

GenePattern Notebooks

An integrative analytical environment for cancer research

www.genepattern.org

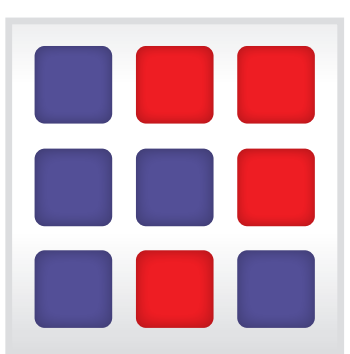


Michael Reich, Thorin Tabor, Peter Carr, David Eby, Helga Thorvaldsdóttir, Barbara Hill, Ted Liefeld, Pablo Tamayo, Jill P. Mesirov

Background

As the availability of genetic and genomic data and analysis tools from large-scale cancer initiatives continues to increase, the need has become more urgent for a software environment that supports the entire “idea to dissemination” cycle of an integrative cancer genomics analysis.

To address these needs, we have developed GenePattern Notebook, based on the GenePattern environment for integrative genomics and the Jupyter Notebook system.



GenePattern

GenePattern Notebook presents a familiar lab notebook format that allows researchers to build a record of their work by creating “cells” containing text, graphics, or executable analyses.

Researchers add, delete, and modify cells as the research evolves.

When an analysis is ready for publication, the same document that was used in the design and analysis phases becomes a research narrative that interleaves text, graphics, data, and executable analyses, serving as the complete, reproducible, *in silico* methods section for a publication.

Notebooks for Cancer Research

We are working with researchers to formulate and deploy GenePattern notebooks, embodying complete analysis studies that investigators can tailor to their own work. Notebooks in development include:

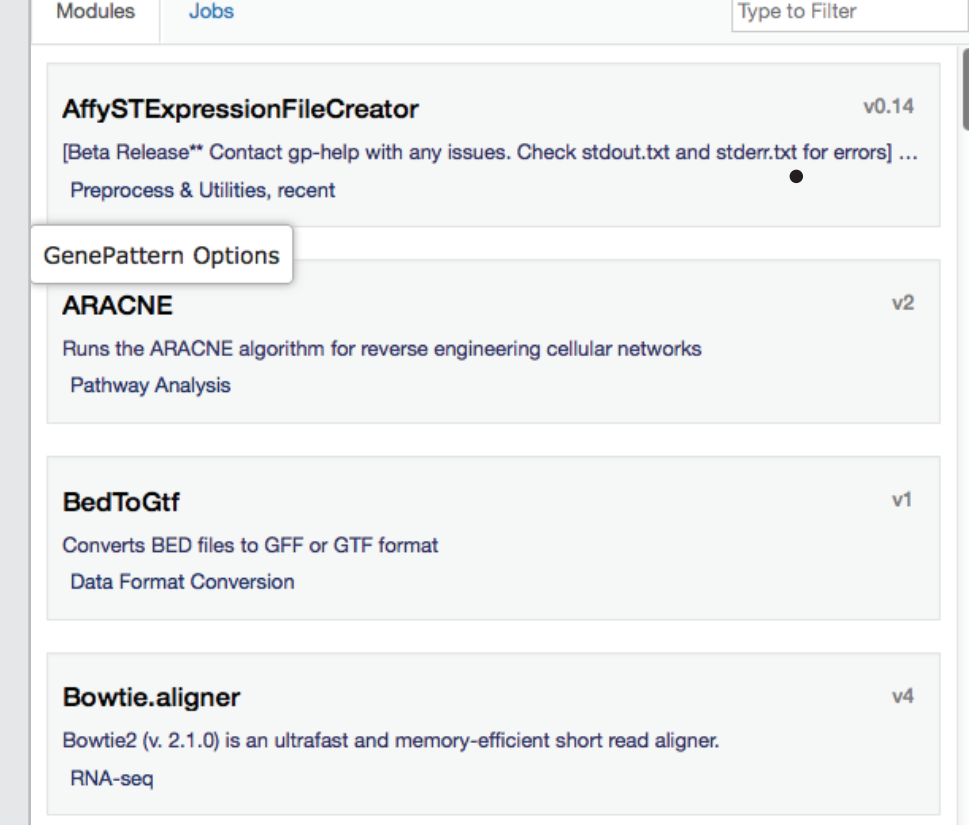
- Intratumoral Heterogeneity Analysis
Characterize intratumoral heterogeneity from single cell RNA-Seq data for ~100 cells from each of 5 or more tumors.
- Actionable Somatic Variants
Provide effective clinical interpretation of comprehensive genomic profiling from whole exome sequencing of a patient’s tumor and germline samples.
- Oncogenic Activation Master Regulator Analysis:
Identify master regulators/ transcription factors associated with the downstream transcriptional effects associated with the activation of an oncogene.



Example Notebook: Identifying Medulloblastoma Subtypes

This notebook represents an excerpt of the analytical steps and scientific narrative of a recently published work aimed at identifying the link between medulloblastoma subtype and likelihood of relapse. Because of the integration of text, graphics, and executable analysis, the notebook contains all of the information required to understand, execute, and reproduce the work described.

Notebooks can be created, viewed, and edited in any current browser.



A pullout displays all available analyses. Users can scroll to the desired analysis or type into the search bar.

Text formatting options are inherited from the flexible Jupyter Notebook interface and include headers, bullet lists, font formatting, tables, and Web links.

The notebook environment supports graphics as well as videos and other multimedia formats.

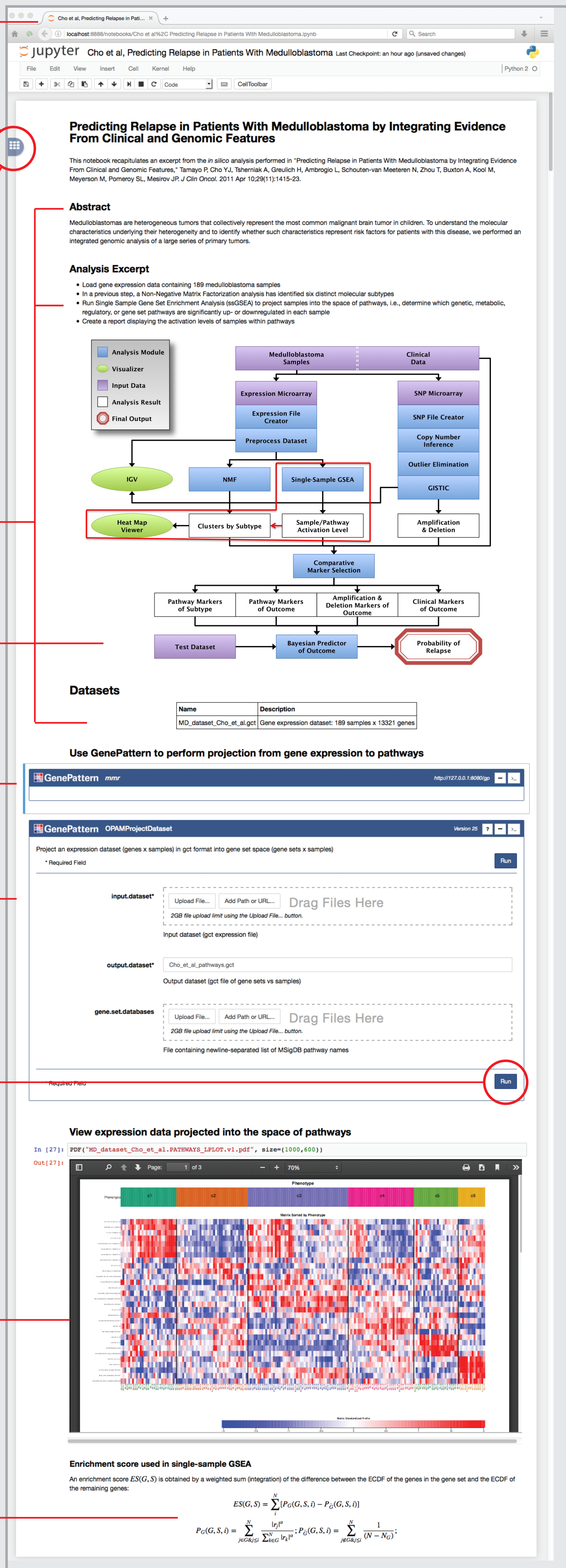
A login cell allows a user to connect to any available GenePattern server, including a user-installed local one.

To run an analysis, the researcher enters the requested parameters and uploads the input datasets. Datasets can come from local storage or can be Web links.

The Run button submits the analysis to the GenePattern server. A status indicator displays whether the analysis is pending, running, or completed.

Analysis results can be displayed in a wide variety of formats, integrated with the rest of the notebook narrative.

The notebook environment typesets complex mathematics.

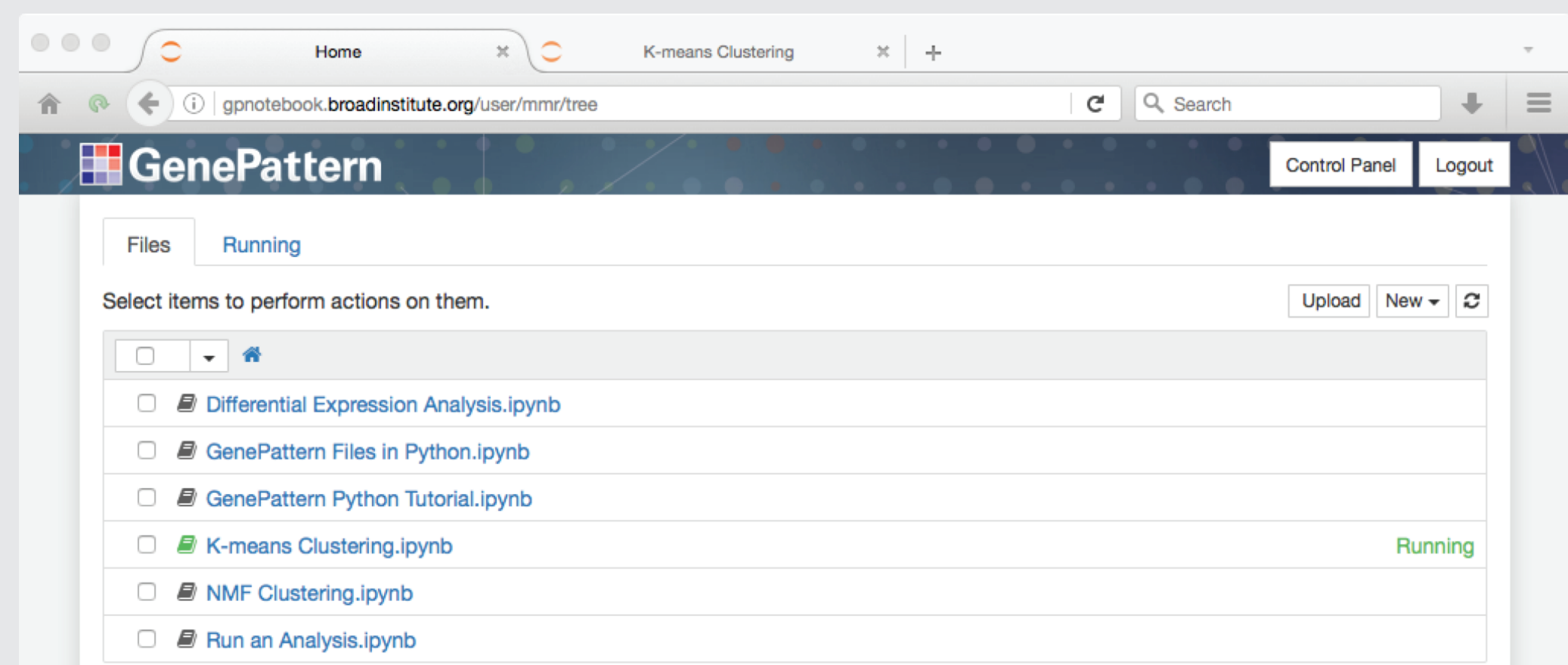


GenePattern Notebooks Online Repository

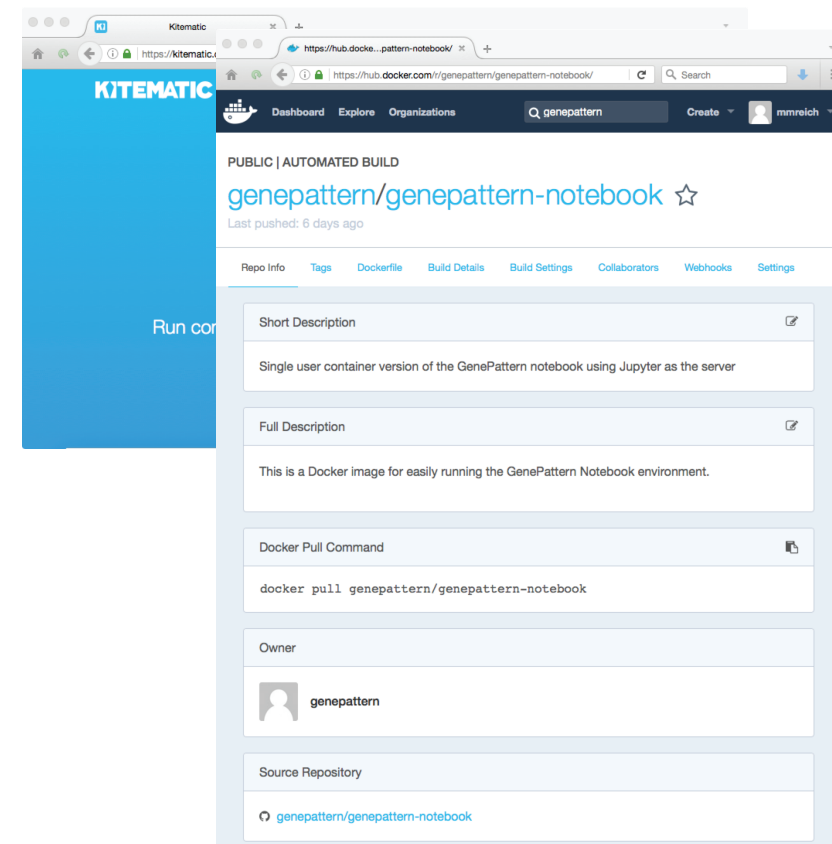
We are developing an online repository where researchers will be able to:

- Create and edit notebooks using only a web browser
- Find notebooks contributed by others and adapt them
- Share notebooks with collaborators and/or the general public

A preview version of the Notebook Repository is available at notebook.genepattern.org



GenePattern Notebooks Local Installation



`pip install genepattern-notebook`

A **docker image** provides a complete environment including GenePattern Notebooks, Jupyter, and all requirements, in a single installation. There is also a Docker image that includes an installable version of the GenePattern Notebook repository.

Programmers can install the notebook extension through the **pip** and **conda** package managers.

Connecting to GenePattern

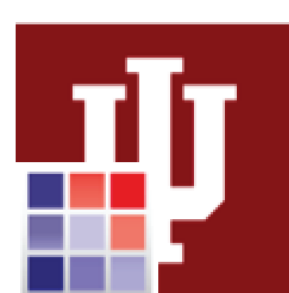
GenePattern is available online as well as for a local or cloud-based installation:



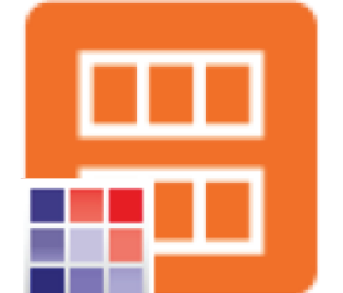
Broad Institute server
genepattern.
broadinstitute.org



Laptop, desktop, or
local server
installation



Indiana University
supercompute cluster
gp.indiana.edu



Amazon Machine Instance
(AMI)



Collaborations in progress with XSEDE resources at Pittsburgh Supercomputing Center (PSC) and Texas Advanced Computing Center (TACC)