

# Integrative Genomic Analysis with



# GenePattern

**August 22, 2025**

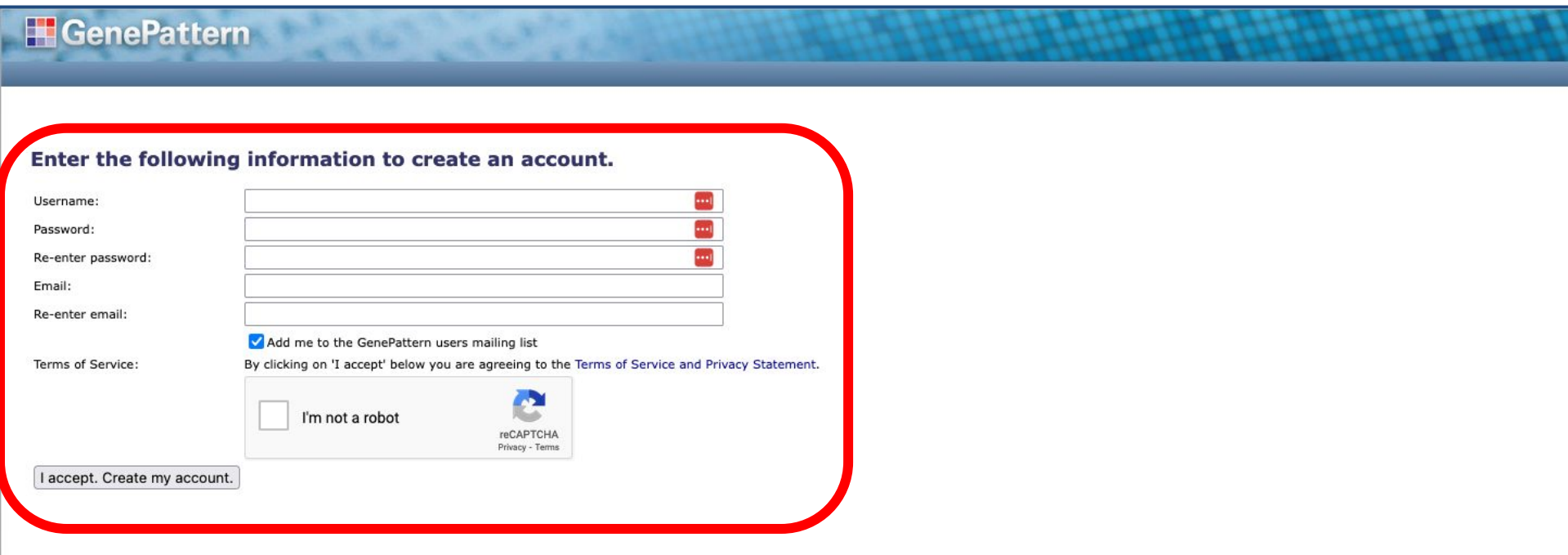
**Thorin Tabor**

**Anthony Castanza**

**Alex Wenzel**

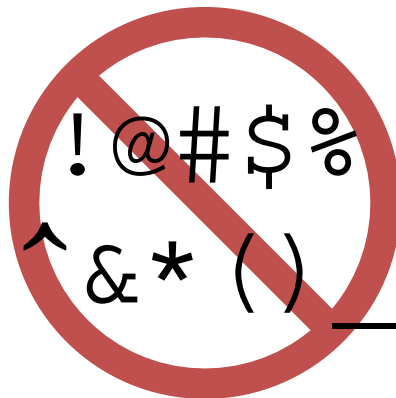
**Ted Liefeld**

# 1. Register for an account at: **cloud.genepattern.org**



The screenshot shows the GenePattern registration page. The header features the GenePattern logo. A red rounded rectangle highlights the registration form, which includes the following elements:

- Enter the following information to create an account.**
- Username: [text input field]
- Password: [text input field]
- Re-enter password: [text input field]
- Email: [text input field]
- Re-enter email: [text input field]
- ☒ Add me to the GenePattern users mailing list
- Terms of Service: By clicking on 'I accept' below you are agreeing to the [Terms of Service and Privacy Statement](#).
- ☐ I'm not a robot (reCAPTCHA widget)
- 



No spaces or  
special characters  
in usernames

# Agenda

- Overview
- Data Prep
- Differential Analysis
- GSEA
- MSigDB
- **Break**
- Classification & Prediction
- IGV
- Single-cell RNA-Seq Clustering
- GenePattern Copilot
- Closing



# 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/editing/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 htslib so they can be built independently.



## Bowtie 2

Fast and sensitive read alignment



Home

Installation

Documentation

Examples

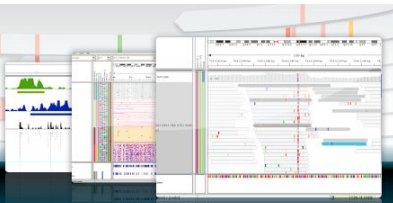
## 1.4. Support Vector Machines

## Hierarchical Clustering / Dendrograms

## Cufflinks

Transcriptome assembly and differential expression analysis for RNA-Seq.

## Integrative Genomics Viewer



## MAGeCK

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

## Cytoscape

Network Data Integration, Analysis, and Visualization in a Box

## NMF: Non-negative Matrix Factorization

### What is HAPSEG?

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

### What is RNA-SeQC?

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

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



**MSigDB**

Molecular Signatures Database

**FLAME**

Flow analysis with Automated Multivariate Estimation

## DESeq2

platforms all

downloads top 5%

in Bioc 4.5 years

build ok

DOI: [10.18129/B9.bioc.DESeq2](https://doi.org/10.18129/B9.bioc.DESeq2) [f](#) [t](#)

Differential gene expression analysis based on the negative binomial distribution

## Classification And Regression Trees for Machine Learning

## 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.



## Gene Set Enrichment Analysis

Constellation Map: Downstream visualization and interpretation of gene set enrichment results [version 1; referees: 2 approved]

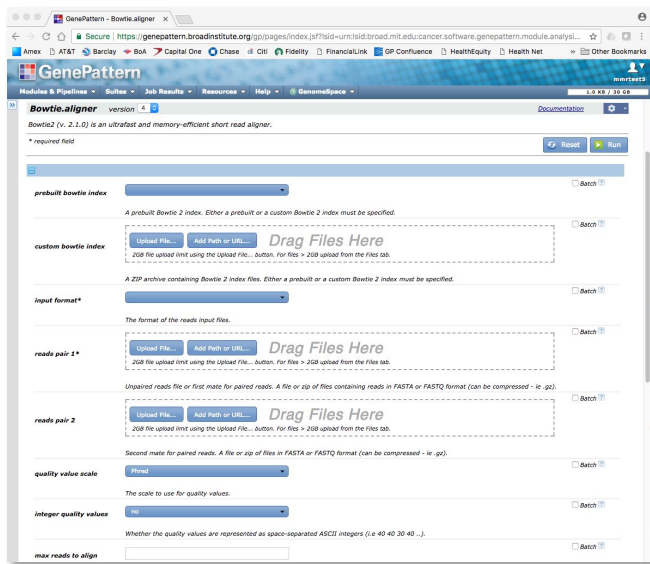
# Problems with bioinformatics tool use and interoperability

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

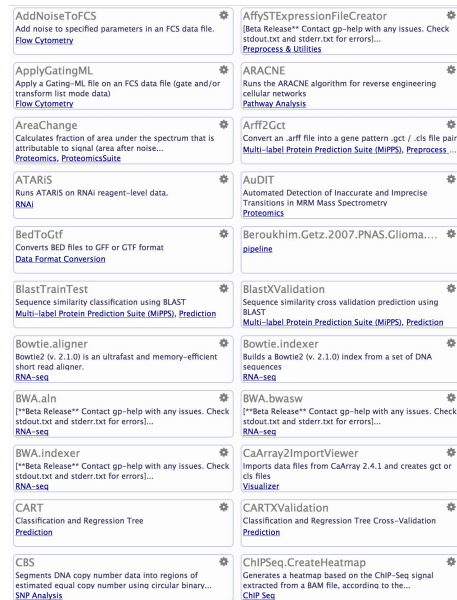
```
bowtie -a --best --strata -S -m 100 -X 400 --chunkmbs 256 --fullref -p 4
Dmel.BDGP5-transcripts \ -1 SRR031714_1.fastq -2 SRR031714_2.fastq |
samtools view -F 0xC -bS - | \ samtools sort -n -
~/Desktop/untreated3-transcriptome
```



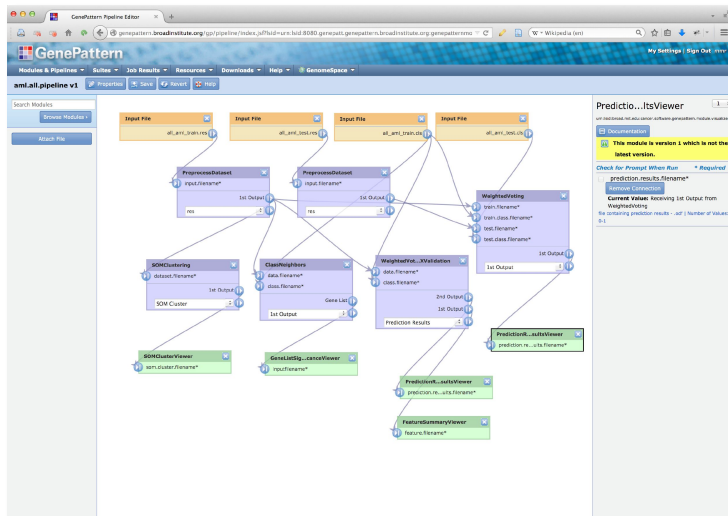
## Solution : GenePattern capabilities



## Wrap tools in a user-friendly interface



Provide a collection of hundreds of 'omics tools



## Provide features for reproducibility

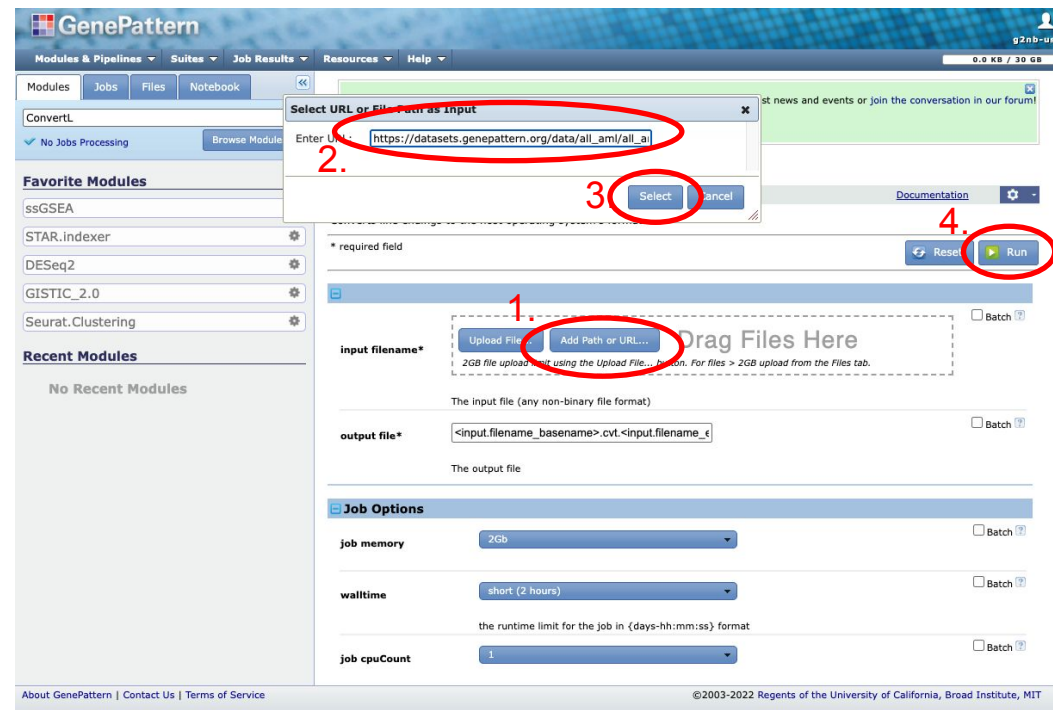
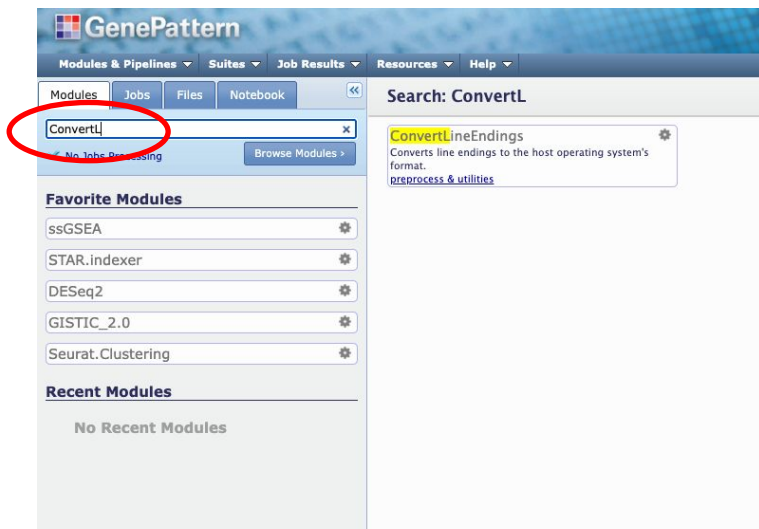
- Record and replay of all analyses
- Retain all versions of code
- Chain analyses into “pipelines” that can be shared and published



# **Exercise #1:**

## **Using Classic GenePattern**





- Go to <https://cloud.genepattern.org>
- Sign in using your GenePattern username and password.
- On the left, search for ConvertLineEndings and click it in the list when it comes up.
- Click “Add Path or URL” next to the “input filename” parameter and enter:  
[https://datasets.genepattern.org/data/all\\_aml/all\\_aml\\_train.gct](https://datasets.genepattern.org/data/all_aml/all_aml_train.gct)



# **g2nb**

**The Genomics to Notebook environment**



# g2nb Workspace

g2nb

File Edit View Run Kernel Git Tabs Settings Help

SEARCH

▼ +

Cytoscape Network  
Displays a Cytoscape network in the notebook.

Galaxy Login  
Log into a Galaxy server

GenePattern Login  
Log into a GenePattern server

Integrative Genomics Viewer ...  
Use igv.js to embed an interactive genome visualization

▼ GENE\_PATTERN\_CLOUD

AddFCSEventIndex  
Adds indexes to events in a Flow Cytometry Standard (FCS) data ...

AddFCSPParameter  
Add parameters and their values to a FCS data file

AddNoiseToFCS  
Add noise to specified parameters in an FCS data file.

AffySTExpressionFileCreator  
Creates a GCT file from a set of CEL files from Affymetrix ST

Amaretto  
Discovery of driver genes using epigenomic, genomic and

AmpliconSuite  
A multithread-enabled quickstart tool for AmpliconArchitect.

AmpliconSuiteAggregator

g2nb Overview and Tutorial

Python 3 (ipykernel)

Code

git

GenePattern Notebook tablor

https://cloud.genepattern.org/gp

Version 2

Run

ConvertLineEndings

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

input\_filename\*

Upload

The input file (any non-binary file format)

output\_file\*

<input.filename\_basename>.cvt.<input.filename\_extension>

The output file

Run

## Authentication Cells

The first GenePattern cell that you have encountered is an Authentication Cell. This cell allows a user to sign into a GenePattern server. Doing this allows GenePattern to keep a user's results private, and to remember a user's settings.

Authentication cells look like a login form with the additional option of selecting which GenePattern server to sign into. If the user has already authenticated, such as when using the g2nb Notebook Workspace, the user may instead be prompted to either sign in as the current user or to cancel and sign in as a different user

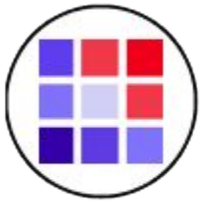
### Instructions

- Sign into the GenePattern server by clicking **Login as [username]** in the authentication cell above.
- Alternatively: fill in your credentials on the authentication cell above and click **Log into GenePattern**.

## Analysis Cells

After signing into an Authentication Cell, users can now access GenePattern analysis cells. Going to the g2nb Toolbox (Tool/s) will display the list of available GenePattern analyses. Search for a desired analysis and simply click on it to add to your current notebook.

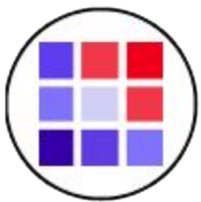
Every Analysis Cell contains form inputs for the required parameters of each analysis. Clicking Run will upload any files as well as launch a job on the GenePattern server. The status of the job in GenePattern's queue will be displayed below. Upon completion, the cell will show a list of outputs, which can be displayed in the browser, downloaded, or sent as input to another GenePattern analysis. Outputs are indicated by the icon. If this analysis includes a visualization, the visualization will load and appear inside as well.



# Jupyter

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





# Complete Research Narrative

- Leverages the best of Jupyter, GenePattern and other popular bioinformatics platforms
- Interleave text, visualization, graphics and analytical aspects

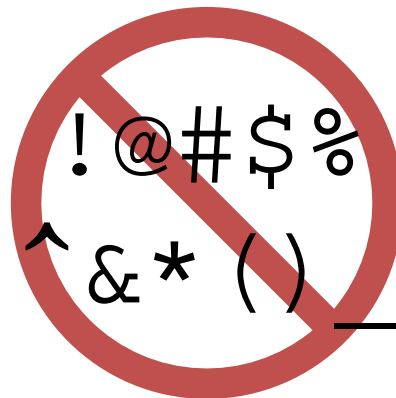
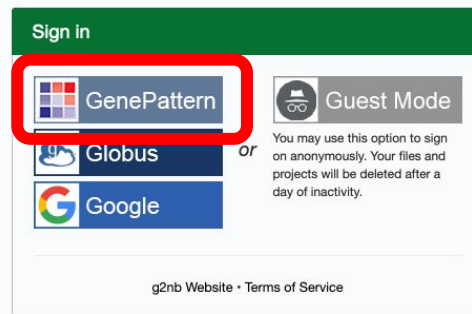




# **Exercise #2:**

## **Using g2nb**

# 1. Sign in at: **workspace.g2nb.org**



No spaces or  
special characters  
in usernames



# 2. Run the workshop project: 2025-08-22 MSTP Bootcamp

2022-09-01 19:35:20.322736

testgalaxy

Project containing edited Galaxy notebook

Shared by edwin5588

2022-10-17 18:45:22.197324

TestAuth

Shared by thorin

## Public Library

Browse, launch and reproduce public projects.

Search Library

Q

featured

tutorial

workshop

all projects

RNA-velocity analysis

RNA-velocity analysis using Alevin, Scanpy, and scVelo

featured

Analyze and Quantify RNA-seq Data using Salmon

End to end RNA-seq quantification, differential expression, and gene set enrichment analysis using GenePattern

1 copy

featured

Proteogenomics Workflow

Proteogenomics Galaxy Workflow

featured

galaxy

ATAC-Seq Workflow

ATAC-Seq Galaxy Workflow

1 copy

featured

galaxy

The \*AMARETTO framework in GenePattern Notebook

Multiscale and multimodal inference of regulatory networks to identify cell circuits and their drivers shared/distinct within/across biological systems of human disease, especially cancer

2023-11-09 Solid Tumor Therapeutics Workshop

Workshop for Moores Cancer Center Solid Tumor Therapeutics Program

1 copy

workshop

g2nb Tutorial

A tutorial on how to use the basic features of the g2nb environment.

2 copies

tutorial

Seurat Suite

QC, preprocessing, clustering, and visualization of scRNA-seq data using Seurat

1 copy

scrnaseq

single-cell

seurat

ssGSEA\_ROC Analysis of TCGA Samples

Perform sample selection and ROC classification of single sample GSEA scores

Trajectory Inference from single-cell data using STREAM

STREAM is a trajectory inference method that can accurately reconstruct complex developmental trajectories from gene expression or epigenomic ...

featured

single-cell

ma-seq

NMF Clustering

1 copy