

Introduction to GenePattern

Rehan Akbani

rakbani@mdanderson.org

Overview

1. What is GenePattern and why do I care?
2. How do I convert my script into a GenePattern module?
3. How do I create customized GenePattern pipelines?
4. How do I share my software/data with others?
5. How do I make my research reproducible using GenePattern?

What is GenePattern and why do I care?

- GenePattern (GP) is server software created by the Broad Institute
- What does it do?
 1. Browser based client side
 2. Allows interoperability between software tools (modules)
 3. Modules can be heterogeneous; using different languages and libraries
 4. Easily converts a module into a web service
 5. Allows the creation of workflows (pipelines)
 6. Modules/pipelines can be called directly from Java, Matlab or R
 7. Modules/pipelines can easily be shared
 8. Allows reproducible research

What is GenePattern and why do I care?

- TCGA will be using a GenePattern/Firehose pipeline to perform their monthly analysis runs (branded under NIH/NCI)
- GP will allow MDACC GDAC Analysis group to easily share tools internally
- GP server at Broad (free registration required):
<http://genepattern.broadinstitute.org/>
- MDACC local GP server behind firewall:
<http://mdadqsgdac1.mdanderson.edu:8080/gp/>

How do I convert my script into a GP module?

1. Analysis Modules

- Non-interactive, command line based
- Runs on the GP server

2. Visualization Module

- Can be interactive
- Runs on client's machine using Java applets

How do I convert my script into a GP module?

- Write a command line tool using any language
- Read/write input/output files from the current working directory
- Write messages to standard error and standard output
- Read module data files from <libdir>
- Read and write standard GenePattern file formats (e.g. gct, res, odf)
- Use command line parameter flags, instead of location
- Avoid absolute pathnames

- Ref:

http://www.broadinstitute.org/cancer/software/genepattern/tutorial/gp_programmer.html#_Writing_Modules_for_GenePattern

How do I convert my script into a GP module?

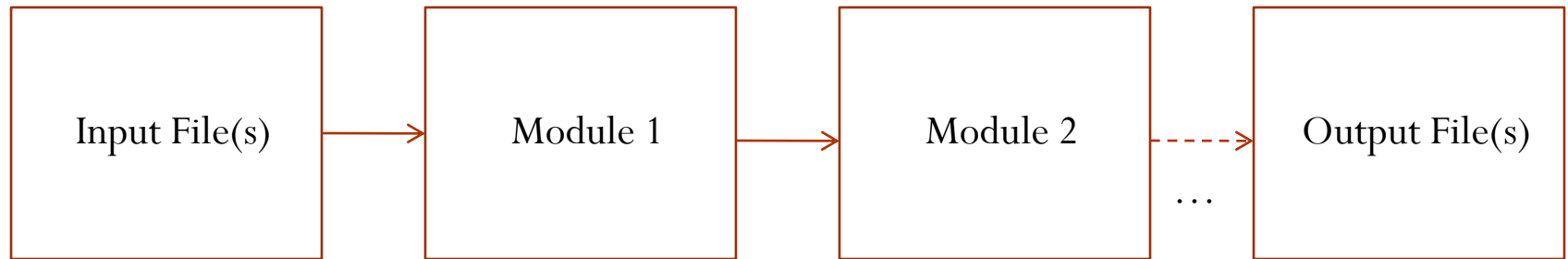
- Install your module into GP (if you have access rights)

e.g. invocation:

```
<perl> <libdir> myProg.pl -F <input.filename> -o <output.file>
```

- Click “Modules & Pipelines”->”New Module”
- Easily possible in MDACC GP, but not Broad due to access rights
- Decide if you want it to be public or private
- Ref:
http://www.broadinstitute.org/cancer/software/genepattern/tutorial/gp_web_client.html#creating_tasks

How do I create customized GP pipelines?



GP Pipeline

- See demo

How do I share my software/data with others?

- See demo

How do I make my research reproducible using GP?

- Jill Mesirov, “Accessible Reproducible Research,” Science, 22 January 2010: Vol. 327. no. 5964, pp. 415 – 416

Reproducible Research System (RRS)

