 'HeatMapView', and 'HierarchicalClusteringViewer'. The main content area is titled 'Browse Modules > All Modules' and displays a grid of 25 module cards. A red box highlights the first two columns of the grid. Each card includes the module name, a brief description, and a 'SNP Analysis' tag. The modules listed include ABSOLUTE, ABSOLUTE.review, ABSOLUTE.summarize, AddFCSEventIndex, AddFCSParameter, AddNoiseToFCS, aml.all.pipeline, ApplyGatingML, ARACNE, AreaChange, Arff2Gct, ATARIS, AuDIT, BedToGtf, Beroukhim.Getz.2007.PNAS.Glioma.GI pipeline, Birdseed, BirdseedCallRate, BirdseedDataPreparation, BlastTrainTest, BlastXValidation, Bowtie.aligner, Bowtie.indexer, BWA.aln, and BWA.bwasw.

**Browse Modules > All Modules**

ABSOLUTE	Extracts absolute copy numbers per cancer cell from a mixed DNA population. Use this module... SNP Analysis	ABSOLUTE.review	Extracts the absolute copy number per cancer cell from a mixed DNA population. Use this module... SNP Analysis
ABSOLUTE.summarize	Summarizes the results from multiple ABSOLUTE runs so that an analyst can manually review the solutions. SNP Analysis	AddFCSEventIndex	Adds indexes to events in a Flow Cytometry Standard (FCS) data file. Flow Cytometry
AddFCSParameter	Add parameters and their values to a FCS data file Flow Cytometry	AddNoiseToFCS	Add noise to specified parameters in an FCS data file. Flow Cytometry
aml.all.pipeline	ALL/AML methodology, from Golub and Slonim et al., 1999 pipeline	ApplyGatingML	Apply a Gating-ML file on an FCS data file (gate and/or transform list mode data) Flow Cytometry
ARACNE	Runs the ARACNE algorithm for reverse engineering cellular networks Pathway Analysis	AreaChange	Calculates fraction of area under the spectrum that is attributable to signal (area after noise)... Proteomics, ProteomicsSuite
Arff2Gct	Convert an .arff file into a gene pattern .gct / .cls file pair Multi-label Protein Prediction Suite (MiPPS), Preprocess ...	ATARIS	Runs ATARIS on RNAi reagent-level data RNAi
AuDIT	Automated Detection of Inaccurate and Imprecise Transitions in MRM Mass Spectrometry Proteomics	BedToGtf	Converts BED files to GFF or GTF format Data Format Conversion
Beroukhim.Getz.2007.PNAS.Glioma.GI pipeline		Birdseed	SNP genotyping algorithm that runs on the Affymetrix 500K, SNP5.0, and SNP6.0 platforms SNP Analysis
BirdseedCallRate	Computes the call rate of the Birdseed algorithm SNP Analysis	BirdseedDataPreparation	Prepare a bspn file for running Birdseed SNP Analysis
BlastTrainTest	Sequence similarity classification using BLAST Multi-label Protein Prediction Suite (MiPPS), Prediction	BlastXValidation	Sequence similarity cross validation prediction using BLAST Multi-label Protein Prediction Suite (MiPPS), Prediction
Bowtie.aligner	Bowtie2 (v. 2.1.0) is an ultrafast and memory-efficient short read aligner. RNA-seq	Bowtie.indexer	Builds a Bowtie2 (v. 2.1.0) index from a set of DNA sequences RNA-seq
BWA.aln		BWA.bwasw	

About GenePattern | Contact Us



# GenePattern Notebook Environment

The screenshot displays the GenePattern Notebook environment. At the top, there is a navigation bar with the GenePattern Notebook logo, the title "GenePattern Notebook Workshop (autosaved)", and links for "Control Panel" and "Logout tabor@broadinstitute.org". Below the navigation bar is a toolbar with various icons for file operations like opening, saving, and printing, as well as a "Code" button.

The main content area features a "GenePattern Notebook Tutorial" section. It includes a brief introduction: "If you are new to the GenePattern Notebook environment, this tutorial should help familiarize you with some of its most important features." and a call-to-action: "Click below to begin a quick tour of interface. Read onward for an overview of other features." A blue button labeled "Click here for a tour!" is present.

Below the tutorial, a note states: "Below is an example GenePattern authentication cell." This is followed by a login form titled "GenePattern Login". The form has fields for "GenePattern Server" (set to "Broad Institute"), "GenePattern Username" (set to "tabor@broadinstitute.org"), and "GenePattern Password" (with a masked input field). At the bottom of the form are two buttons: "Log into GenePattern" and "Register an Account".



# Jupyter Notebook

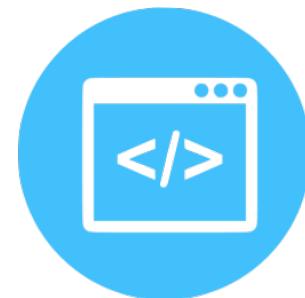
- Popular and well-supported framework for scientific computing
- Ecosystem of available extensions and resources
- Open source

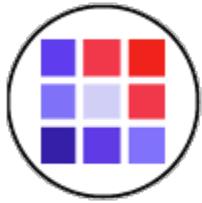




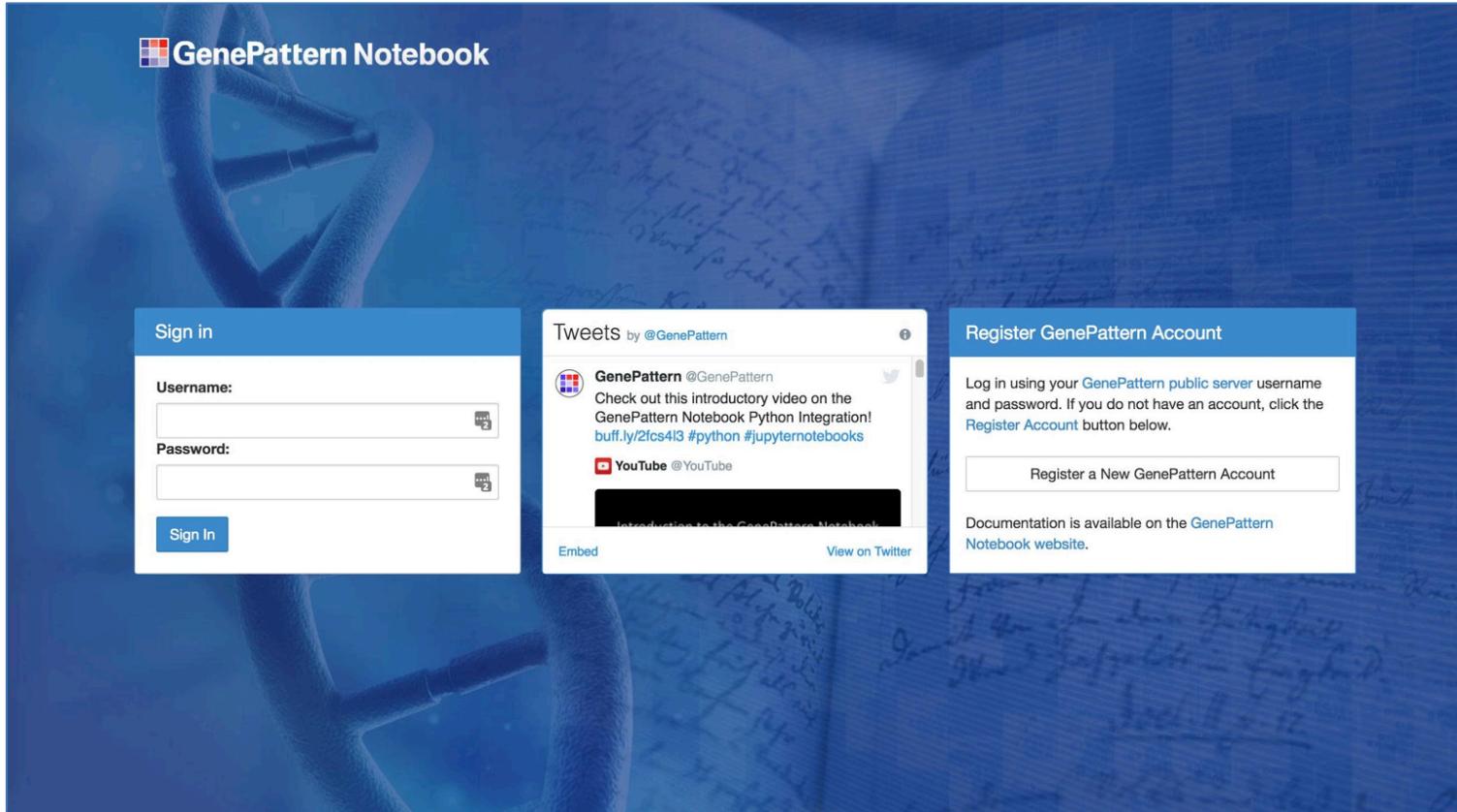
# Complete Research Narrative

- Leverages the best of Jupyter and GenePattern
- Interleave text, visualization, graphics and analytical aspects

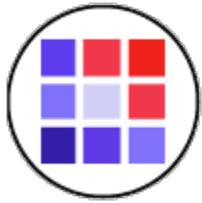




# GenePattern Notebook Repository



- <https://notebook.genepattern.org>

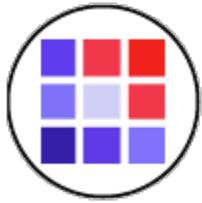


# Notebook Workspace

- Lists your private notebooks and associated files.
- Copy, move, rename, delete, download and publish notebooks from here.

The screenshot shows the GenePattern Notebook interface. At the top, there's a navigation bar with tabs for "Files" (which is selected), "Running", and "Public Notebooks". On the right side of the bar are "Control Panel" and "Logout genepattern" buttons. Below the navigation bar, a message says "Select items to perform actions on them." A file list table follows:

		Name	Last Modified
<input type="checkbox"/>		GenePattern Notebook Tutorial.ipynb	seconds ago
<input type="checkbox"/>		GenePattern Python Tutorial.ipynb	seconds ago

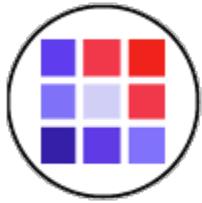


# Browse Public Notebooks

- A variety of public notebooks are available in the GenePattern Notebook Repository.
- Anyone can make a copy of these notebooks to read, run and reproduce.

The screenshot shows the GenePattern Notebook interface. At the top, there is a navigation bar with tabs for "Files", "Running", and "Public Notebooks". The "Public Notebooks" tab is currently selected. On the left, there is a sidebar titled "Public Notebooks" with a search bar labeled "Search Repository". The main content area displays a list of public notebooks:

Notes	Title	Author
	GenePattern Notebook Tutorial A tutorial on how to use the GenePattern Notebook environment.	Thorin Tabor
	GenePattern Python Tutorial A short tutorial on how to use the programmatic GenePattern client for Python.	Thorin Tabor
	K-means Clustering Tutorial on K-means clustering in GenePattern	GenePattern Team
	GenePattern Files in Python How to use GenePattern files with scipy, pandas and matplotlib.	Thorin Tabor
	Classification and Prediction Classification and prediction analysis in GenePattern.	GenePattern Team



# Run an Analysis Notebook

2018-01-23\_01\_UBIC GenePattern Notebook Introduction

The screenshot shows the GenePattern Notebook environment. At the top, there's a header bar with the GenePattern logo, the title "2017-11-07\_02\_CCMI\_GenePattern\_Notebook", the last checkpoint time ("Last Checkpoint: 17 hours ago (unsaved changes)"), and user information ("Logout tabor@broadinstitute.org"). Below the header is a toolbar with various icons for file operations, cell selection, and tools.

The main content area displays a document titled "Introduction to GenePattern Notebook". The text in this document reads:

This document should help you understand how to run an analysis in the GenePattern Notebook environment. In it you will perform a simple preprocessing step and then view the results in a heat map. Instructions to follow are given in blue boxes, such as with the one below.

A blue box contains the text: "Sign in to GenePattern by clicking the login button or entering your username and password into the form below."

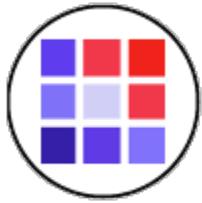
Below the document, a "Login" dialog box is overlaid. It has fields for "GenePattern Server" (set to "Broad Institute"), "GenePattern Username" (set to "tabor@broadinstitute.org"), and "GenePattern Password" (with a masked value). A message in the dialog says: "You have already authenticated with the GenePattern Public Server. Would you like to automatically sign in now?". It includes two buttons: "Login as tabor@broadinstitute.org" and "Cancel".

At the bottom of the page, there's a section titled "Step 1: PreprocessDataset" with a descriptive text about preprocessing gene expression data.



# Understanding Notebooks

- Notebooks encapsulate a workflow, including analysis, documentation and other considerations, so that it can be easily reproduced.
- To achieve this, all notebooks are backed by a “kernel,” which is a complete contained computational environment.
- The kernel provides programmatic capabilities for users who want to code, and also allows for interactive widgets for users who don’t want to code.



# Name, Save & Checkpoint Notebooks

- Name or rename notebooks
- Save or revert to a checkpoint
- Make a duplicate notebook

The image displays two side-by-side screenshots of the GenePattern Notebook software interface. Both screenshots show the 'File' menu open, revealing options such as 'New Notebook', 'Open...', 'Make a Copy...', 'Rename...', 'Save and Checkpoint', 'Revert to Checkpoint', 'Print Preview', 'Download as', 'Trusted Notebook', and 'Close and Halt'. The left screenshot shows the notebook title as 'Untitled3' and the right one as 'Untitled4'. Both screenshots also show the same content in the main workspace, which includes a section titled 'ustering Expression Files' with a date of 'Friday, September 30, 2016 11:53 AM' and a description of the notebook's purpose.



# GenePattern Cells

Auth  
Cell

**GenePattern Login**

**GenePattern Server**  
Broad Institute

**GenePattern Username**  
Username

**GenePattern Password**  
Password

[Log into GenePattern](#) [Register an Account](#)

Analysis  
Cell

**GenePattern ConvertLineEndings**

Converts line endings to the host operating system's format.

\* Required Field [Run](#)

**input.filename\***  [Upload File...](#) [Add Path or URL...](#) **Drag Files Here**  
2GB file upload limit using the Upload File... button.

The input file (any non-binary file format)

**output.file\***  <input.filename\_basename>.cvt.<input.filename\_extension>

The output file

\* Required Field [Run](#)

Job  
Cell

**GenePattern 1251770. ConvertLineEndings**

Submitted by tabor on 2016-03-03T12:09:39-05:00

[test.cvt.txt](#) [gp\\_execution\\_log.txt](#)

[Completed](#)



# Authentication Cells

**GenePattern Login**

**GenePattern Server**  
Broad Institute

**GenePattern Username**  
Username

**GenePattern Password**  
Password

**Log into GenePattern**   **Register an Account**

**GenePattern tabor** <http://genepattern.broadinstitute.org/gp>

-- Sun 5:00 pm -- Update: The job queue is back online and accepting new jobs. For best results you should cancel any jobs which you had started before today at 5:00 pm. We can not make any guarantees about results obtained for jobs that had not yet completed before the start of the maintenance window. Thanks, The GenePattern Team -- Sat 5:00 pm -- Update: The job queue is not yet ready to accept new jobs. Please refrain from starting new jobs until further notice. We expect it to be ready during the day Sunday. Thanks, The GenePattern Team Important message: The GenePattern Server will go offline for quarterly maintenance just before 8:00 am, Saturday March 5. We expect the maintenance to last the majority of the day. Thanks, The GenePattern Team -- March 7 -- New Blog Post: Older Java Applet Visualizers Blocked by Default in Updated Firefox Older Java Applet visualizers are no longer supported in Chrome. Please read our blog post for more information.

Experiencing a bug? Have thoughts on how to make GenePattern Notebook better? Let us know by leaving feedback.

**Leave Feedback**



# Analysis Cells

GenePattern ExtractComparativeMarkerResults Version 4 ? - >\_

Creates a derived dataset and feature list file from the results of ComparativeMarkerSelection

\* Required Field Run

**comparative.marker.filename\***  Upload File... Add Path or URL... Drag Files Here  
2GB file upload limit using the Upload File... button.  
The results from ComparativeMarkerSelection - .odf

**dataset.filename\***  Upload File... Add Path or URL... Drag Files Here  
2GB file upload limit using the Upload File... button.  
The dataset file used to select markers - .gct, .res, Dataset

**statistic**   
The statistic to filter features on

**min**   Select features with statistic  $\geq$  min

**max**   Select features with statistic  $\leq$  max

**number.of.neighbors**   
Number of neighbors to select by score in each direction

**base.output.name\***  <comparative.marker.selection.filename\_basename>.filt  
The base name for the output files

\* Required Field Run



# Job/Result Cells

 GenePattern 1251770. ConvertLineEndings ⟳ ⟲ ⟳ ⟶ ⟷

GenePattern 129041. HeatMapView

Submitted by tabor on 2016-03-11T13:18:04-05:00

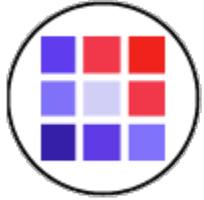
Completed

gp\_execution\_log.txt

Columns: Samples

Rows: Feature Name

The heatmap displays gene expression levels for approximately 20 samples (rows) and 20 features (columns). The samples include various cell types (e.g., T-cell, B-cell, ALL, AML) and tissues (e.g., BioB, BioC, BioDn, CreX, DapX, LysX, PheX). The features represent different genes or probes. The color scale ranges from blue (low expression) to red (high expression), with white indicating intermediate values. The heatmap shows distinct patterns of gene expression across the different samples and features.

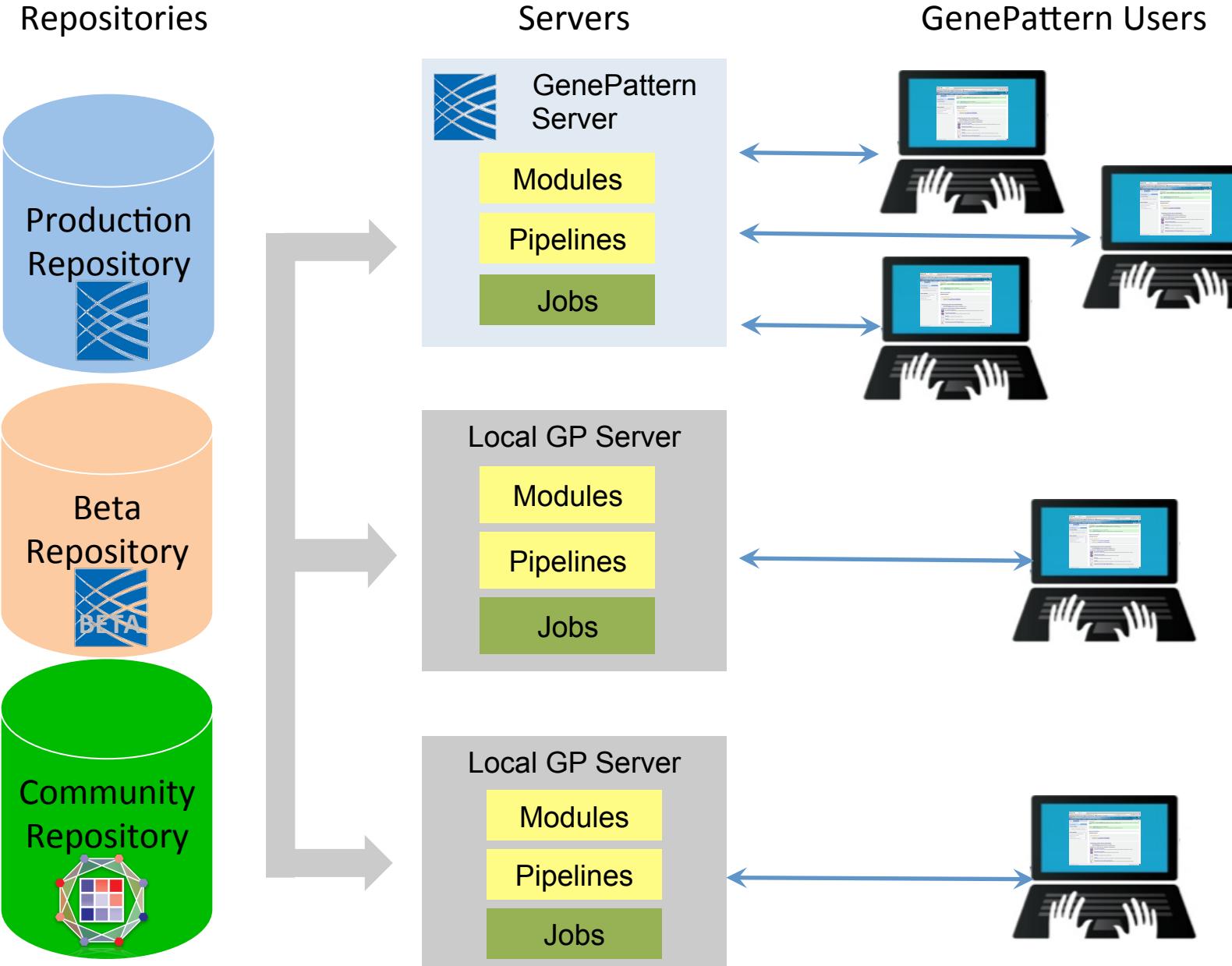


# Running GenePattern Notebook

- Run using the GenePattern Notebook Repository.
- Install on your own computer by installing through the pip or conda package repositories.
- A GenePattern Notebook Docker image is available.

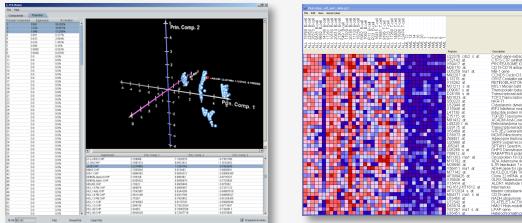
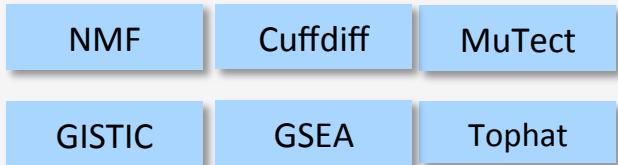


# How GenePattern works

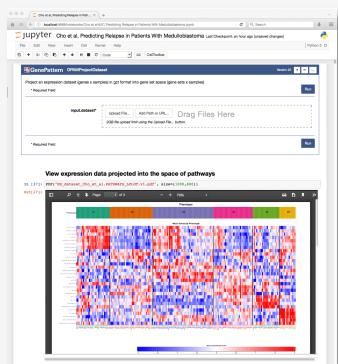


# The GenePattern Ecosystem: Architecture

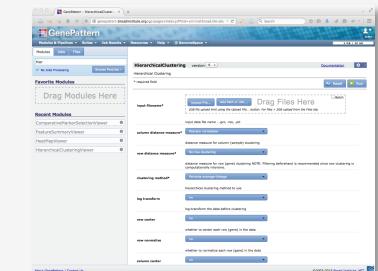
## Module Repository



## Clients



Notebook



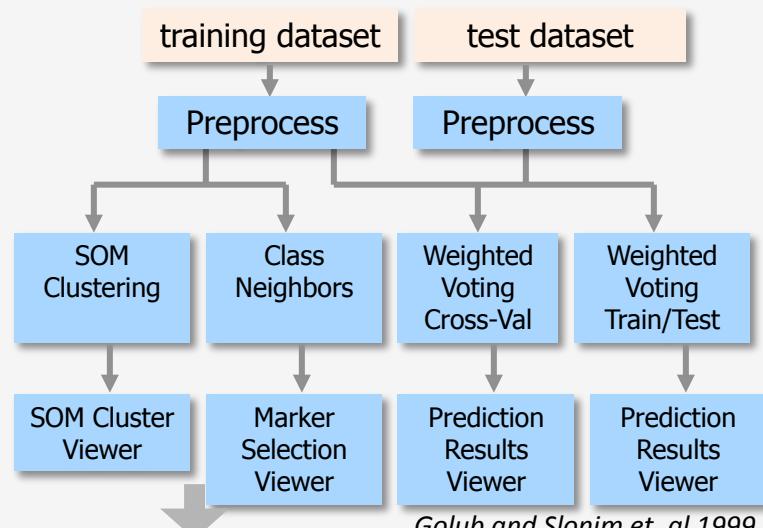
Web

## Analysis Engine

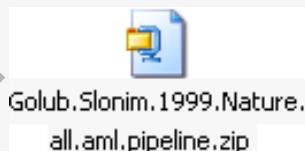
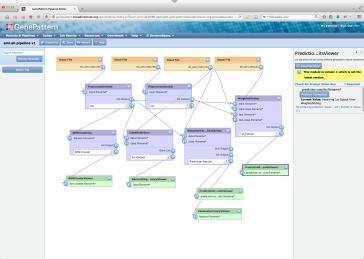


Record/replay analyses  
Versioning of methods  
Web service access

## Pipeline Environment

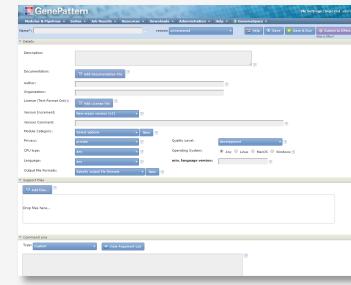


Golub and Slonim et. al 1999

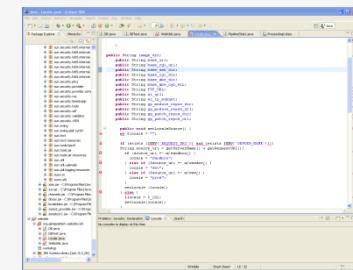


Support for *in silico* reproducible research

## Module Integrator



Easy addition of new tools

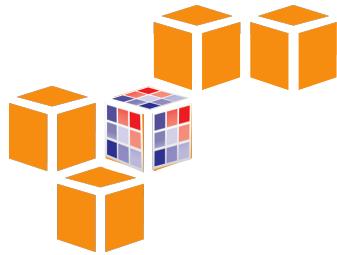


Programming  
Access for all levels of user

# Availability



- Public server at Broad Institute ([genepattern.broadinstitute.org](http://genepattern.broadinstitute.org))



- Public server running on Amazon Web Services (beta) ([gp-dev-ami.genepattern.org](http://gp-dev-ami.genepattern.org))



- Public server at Indiana U, backed by Mason HPC cluster ([gp.indiana.edu](http://gp.indiana.edu))



- Downloadable server (laptop to compute farm) ([www.genepattern.org](http://www.genepattern.org))

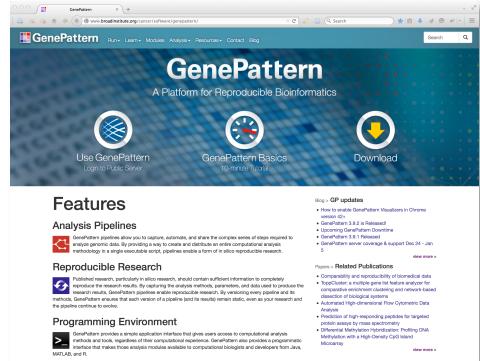


- Amazon Machine Image (AMI)

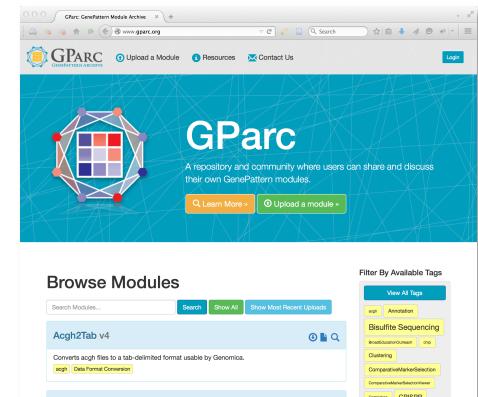
# Community Activity

- Current version: 3.9.10 (5/2017)
- >50,000 registered users
- Open source, BSD-style license
- Public server runs ~4,000 analyses/week
- GParc: GenePattern community repository
  - ~100 community-contributed methods
  - CRISPR analysis
  - Bisulfite sequencing
  - Flow cytometry
  - RNAi screens

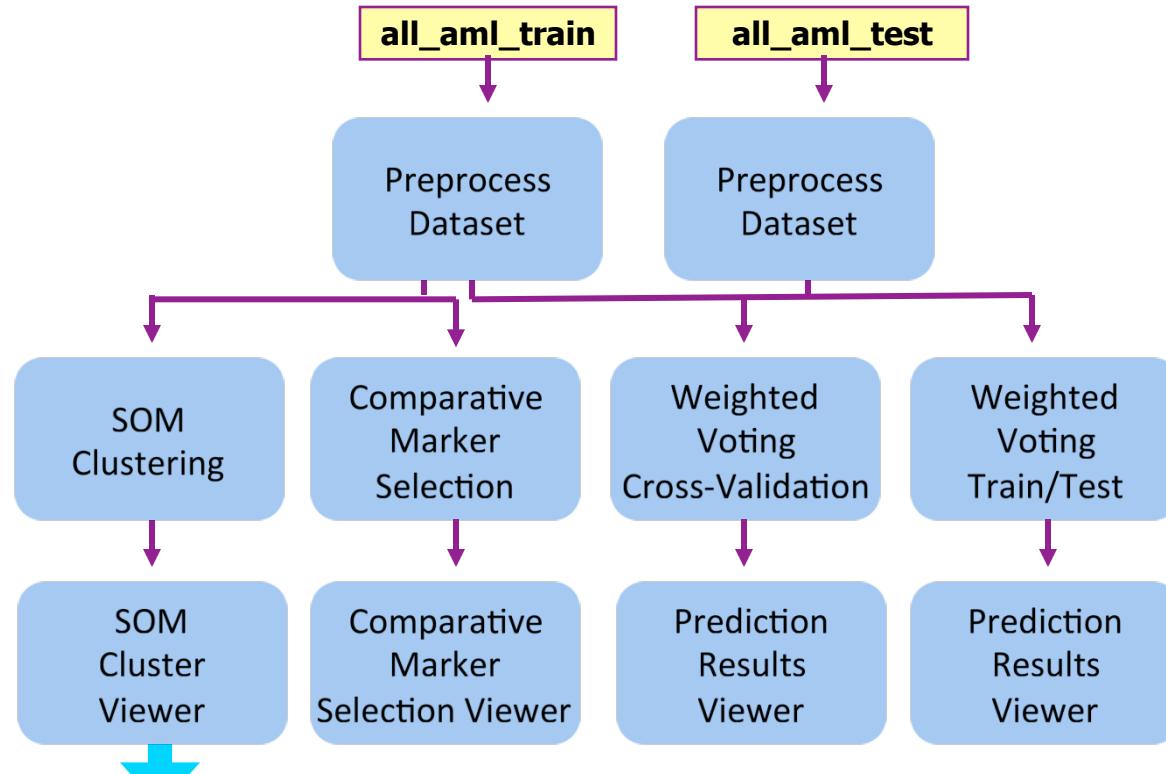
[www.genepattern.org](http://www.genepattern.org)



[www.gparc.org](http://www.gparc.org)



# GenePattern vocabulary: Pipelines



Golub.Slonim.1999.Nature.  
all.aml.pipeline.zip

# Pipelines in GenePattern

The screenshot shows the GenePattern web interface. A red box highlights the 'Browse Modules > pipeline' section. This section contains a grid of 20 pipeline entries, each with a title, a brief description, and a 'details' link.

Module Title	Description	Link
aml.all.pipeline	ALL/AML methodology, from Golub and Slonim et al., 1999 pipeline	<a href="#">details</a>
CBSWrapperPipeline	A one step pipeline that runs CBS pipeline	<a href="#">details</a>
CopyNumberInferencePipeline.Part2of:	A pipeline that runs CopyNumberInferencePipeline.Part2of2 – Part of the pipeline	<a href="#">details</a>
FLAMEContourViewer.Pipeline	Pipeline which runs the FLAMECounterDataGenerator and the FLAMEViewer pipeline	<a href="#">details</a>
Golub.Slonim.1999.Nature.all.aml.pipe	ALL/AML methodology, from Golub and Slonim et al., 1999 pipeline	<a href="#">details</a>
IlluminaDASLPipeline	creates a GenePattern gct file from raw Illumina scan data pipeline	<a href="#">details</a>
ImmPort_FLOCK_Individual_FCS	ImmPort FLOCK and Individual FCS pipeline	<a href="#">details</a>
job212786	describe it here pipeline	<a href="#">details</a>
job437446	describe it here pipeline	<a href="#">details</a>
MGED_Reich	test pipeline	<a href="#">details</a>
PWRGPTTestAuto_InheritType_Vis	Automated pipeline with file input as stored path (ie saved with the pipeline) and text inputs.... pipeline	<a href="#">details</a>
Rot13Madness		<a href="#">details</a>
Beroukhim.Getz.2007.PNAS.Glioma.GI!	pipeline	<a href="#">details</a>
CopyNumberInferencePipeline.Part2of:	Second half of Pipeline for processing SNP 6 data pipeline	<a href="#">details</a>
CufflinksCuffmergePipeline	**Beta Release** Contact gp-help with any issues. Check stdout.txt and stderr.txt for errors Creates... pipeline	<a href="#">details</a>
GetDataSetInSilico	downloads a compressed .tgz file from the Insilico servers and extract it pipeline	<a href="#">details</a>
Golub_Slonim	ALL/AML methodology, from Golub and Slonim et al., 1999 pipeline	<a href="#">details</a>
ImmPort_FLOCK_CrossSample	ImmPort FLOCK and CrossSample pipeline	<a href="#">details</a>
job212108	describe it here pipeline	<a href="#">details</a>
job298686	describe it here pipeline	<a href="#">details</a>
Lu.Getz.Miska.Nature.June.2005.mous	Normal/tumor classifier and kNN prediction of mouse lung samples LuGetzMiska.Nature.2005.Suite, pipeline	<a href="#">details</a>
ParallelICBS	Runs CBS algorithm on multiple samples in parallel	<a href="#">details</a>
RNaseQC_CEGS	pipeline	<a href="#">details</a>
ScripturePipeline		<a href="#">details</a>

Below the highlighted section, there is a green box containing release notes and a large empty area for future content.

[About GenePattern](#) | [Contact Us](#)

©2003-2014 Broad Institute, MIT