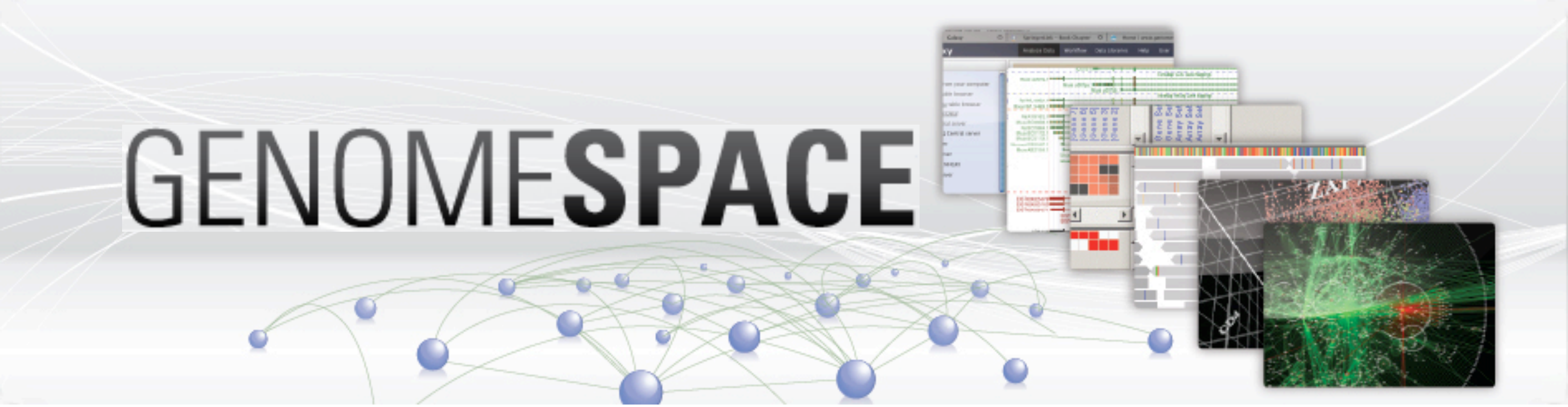
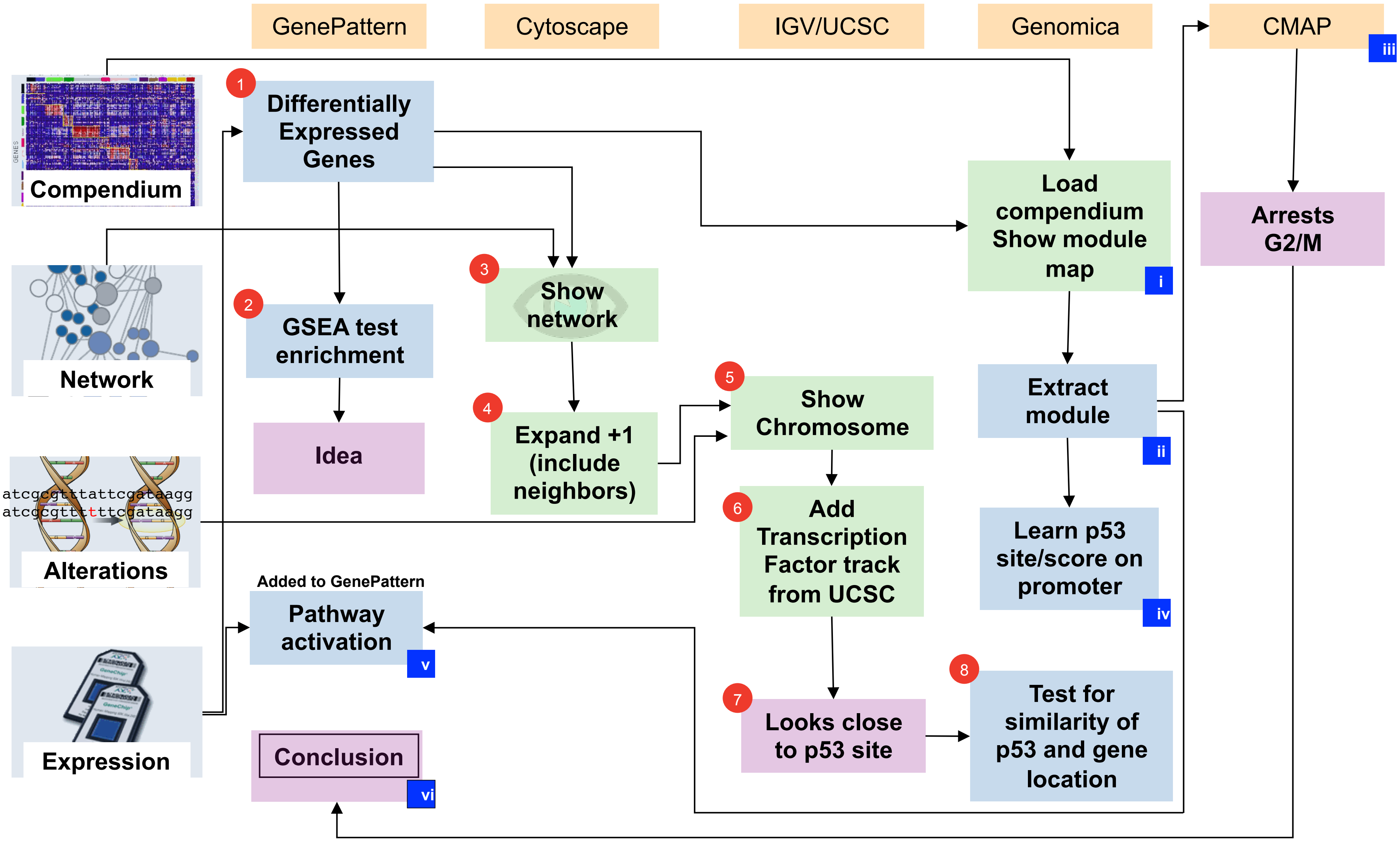

GENOMESPACE

Outline

- Introduction to GenomeSpace
- GenomeSpace Tools and Recipes
- GenomeSpace User Interface
- Integrative analysis exercise
- Other GenomeSpace Tools
- GenomeSpace development
- Q and A

The vision: Integrative Translational Genomics



GenePattern

Cytoscape

IGV/UCSC

Genomica

CMAP iii

Compendium

Network

Alterations

atcgcgtttattcgataagg
atcgcgttttttcgataagg

Expression

1 Differentially Expressed Genes

2 GSEA test enrichment

Idea

3 Show network

4 Expand +1 (include neighbors)

5 Show Chromosome

6 Add Transcription Factor track from UCSC

7 Looks close to p53 site

8 Test for similarity of p53 and gene location

Load compendium Show module map i

Extract module ii

Learn p53 site/score on promoter iv

Arrests G2/M

Added to GenePattern

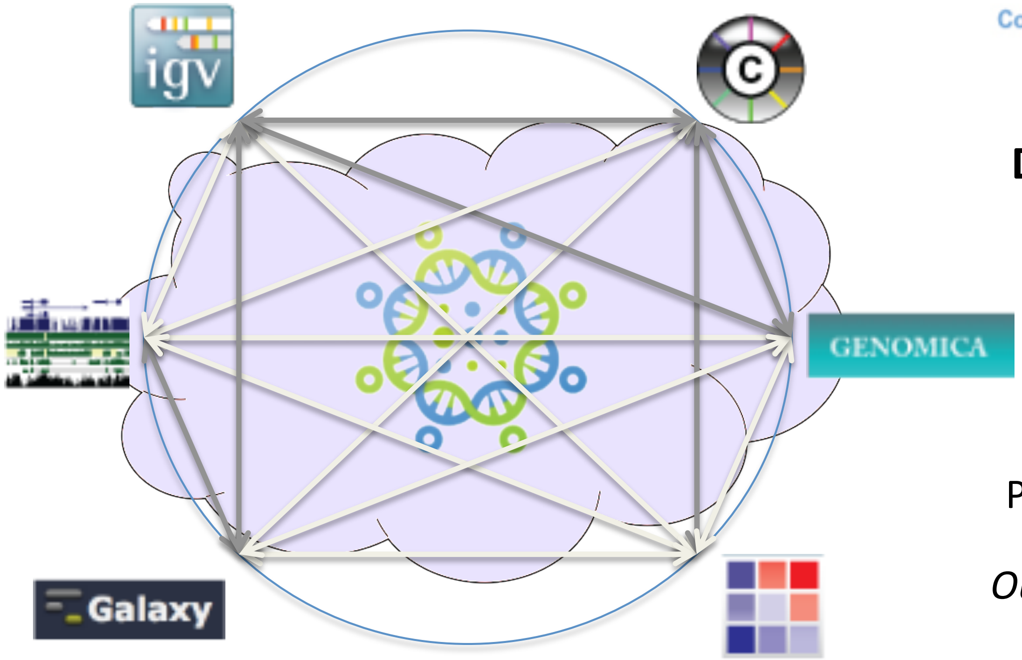
Pathway activation v

Conclusion vi

Online community to share diverse computational tools

GENOMESPACE

Connect Collaborate Combine



Seed Tools

Cytoscape
Galaxy
GenePattern
Genomica
IGV
UCSC Browser

Outreach to new tools

Driving Biological Projects

lincRNAs

Cancer stem cells

Patient Stratification

Outreach to new DBPs

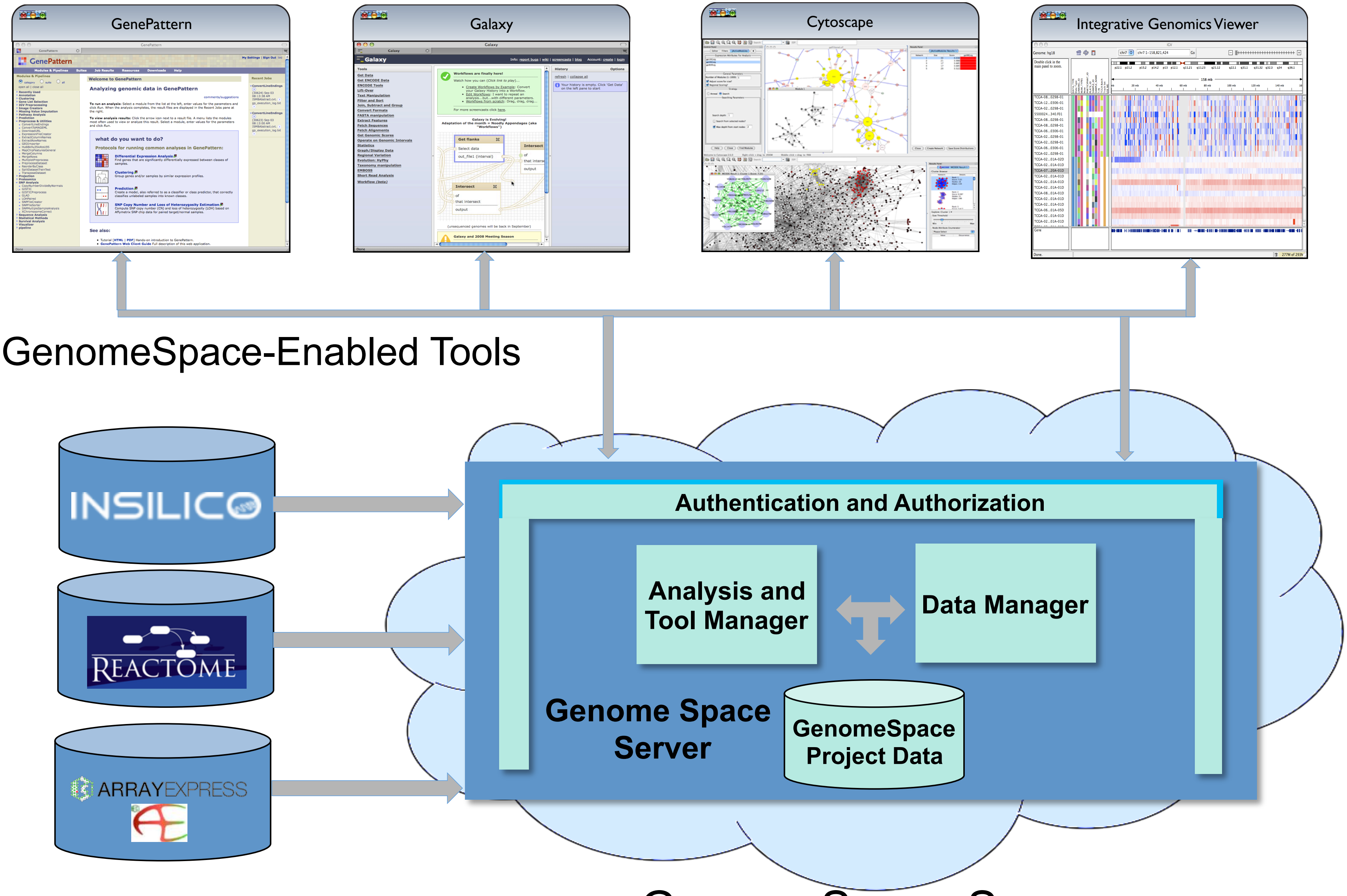
www.genomespace.org

UC SANTA CRUZ
UCSD
PENNSTATE 1855
BROAD INSTITUTE
מכון ויצמן למדע WEIZMANN INSTITUTE OF SCIENCE
STANFORD UNIVERSITY
National Human Genome Research Institute
amazon web services™

GenomeSpace: a connection layer between integrative analysis tools

- Support for all types of resource: Web-based, desktop, etc.
- Automatic conversion of data formats between tools
- Easy access to data from any location
- Ease of entry into the environment

GenomeSpace Components



GenePattern

Galaxy

Cytoscape

Integrative Genomics Viewer

GenomeSpace-Enabled Tools

INSILICO

REACTOME

ARRAYEXPRESS

Authentication and Authorization

Analysis and Tool Manager

Data Manager

Genome Space Server

GenomeSpace Project Data

GenomeSpace Server

Register

GENOMESPACE BETA

Frictionless connection of bioinformatics tools

STATUS 03.18.13 07:01AM 🟩 All systems are operating normally.

WHAT'S NEW

GenomeSpace at BOSC-Broad Interoperability Hackathon

GenomeSpace will be one of the featured tools at the BOSC/Broad interoperability hackathon on Sunday and Monday, April 7 and 8 at the Broad Institute. It occurs just prior to Bio-IT World in Boston and is a warm up event for BOSC 2013 occurring July 19-20th in Berlin. The goal of the hackathon will be developing open source tools to help enhance bioinformatics interoperability. See http://www.broadinstitute.org/events/interoperability-hackathon-2013/home for more information as

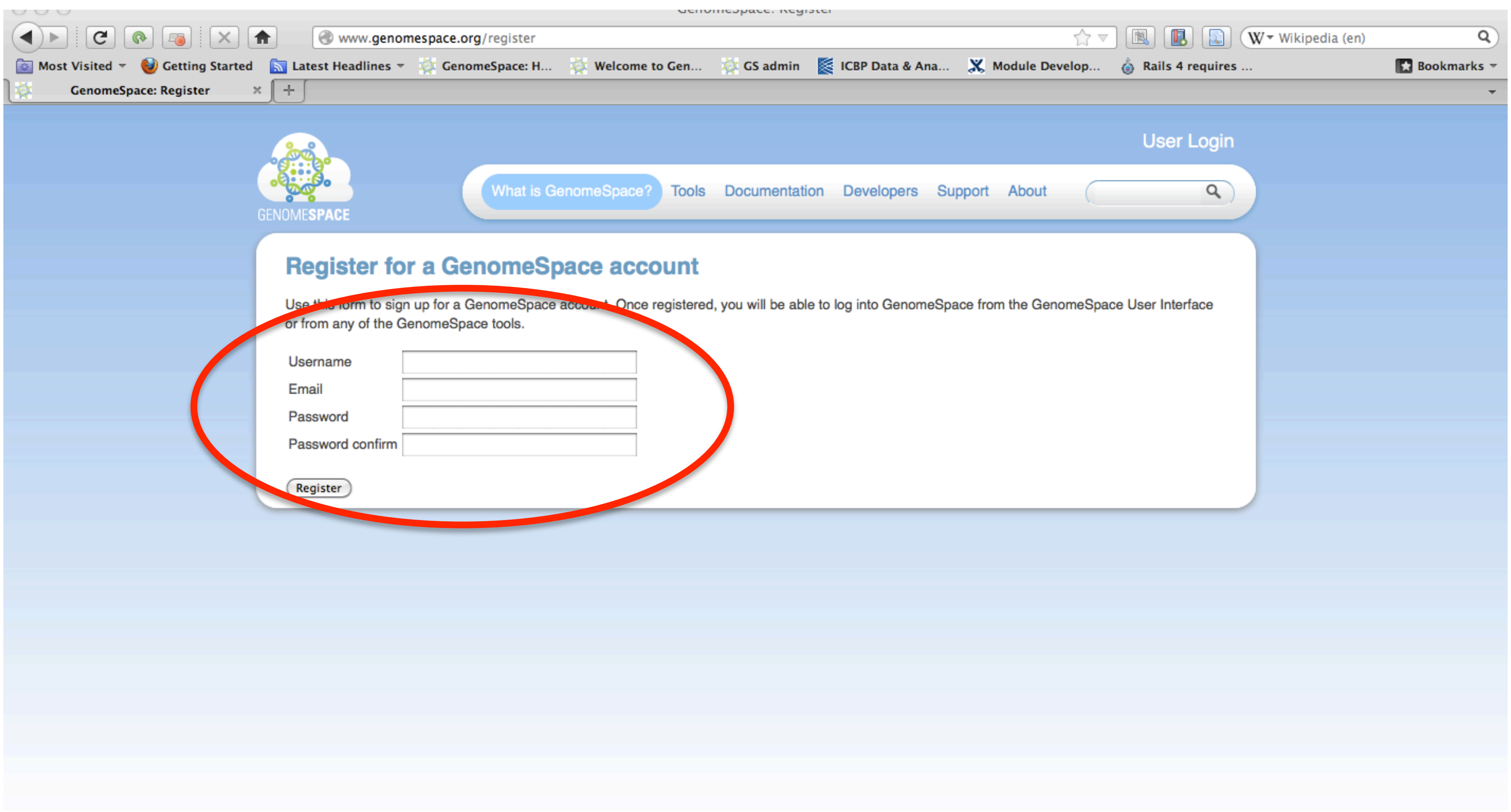
genomespace Learn GenomeSpace at VIZBI: Visualizing Biological Data, March 19–22 – workshop space still available lanyrd.com/2013/vizbi/ #bioinformatics

12 days ago · reply · retweet · favorite

genomespace Join GenomeSpace and others at the Bioinformatics Interoperability Hackathon in Boston, April 7–8. #bioinformatics

www.genomespace.org

Register



Register for a GenomeSpace account

Use this form to sign up for a GenomeSpace account. Once registered, you will be able to log into GenomeSpace from the GenomeSpace User Interface or from any of the GenomeSpace tools.

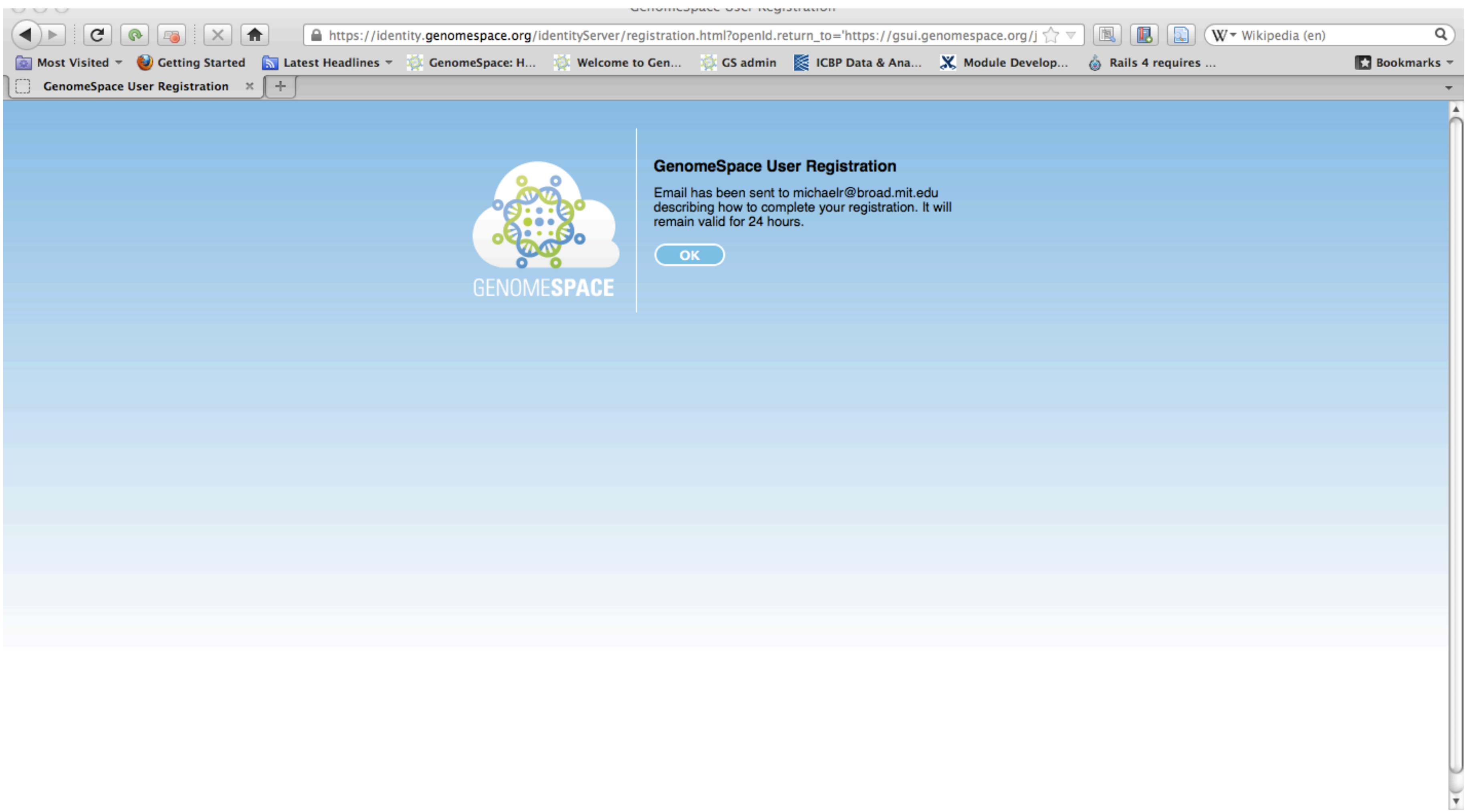
Username

Email

Password

Password confirm

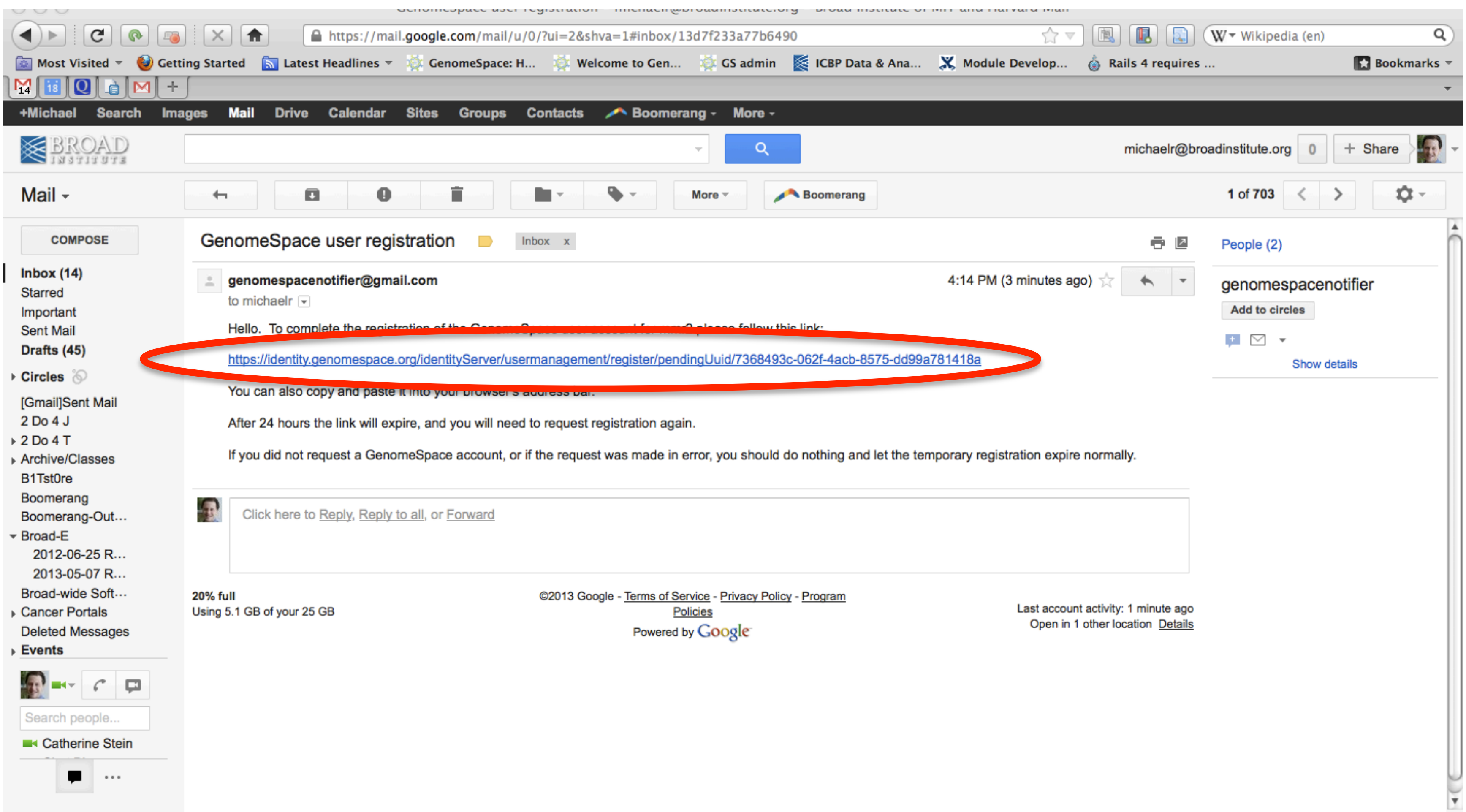
Register



GenomeSpace User Registration

Email has been sent to michaelr@broad.mit.edu describing how to complete your registration. It will remain valid for 24 hours.

Register



GenomeSpace user registration

genomespacenotifier@gmail.com

to michaelr

4:14 PM (3 minutes ago)

Hello. To complete the registration of the GenomeSpace user account for mmr2 please follow this link:

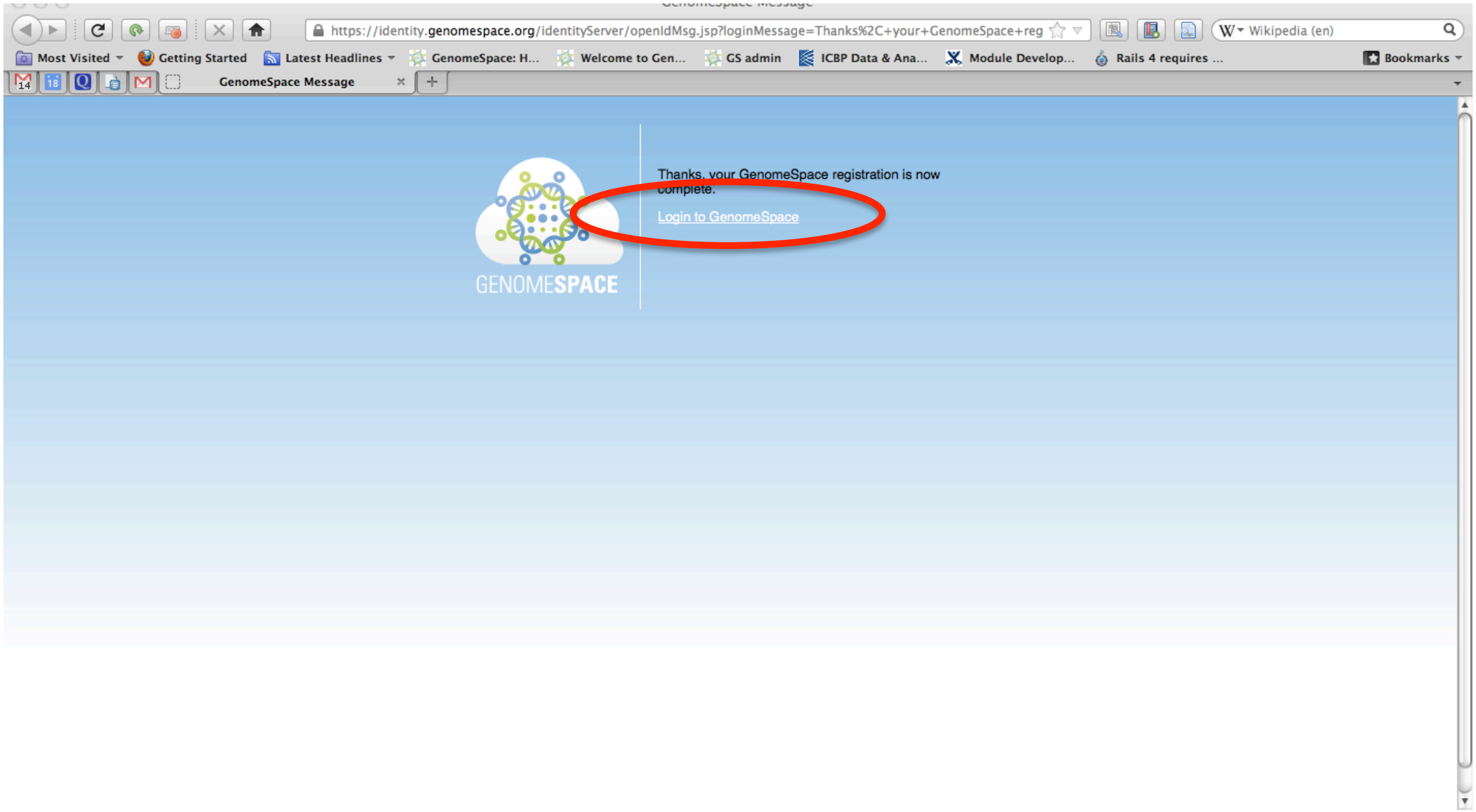
https://identity.genomespace.org/identityServer/usermanagement/register/pendingUuid/7368493c-062f-4acb-8575-dd99a781418a

You can also copy and paste it into your browser's address bar.

After 24 hours the link will expire, and you will need to request registration again.

If you did not request a GenomeSpace account, or if the request was made in error, you should do nothing and let the temporary registration expire normally.

Login

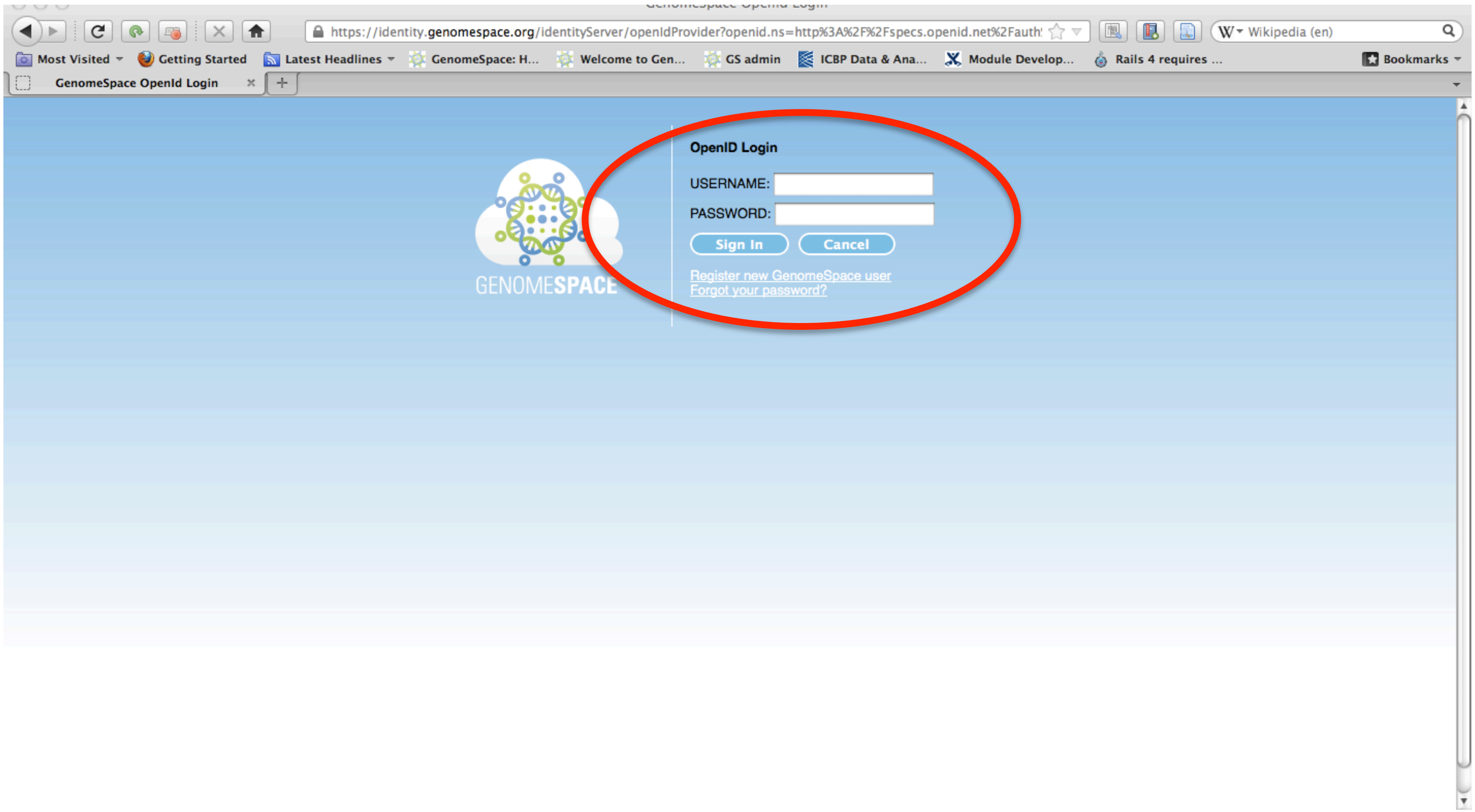


https://identity.genomespace.org/identityServer/openIdMsg.jsp?loginMessage=Thanks%2C+your+GenomeSpace+reg

Thanks, your GenomeSpace registration is now complete.

Login to GenomeSpace

Login



GenomeSpace OpenId Login

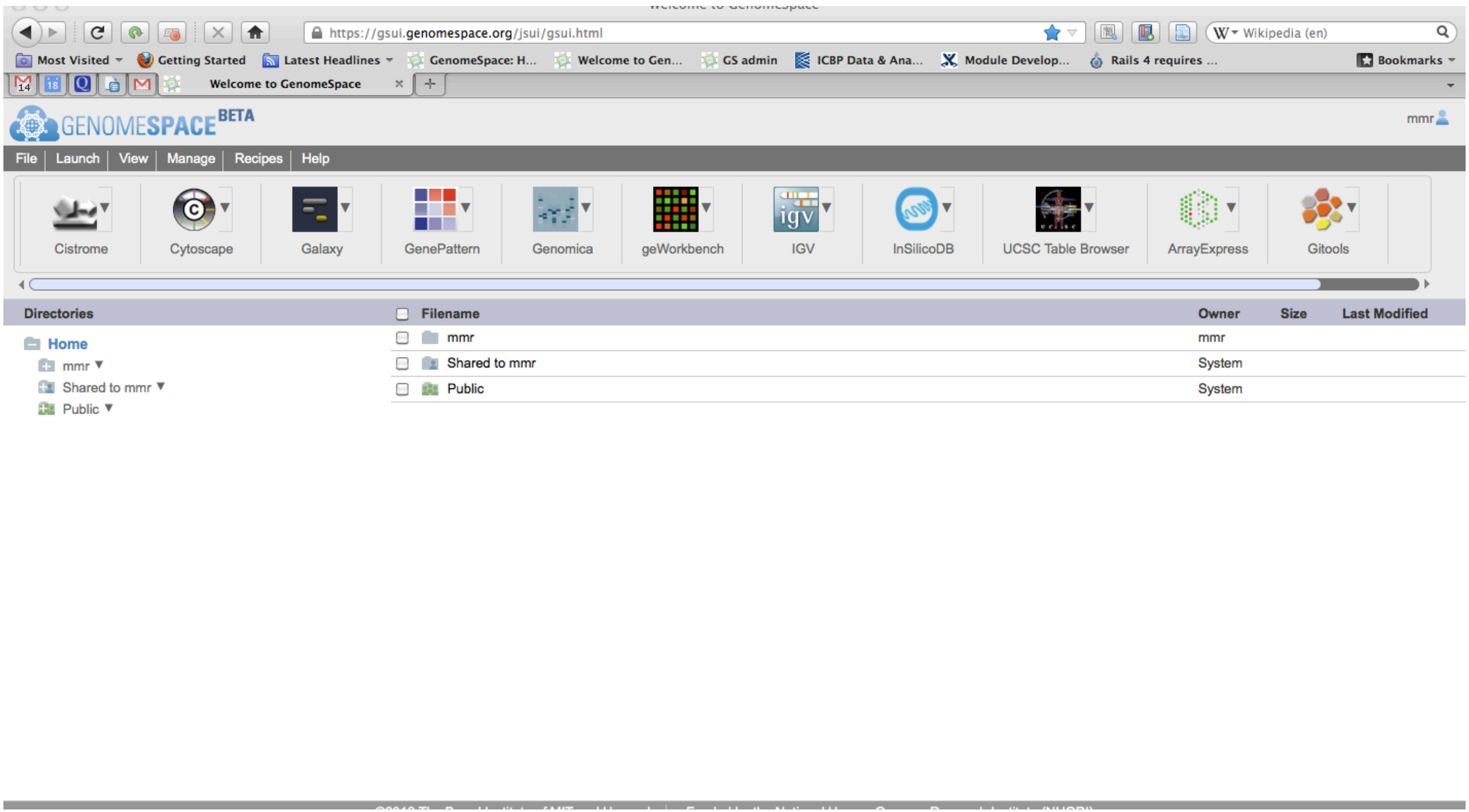
https://identity.genomespace.org/identityServer/openIdProvider?openid.ns=http%3A%2F%2Fspecs.openid.net%2Fauth!

OpenID Login

USERNAME:

PASSWORD:

GenomeSpace UI



☐ Filename	Owner	Size	Last Modified
☐ mmr	mmr		
☐ Shared to mmr	System		
☐ Public	System		

Tools and Recipes

Focus on Kitchen Skills

Agenda

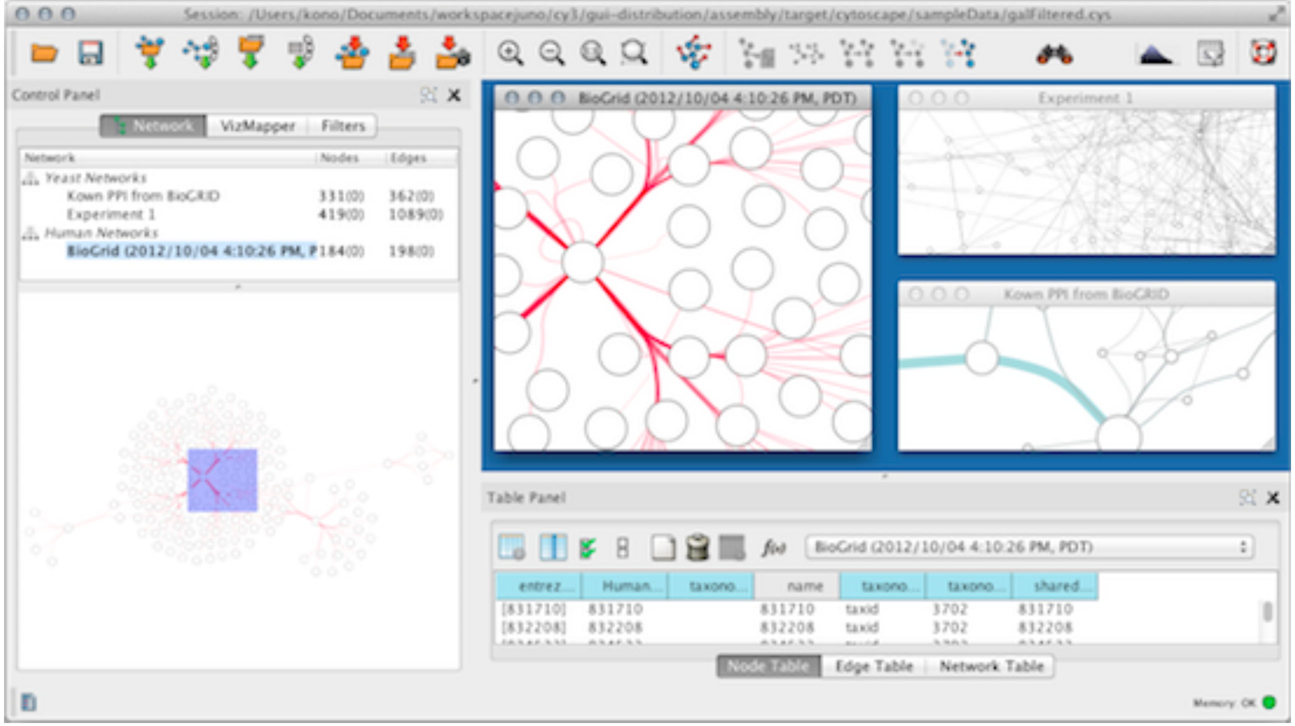
- Review of GenomeSpace tools in the first exercises
- Basic recipes for using GenomeSpace
  - Launching tools
  - Uploading data to GenomeSpace
  - Sending data to tools

GenomeSpace Tools

- ArrayExpress
- Galaxy
- Cistrome
- Cytoscape
- GenePattern
- Genomica
- ISAcreator
- geWorkbench
- Gitools
- IGV
- InSilicoDB
- UCSC Table Browser
- MSigDB

Cytoscape

Cytoscape is an open-source bioinformatics software platform for visualizing molecular interaction networks and biological pathways, and integrating these networks with annotations, gene expression profiles, and other state data.



Session: /Users/kono/Documents/workspacejuno/cy3/gui-distribution/assembly/target/cytoscape/sampleData/galFiltered.cys

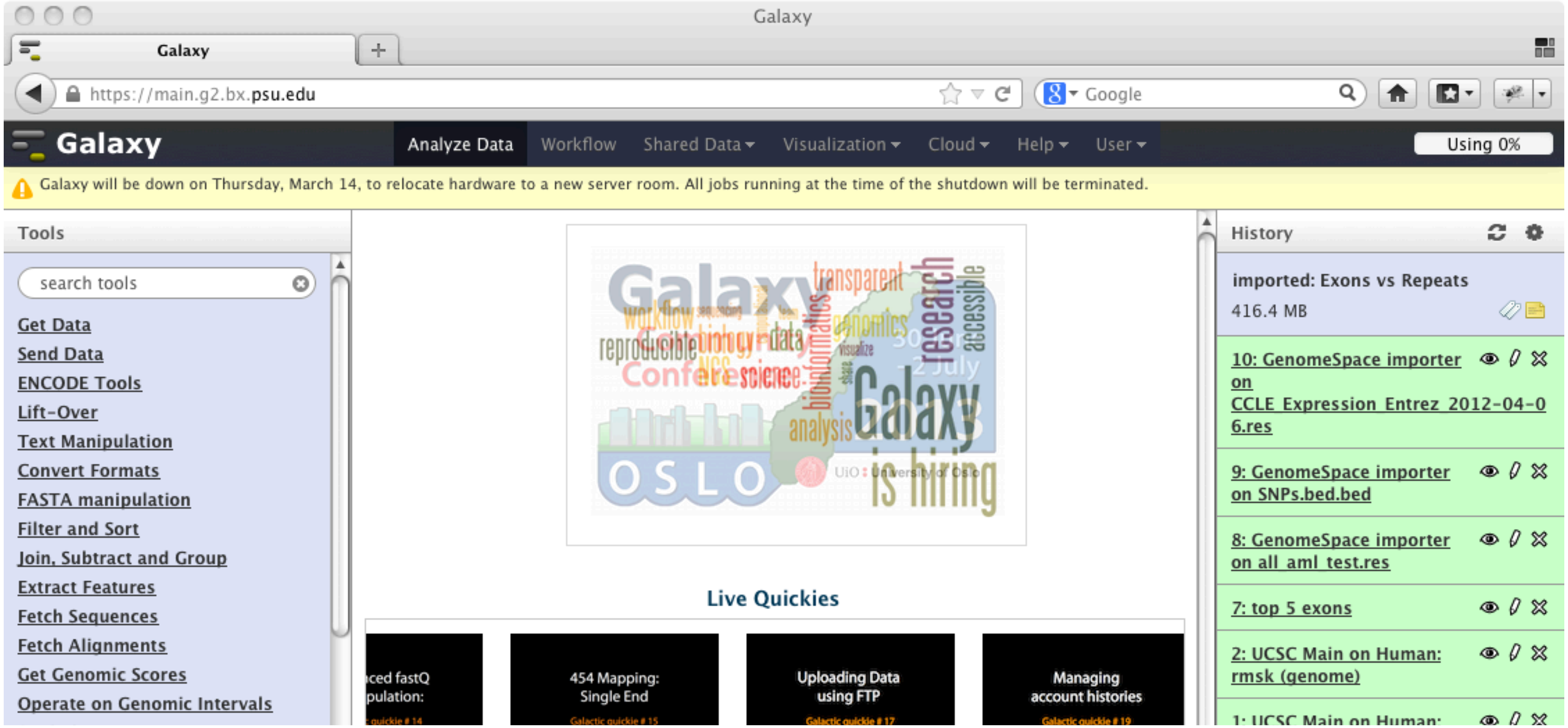
Network	Nodes	Edges
Yeast Networks		
Kown PPI from BioGRID	331(0)	362(0)
Experiment 1	419(0)	1089(0)
Human Networks		
BioGrid (2012/10/04 4:10:26 PM, P	184(0)	198(0)

BioGrid (2012/10/04 4:10:26 PM, PDT)

entrez...	Human...	taxono...	name	taxono...	taxono...	shared...
[831710]	831710		831710	taxid	3702	831710
[832208]	832208		832208	taxid	3702	832208

Galaxy

Galaxy is an open-source, scalable framework for tool integration that allows users to analyze multiple alignments, compare genomic annotations, and profile metagenomic samples, among many possible analyses; workflows allow the linking together of analyses.



Galaxy will be down on Thursday, March 14, to relocate hardware to a new server room. All jobs running at the time of the shutdown will be terminated.

imported: Exons vs Repeats
416.4 MB

10: GenomeSpace importer on CCLE_Expression_Entrez_2012-04-06.res

9: GenomeSpace importer on SNPs.bed.bed

8: GenomeSpace importer on all_aml_test.res

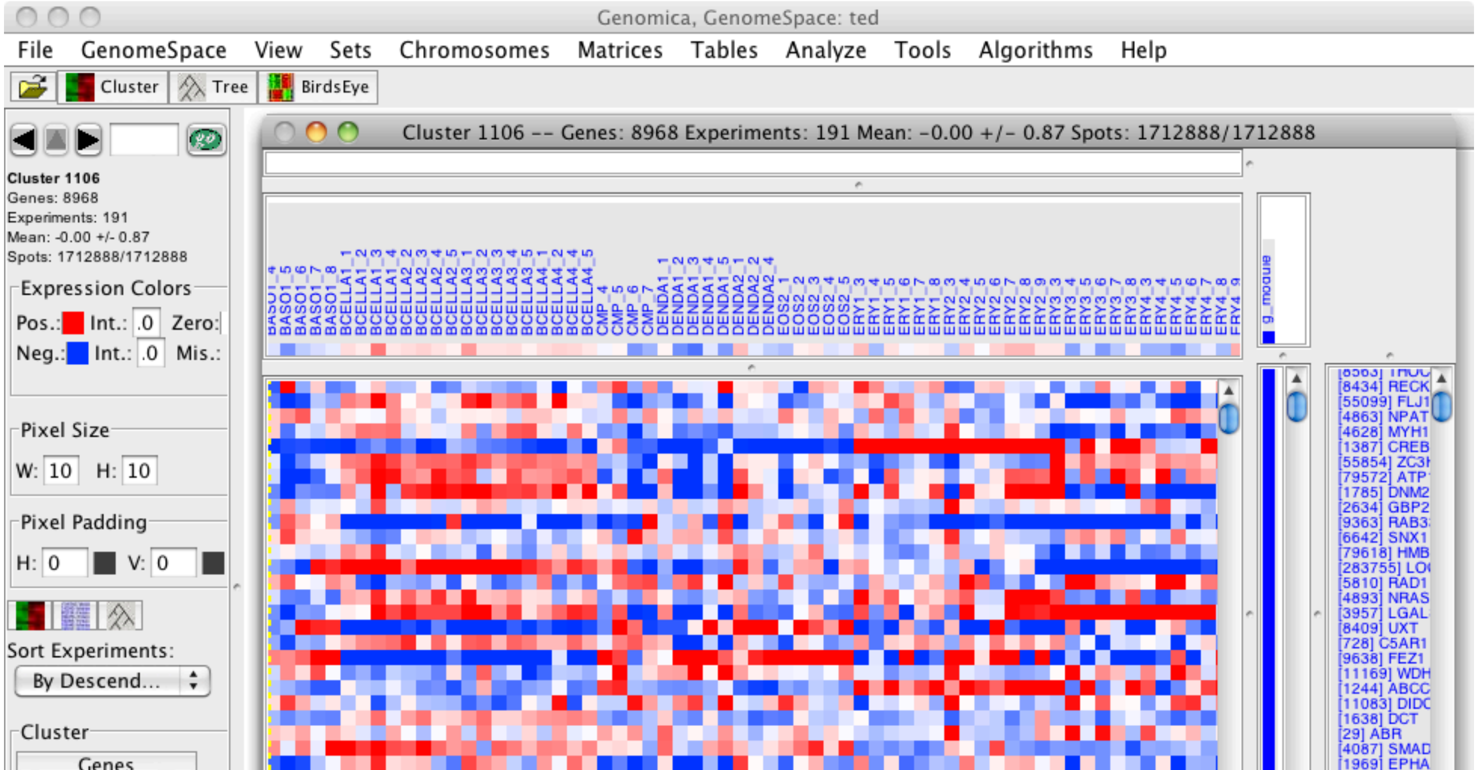
7: top 5 exons

2: UCSC Main on Human: rmsk (genome)

1: UCSC Main on Human:

Genomica

Genomica is an analysis and visualization tool for genomic data that can integrate gene expression data, DNA sequence data, and gene and experiment annotation information.



Genomica, GenomeSpace: ted

Cluster 1106 -- Genes: 8968 Experiments: 191 Mean: -0.00 +/- 0.87 Spots: 1712888/1712888

Cluster 1106
Genes: 8968
Experiments: 191
Mean: -0.00 +/- 0.87
Spots: 1712888/1712888

Expression Colors
Pos.: 🟥 Int.: .0 Zero:
Neg.: 🟦 Int.: .0 Mis.:

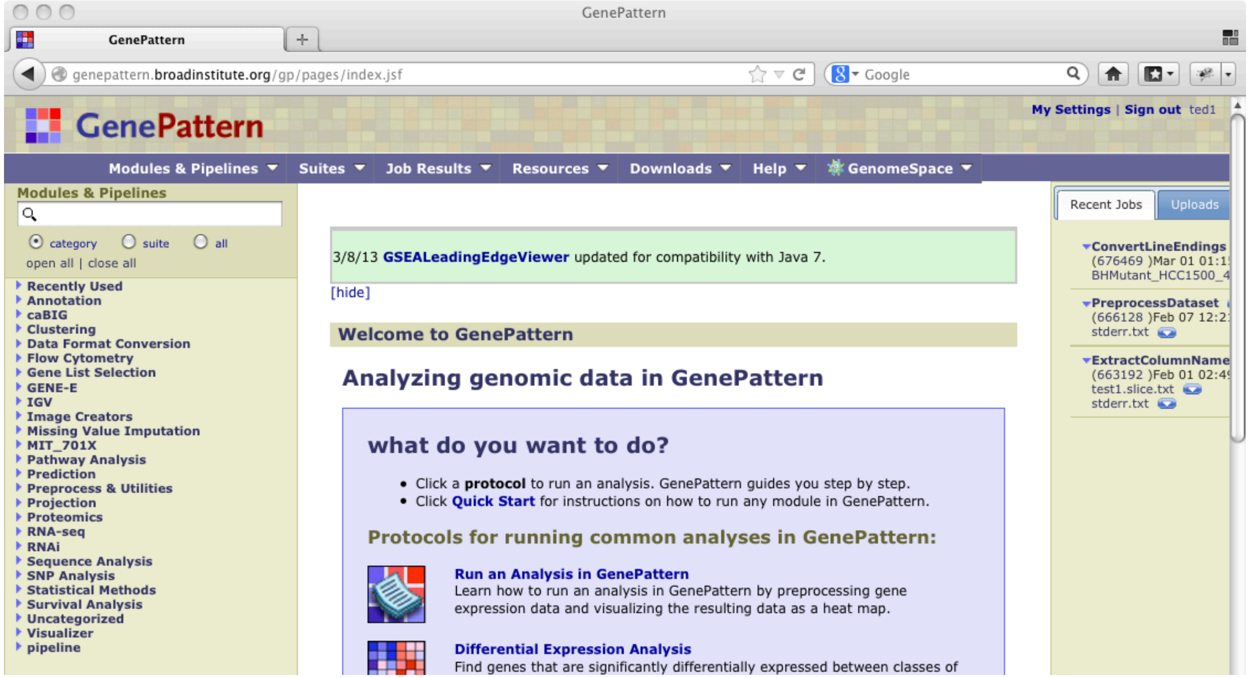
Pixel Size
W: 10 H: 10

Pixel Padding
H: 0 V: 0

Sort Experiments: By Descend...

GenePattern

GenePattern is a powerful genomic analysis platform that provides access to more than 150 tools for gene expression analysis, proteomics, SNP analysis, flow cytometry, RNA-seq analysis, and common data processing tasks. A web-based interface provides easy access to these modules and allows for the creation of multi-step analysis pipelines that enable reproducible *in silico* research.



3/8/13 **GSEALeadingEdgeViewer** updated for compatibility with Java 7.

[hide]

Welcome to GenePattern

Analyzing genomic data in GenePattern

what do you want to do?

- Click a **protocol** to run an analysis. GenePattern guides you step by step.
- Click **Quick Start** for instructions on how to run any module in GenePattern.

Protocols for running common analyses in GenePattern:

Run an Analysis in GenePattern
Learn how to run an analysis in GenePattern by preprocessing gene expression data and visualizing the resulting data as a heat map.

Differential Expression Analysis
Find genes that are significantly differentially expressed between classes of

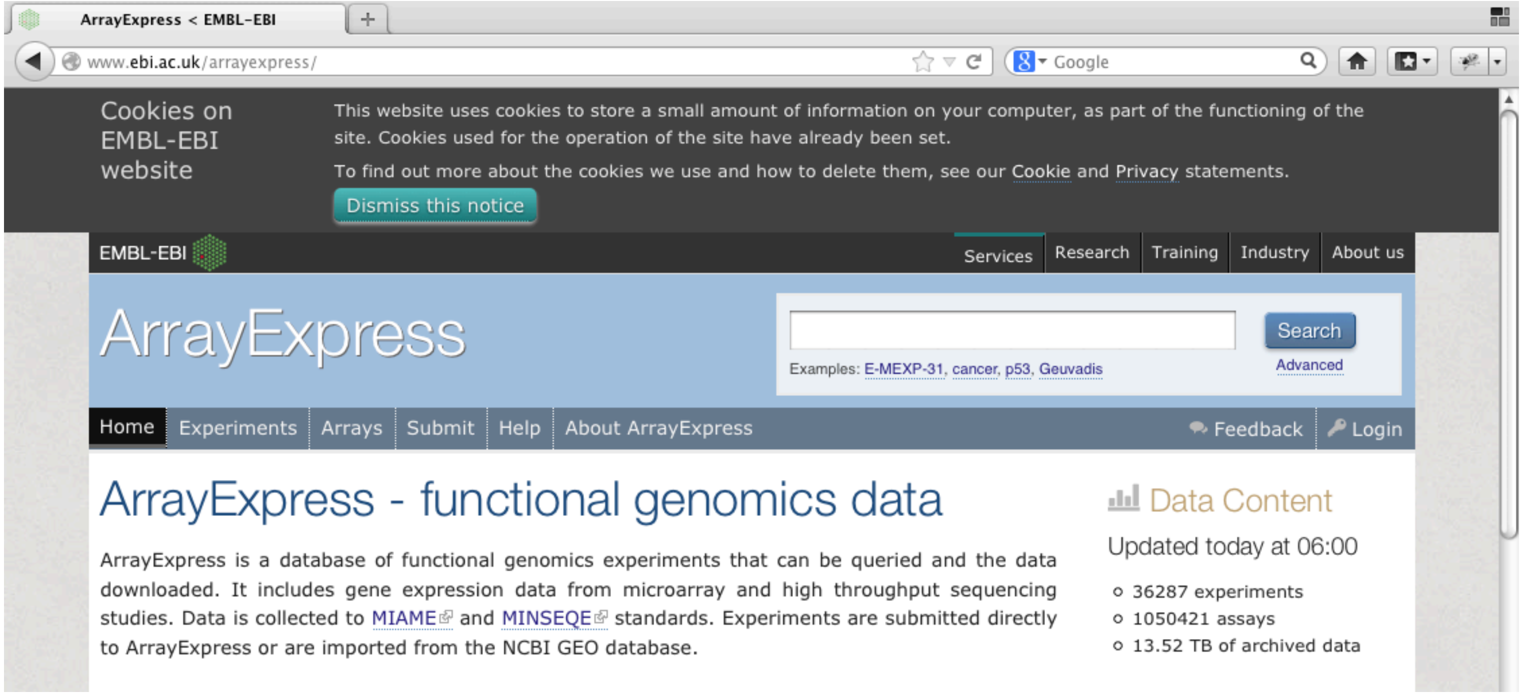
ConvertLineEndings
(676469)Mar 01 01:1
BHMutant_HCC1500_4

PreprocessDataset
(666128)Feb 07 12:2
stderr.txt

ExtractColumnName
(663192)Feb 01 02:4
test1.slice.txt
stderr.txt

ArrayExpress

ArrayExpress is a repository of over 30,000 functional genomics experiments comprising nearly 1 million assays. Users can query and retrieve data in a number of different formats including the MIAME and MINSEQE standards.



Cookies on EMBL-EBI website

This website uses cookies to store a small amount of information on your computer, as part of the functioning of the site. Cookies used for the operation of the site have already been set.

To find out more about the cookies we use and how to delete them, see our Cookie and Privacy statements.

Dismiss this notice

ArrayExpress

Examples: E-MEXP-31, cancer, p53, Geuvadis

ArrayExpress - functional genomics data

ArrayExpress is a database of functional genomics experiments that can be queried and the data downloaded. It includes gene expression data from microarray and high throughput sequencing studies. Data is collected to MIAME and MINSEQE standards. Experiments are submitted directly to ArrayExpress or are imported from the NCBI GEO database.

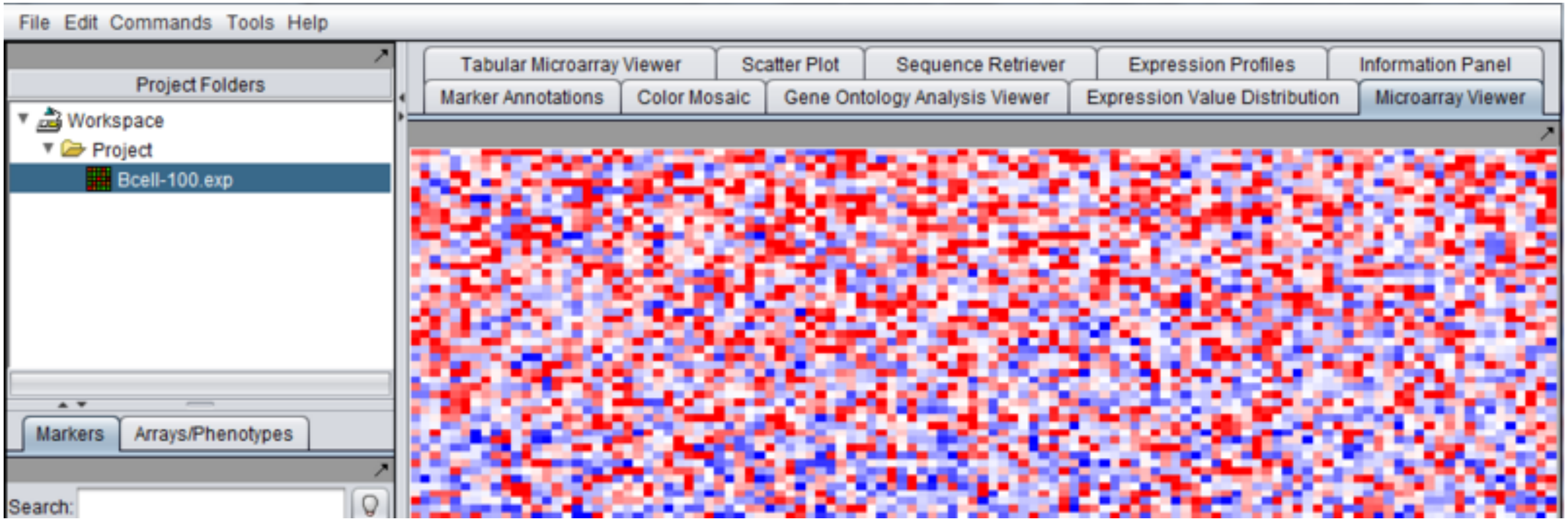
Data Content

Updated today at 06:00

- 36287 experiments
- 1050421 assays
- 13.52 TB of archived data

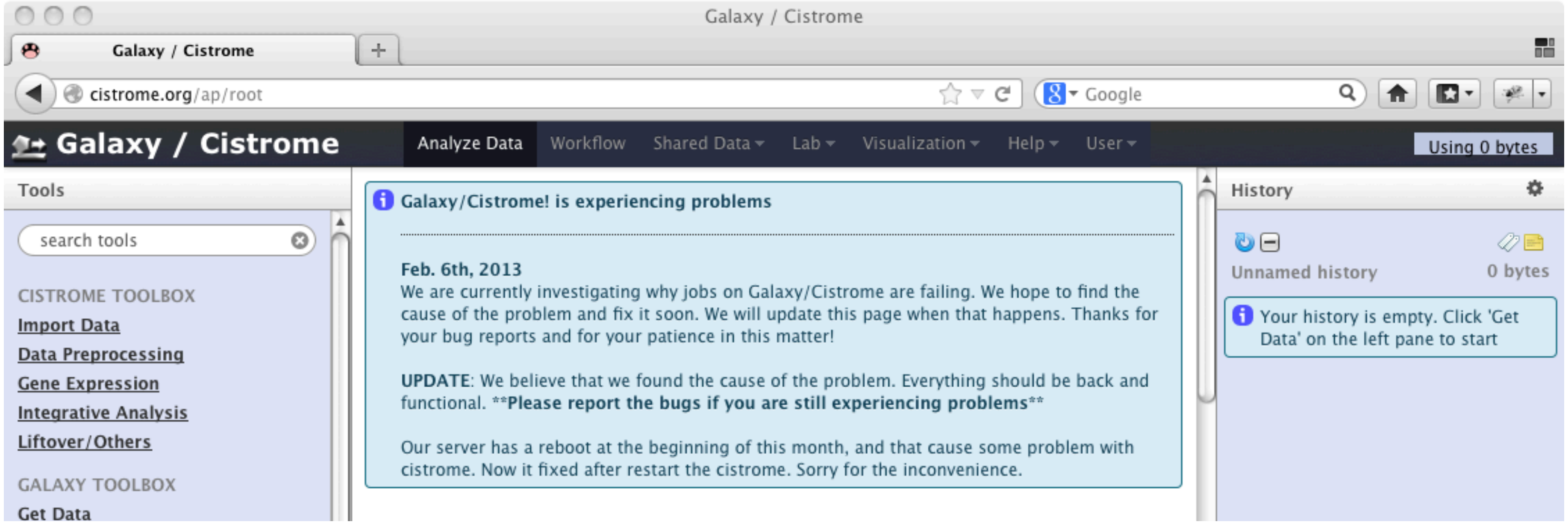
geWorkbench

geWorkbench is an open-source bioinformatics platform that offers a comprehensive and extensible collection of tools for the management, analysis, visualization, and annotation of biomedical data. For microarrays, there are tools for filtering and normalization, basic statistical analyses, clustering, network reverse engineering, as well as many common visualization tools



Cistrome

In addition to the standard Galaxy functions, Cistrome has 29 ChIP-chip- and ChIP-seq-specific tools in three major categories, from preliminary peak calling and correlation analyses, to downstream genome feature association, gene expression analyses, and motif discovery.



Galaxy/Cistrome! is experiencing problems

Feb. 6th, 2013
We are currently investigating why jobs on Galaxy/Cistrome are failing. We hope to find the cause of the problem and fix it soon. We will update this page when that happens. Thanks for your bug reports and for your patience in this matter!

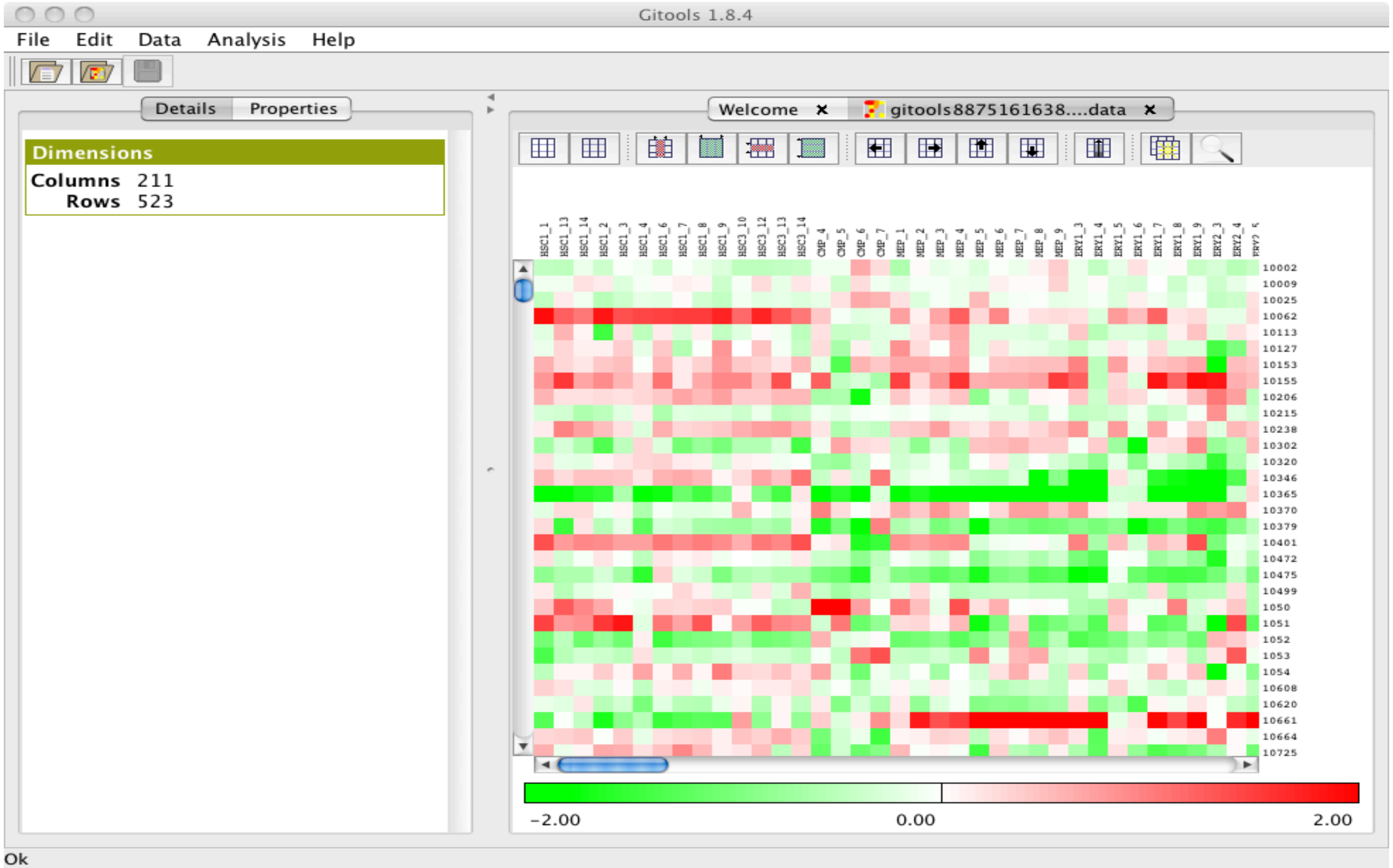
UPDATE: We believe that we found the cause of the problem. Everything should be back and functional. ****Please report the bugs if you are still experiencing problems****

Our server has a reboot at the beginning of this month, and that cause some problem with cistrome. Now it fixed after restart the cistrome. Sorry for the inconvenience.

Your history is empty. Click 'Get Data' on the left pane to start

Gitools

- Gitools is a framework for analysis and visualization of genomic data using interactive heatmaps.

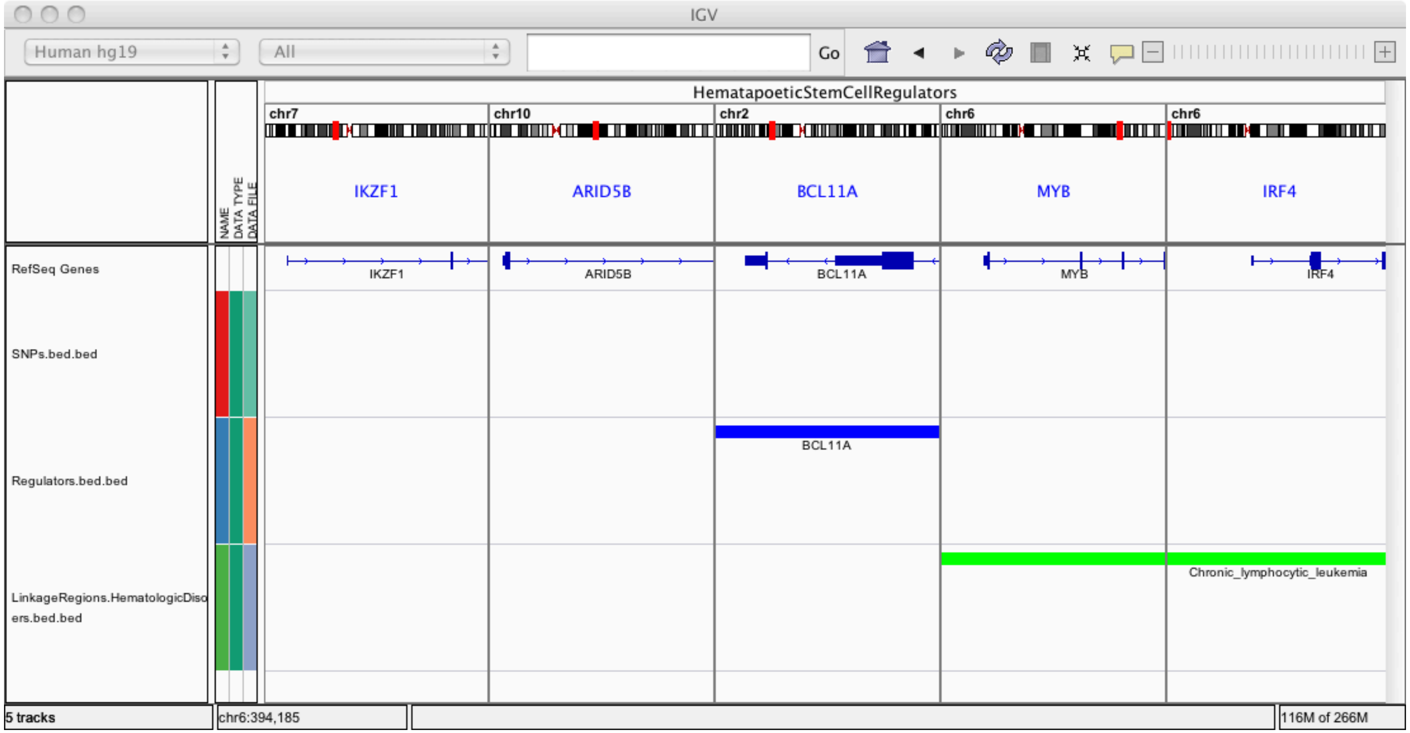


Dimensions

Columns 211
Rows 523

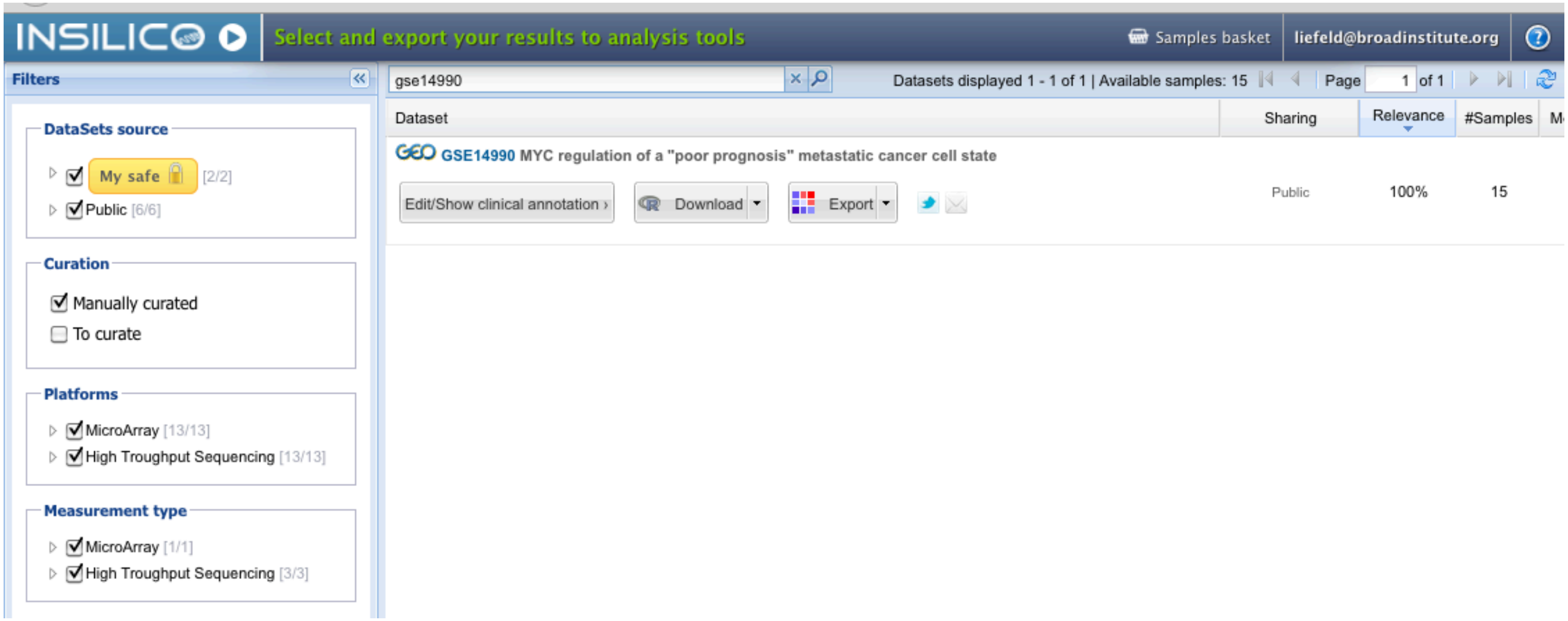
Integrative Genomics Viewer (IGV)

The **Integrative Genomics Viewer (IGV)** is a high-performance visualization tool for interactive exploration of large, integrated genomic datasets. It supports a wide variety of data types, including array-based and next-generation sequence data, and genomic annotations.



InSilicoDB

InSilico DB is a web-based genomics data manager containing thousands of curated public datasets. The datasets can be exported to analysis tools and GenomeSpace.



gse14990

Datasets displayed 1 - 1 of 1 | Available samples: 15 | Page 1 of 1

Dataset	Sharing	Relevance	#Samples
GEO GSE14990 MYC regulation of a "poor prognosis" metastatic cancer cell state	Public	100%	15

DataSets source

- ☑ My safe [2/2]
- ☑ Public [6/6]

Curation

- ☑ Manually curated
- ☐ To curate

Platforms

- ☑ MicroArray [13/13]
- ☑ High Troughput Sequencing [13/13]

Measurement type

- ☑ MicroArray [1/1]
- ☑ High Troughput Sequencing [3/3]

UCSC Table Browser

The Table Browser allows you to retrieve data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. After you select the options for your output file, you can opt to send your output file to your GenomeSpace cloud storage.

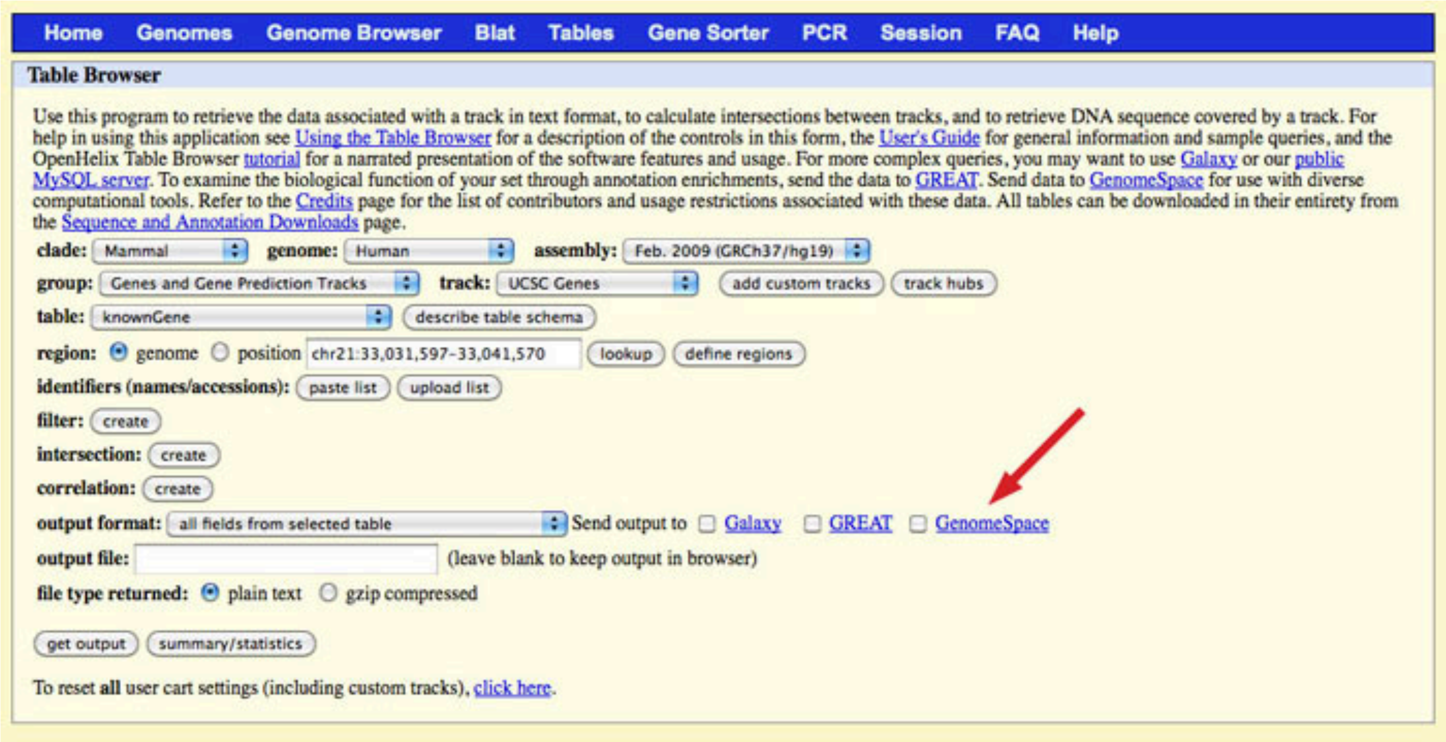


Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see Using the Table Browser for a description of the controls in this form, the User's Guide for general information and sample queries, and the OpenHelix Table Browser tutorial for a narrated presentation of the software features and usage. For more complex queries, you may want to use Galaxy or our public MySQL server. To examine the biological function of your set through annotation enrichments, send the data to GREAT. Send data to GenomeSpace for use with diverse computational tools. Refer to the Credits page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the Sequence and Annotation Downloads page.

clade: Mammal **genome:** Human **assembly:** Feb. 2009 (GRCh37/hg19)

group: Genes and Gene Prediction Tracks **track:** UCSC Genes

table: knownGene

region: ● genome ○ position chr21:33,031,597-33,041,570

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: all fields from selected table Send output to ☐ Galaxy ☐ GREAT ☐ GenomeSpace

output file: (leave blank to keep output in browser)

file type returned: ● plain text ○ gzip compressed

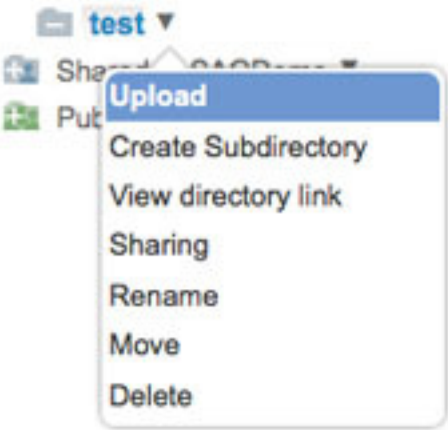
To reset **all** user cart settings (including custom tracks), click here.

Basic GenomeSpace recipes

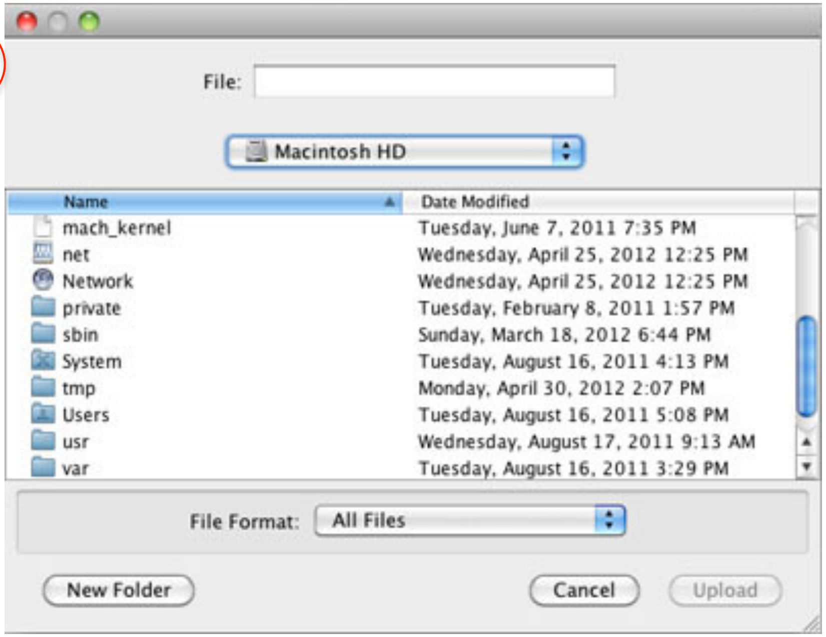
- Uploading data
- Launching tools
- Transitioning across tools

Uploading Data

①



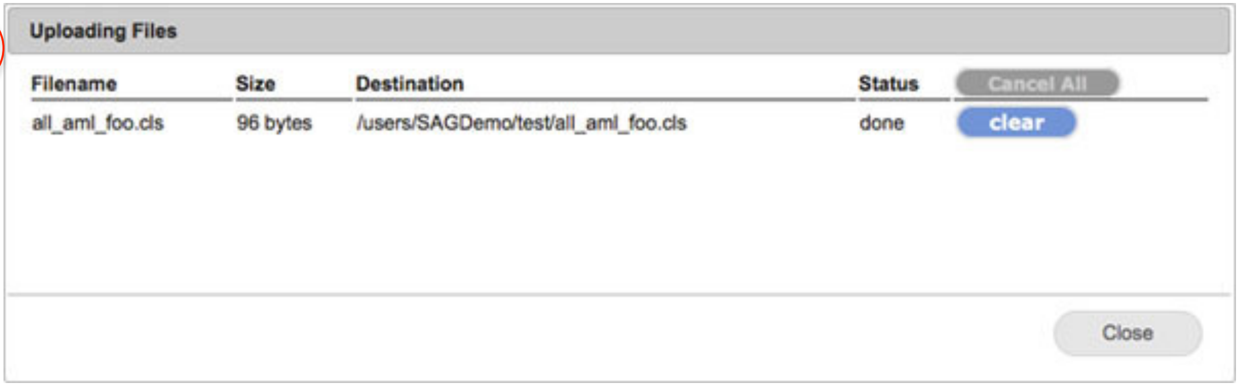
②



Name	Date Modified
mach_kernel	Tuesday, June 7, 2011 7:35 PM
net	Wednesday, April 25, 2012 12:25 PM
Network	Wednesday, April 25, 2012 12:25 PM
private	Tuesday, February 8, 2011 1:57 PM
sbin	Sunday, March 18, 2012 6:44 PM
System	Tuesday, August 16, 2011 4:13 PM
tmp	Monday, April 30, 2012 2:07 PM
Users	Tuesday, August 16, 2011 5:08 PM
usr	Wednesday, August 17, 2011 9:13 AM
var	Tuesday, August 16, 2011 3:29 PM

File Format: All Files

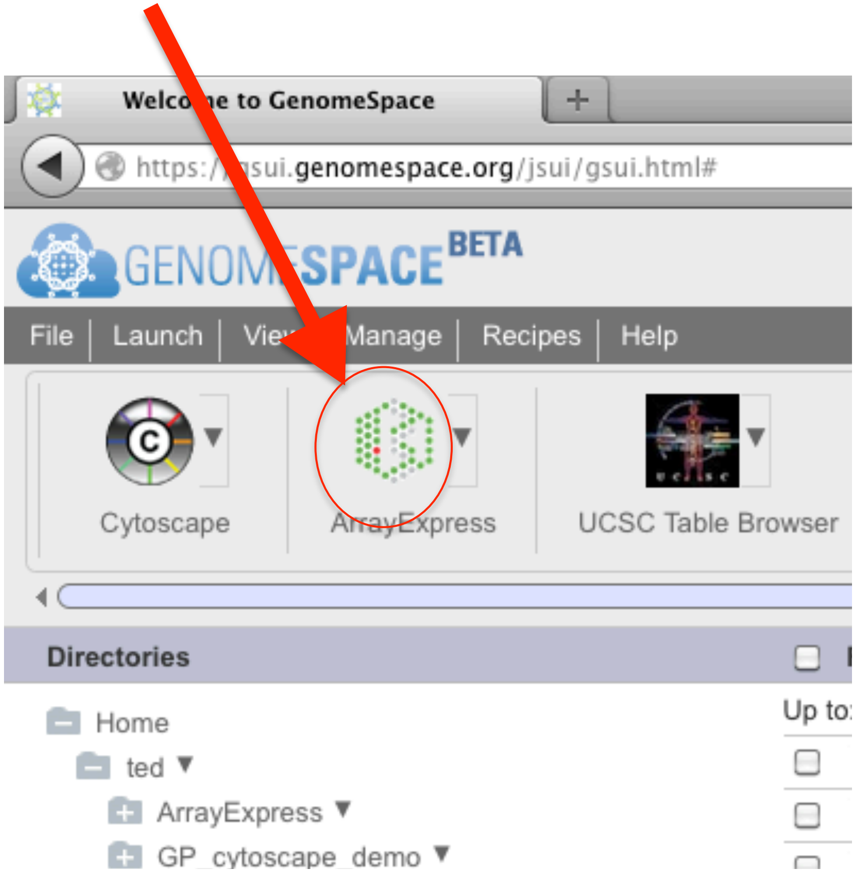
③



Filename	Size	Destination	Status
all_aml_foo.cls	96 bytes	/users/SAGDemo/test/all_aml_foo.cls	done

Launching tools

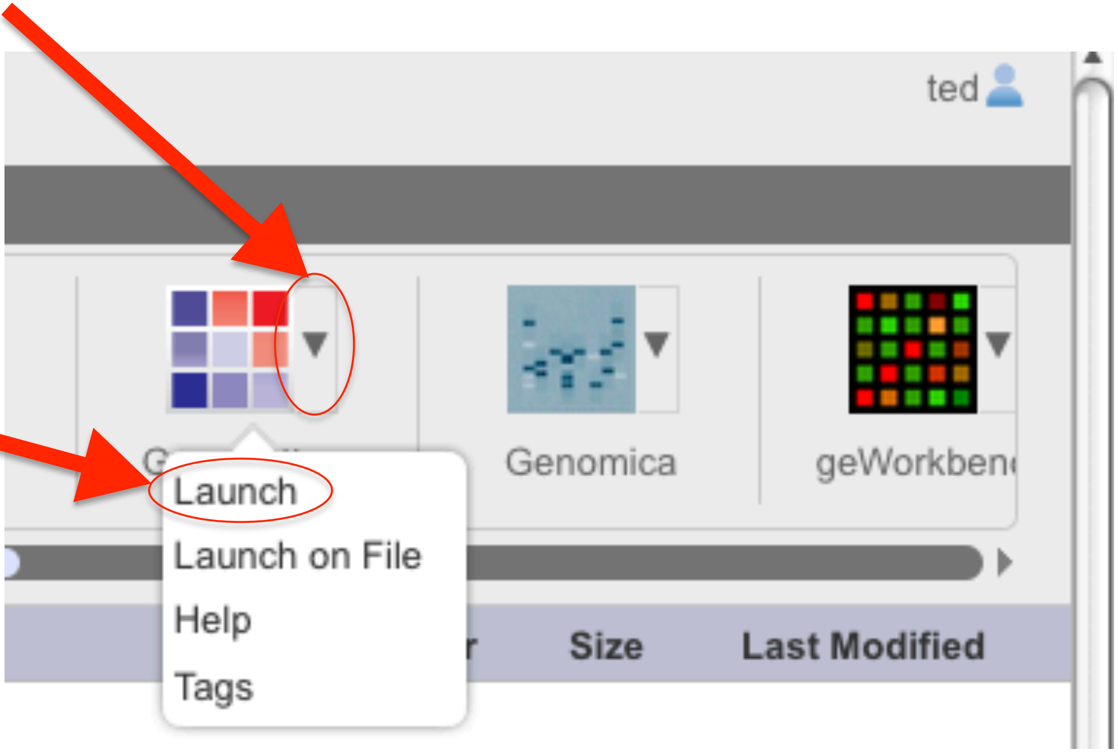
Click on the tool’s icon



Launching tools

Open the tool’s context menu

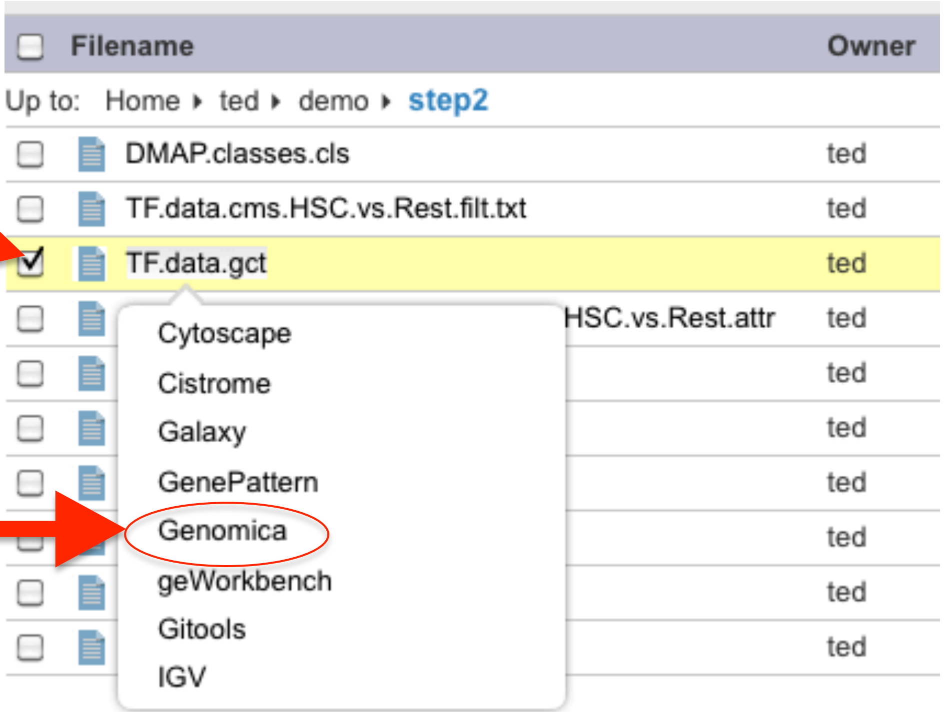
Then click on Launch
(or Launch on File)



Launching Tools

Click the checkbox for one (or more) files

Then click on one of The files to get the Launch menu and pick Your tool

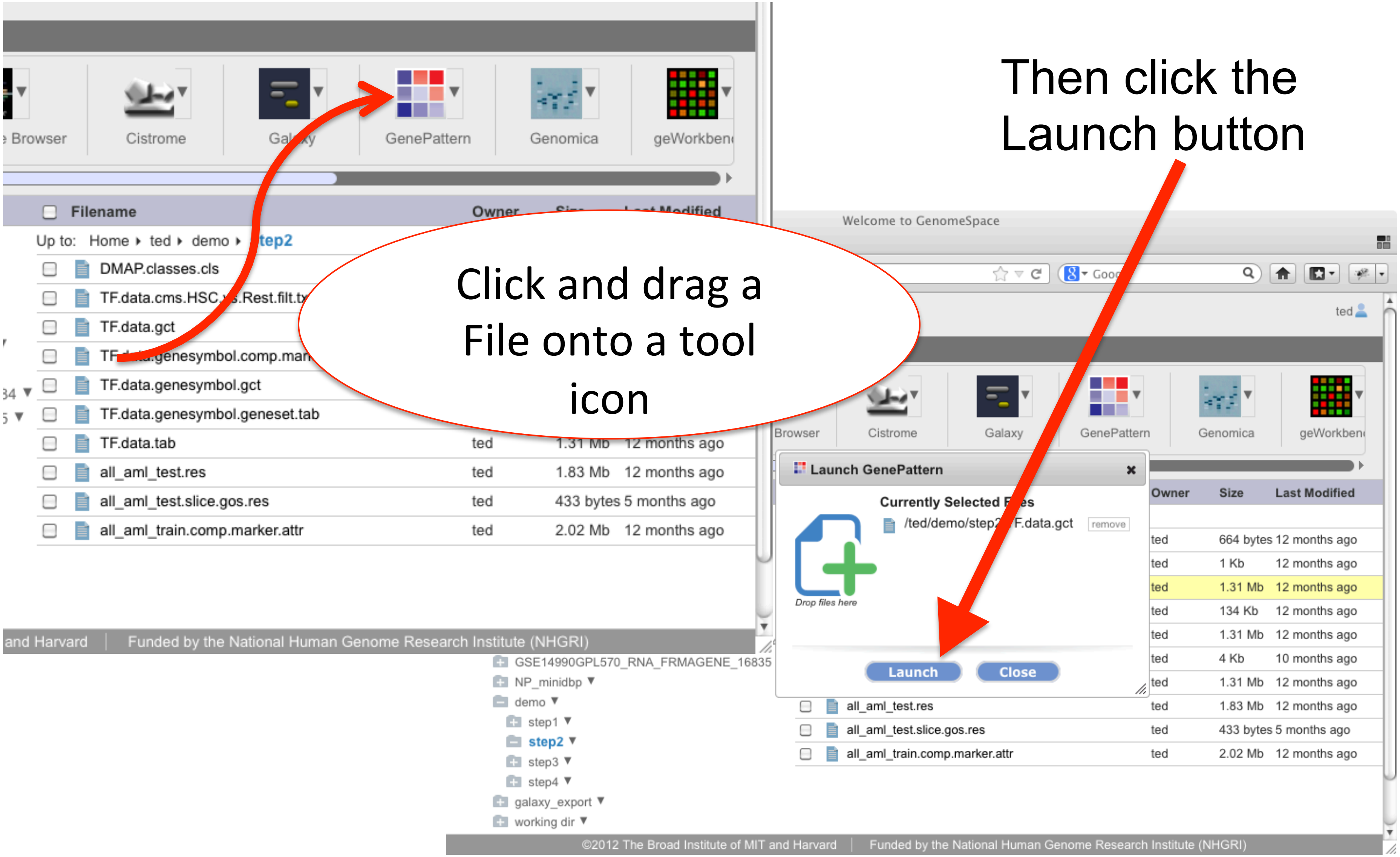


☐ **Filename** **Owner**

Up to: Home ▸ ted ▸ demo ▸ **step2**

	Filename	Owner
☐	DMAP.classes.cls	ted
☐	TF.data.cms.HSC.vs.Rest.filt.txt	ted
☑	TF.data.gct	ted
☐	HSC.vs.Rest.attr	ted
☐		ted
☐		ted
☐		ted
☐		ted
☐		ted
☐		ted

Launching tools



Then click the Launch button

Click and drag a File onto a tool icon

☐ Filename	Owner	Size	Last Modified
Up to: Home ▸ ted ▸ demo ▸ step2			
☐ DMAP.classes.cls			
☐ TF.data.cms.HSC.vs.Rest.filt.tx			
☐ TF.data.gct			
☐ TF.data.genesymbol.comp.mar			
☐ TF.data.genesymbol.gct			
☐ TF.data.genesymbol.geneset.tab			
☐ TF.data.tab	ted	1.31 Mb	12 months ago
☐ all_aml_test.res	ted	1.83 Mb	12 months ago
☐ all_aml_test.slice.gos.res	ted	433 bytes	5 months ago
☐ all_aml_train.comp.marker.attr	ted	2.02 Mb	12 months ago

and Harvard | Funded by the National Human Genome Research Institute (NHGRI)

- GSE14990GPL570_RNA_FRMAGENE_16835
- NP_minidbp ▾
- demo ▾
  - step1 ▾
  - **step2** ▾
  - step3 ▾
  - step4 ▾
- galaxy_export ▾
- working dir ▾

Welcome to GenomeSpace

Launch GenePattern ✖

Currently Selected Files

/ted/demo/step2/TF.data.gct remove

Drop files here

Launch | Close

Owner	Size	Last Modified
ted	664 bytes	12 months ago
ted	1 Kb	12 months ago
ted	1.31 Mb	12 months ago
ted	134 Kb	12 months ago
ted	1.31 Mb	12 months ago
ted	4 Kb	10 months ago
ted	1.31 Mb	12 months ago

Filename	Owner	Size	Last Modified
☐ all_aml_test.res	ted	1.83 Mb	12 months ago
☐ all_aml_test.slice.gos.res	ted	433 bytes	5 months ago
☐ all_aml_train.comp.marker.attr	ted	2.02 Mb	12 months ago

©2012 The Broad Institute of MIT and Harvard | Funded by the National Human Genome Research Institute (NHGRI)

Transitioning across tools

1. Launch Genomica
    - Load (shared) data from GenomeSpace
    - Save it back to a new folder
2. Launch GenePattern on your data
    - Do a simple processing step
    - Save it back to GenomeSpace
    - Send it to IGV
3. Visualize the procesed data IGV

Launch Genomica

- Using one of the options you saw earlier
  - – Click on the icon
  - – or use the context menu
  - – or use the launch menu
- Load data from GenomeSpace

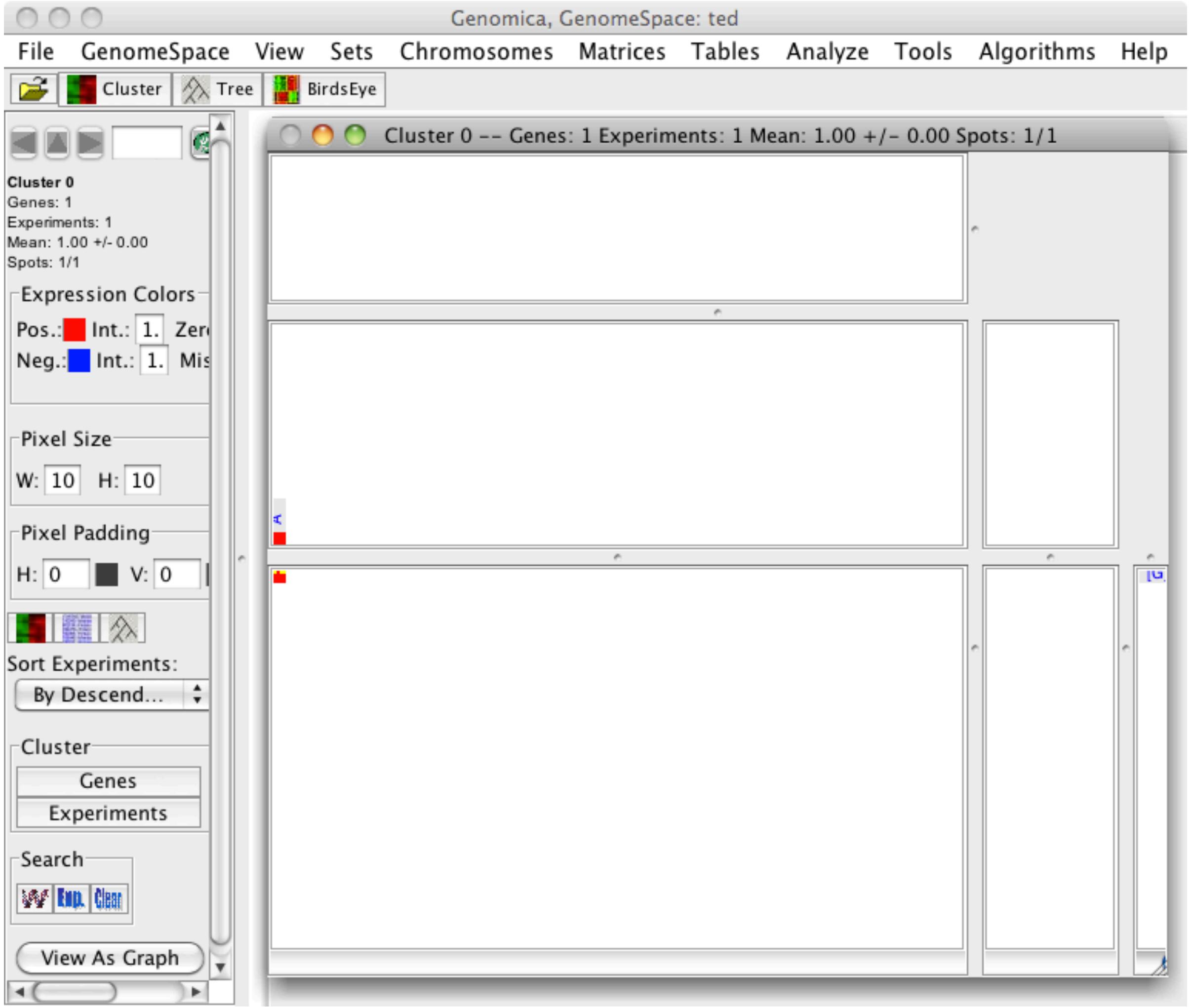
Home ▸ Public ▸ SharedData ▸ Demos ▸ Scenario ▸ step3 ▸ **80_module.gxp**

Or

Home ▸ Shared to <your id> ▸ mmr ▸ FGED ▸ **80_module.gxp**

Loading into Genomica

Home ▸ Shared to <your id> ▸ mmr ▸ FGED ▸ **80_module.gxp**



Cluster 0 -- Genes: 1 Experiments: 1 Mean: 1.00 +/- 0.00 Spots: 1/1

Cluster 0
Genes: 1
Experiments: 1
Mean: 1.00 +/- 0.00
Spots: 1/1

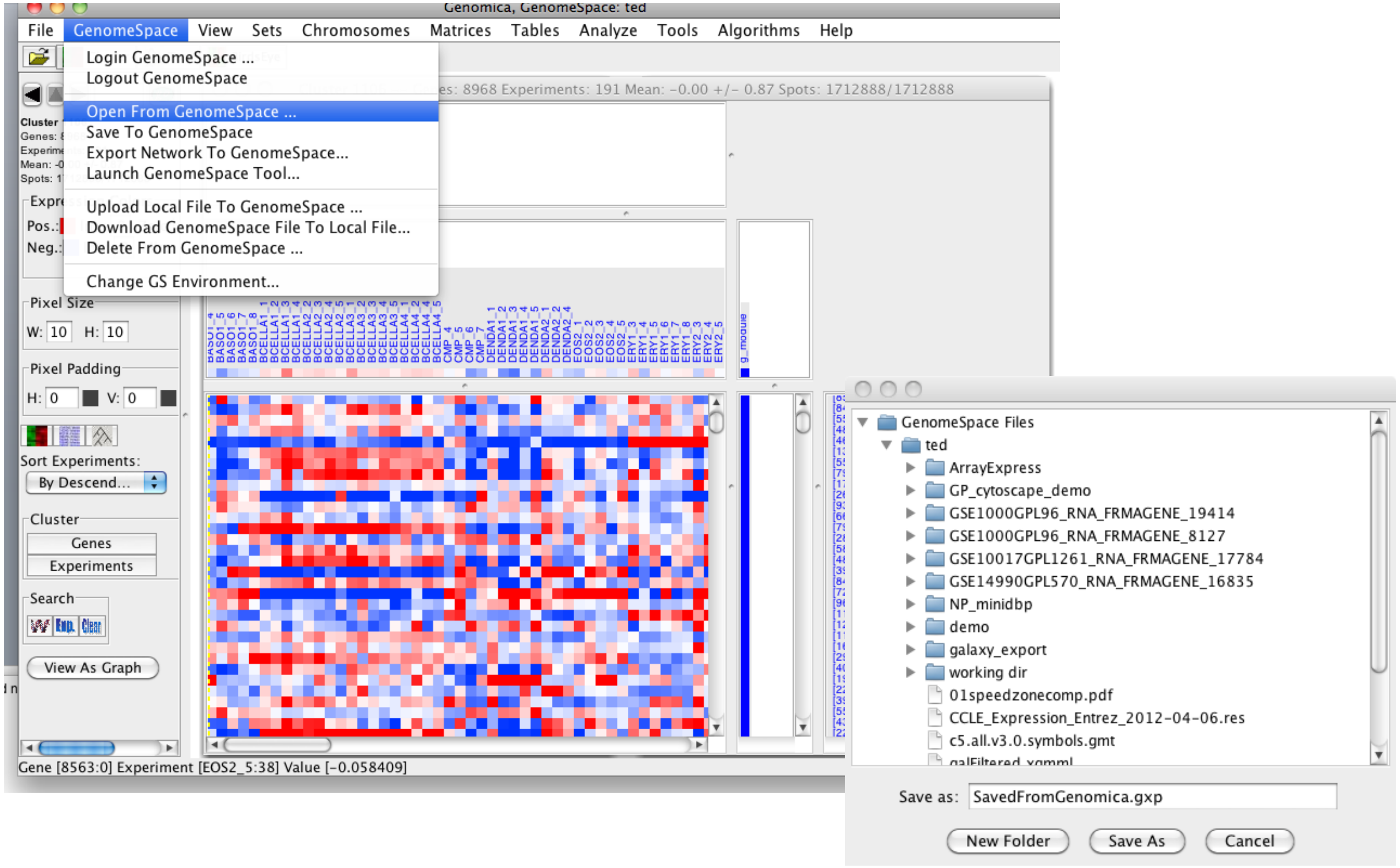
Expression Colors
Pos.: Int.: 1.
Neg.: Int.: 1.

Pixel Size
W: 10 H: 10

Pixel Padding
H: 0 V: 0

Sort Experiments: By Descend...

Saving Back to GenomeSpace



es: 8968 Experiments: 191 Mean: -0.00 +/- 0.87 Spots: 1712888/1712888

W: 10 H: 10

H: 0 V: 0

Gene [8563:0] Experiment [EOS2_5:38] Value [-0.058409]

Save as: SavedFromGenomica.gxp

Launching GenePattern

- You can do this from within Genomica or also from the GenomeSpace interface
- Select "PreprocessDataset" in the send to module

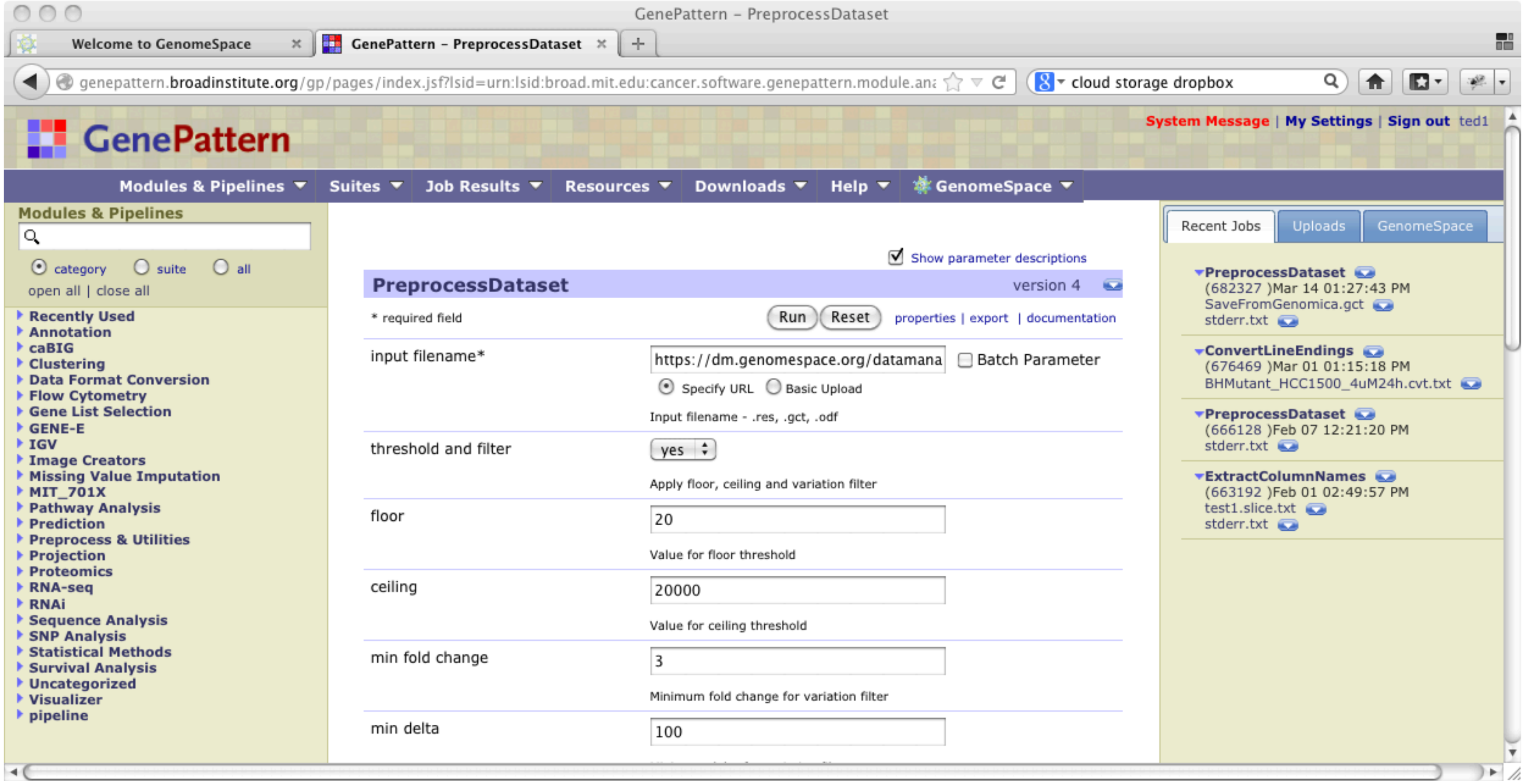
GenomeSpace has sent you a file

SaveFromGenomica.gxp *Converted to gct*
Save a Copy: Upload Directory Submit

Send to Module: PreprocessDataset Submit

Process the data

- Run PreprocessDataset with default parameters



☑ Show parameter descriptions

PreprocessDataset version 4

* required field

input filename*: https://dm.genomespace.org/datamana ☐ Batch Parameter
● Specify URL ○ Basic Upload
Input filename - .res, .gct, .odf

threshold and filter: yes
Apply floor, ceiling and variation filter

floor: 20
Value for floor threshold

ceiling: 20000
Value for ceiling threshold

min fold change: 3
Minimum fold change for variation filter

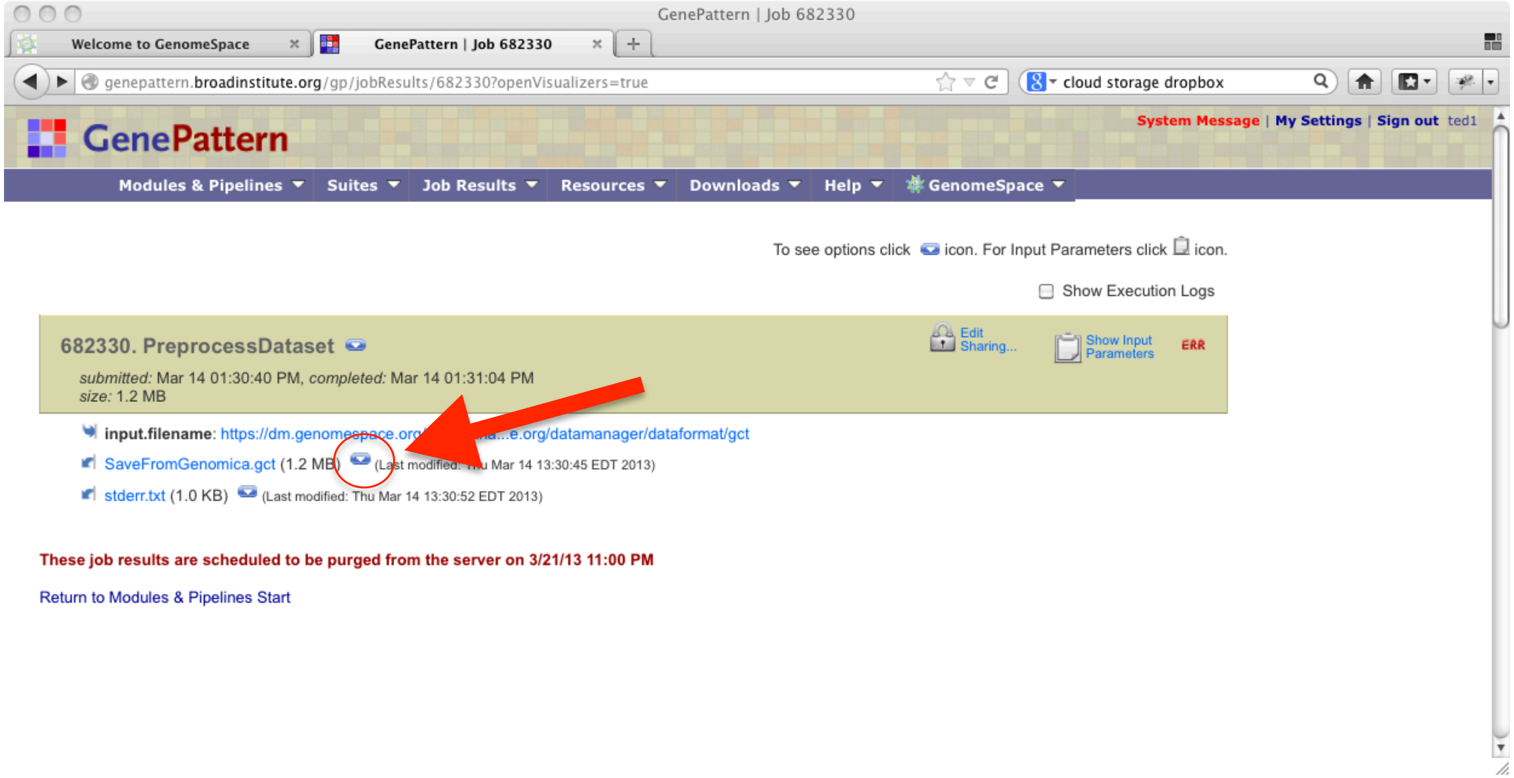
min delta: 100

Recent Jobs:

- PreprocessDataset (682327)Mar 14 01:27:43 PM
  - SaveFromGenomica.gct
  - stderr.txt
- ConvertLineEndings (676469)Mar 01 01:15:18 PM
  - BHMutant_HCC1500_4uM24h.cvt.txt
- PreprocessDataset (666128)Feb 07 12:21:20 PM
  - stderr.txt
- ExtractColumnNames (663192)Feb 01 02:49:57 PM
  - test1.slice.txt
  - stderr.txt

Save the result

Use the context menu for the file on either the job result page …



To see options click icon. For Input Parameters click icon.

☐ Show Execution Logs

682330. PreprocessDataset

submitted: Mar 14 01:30:40 PM, *completed:* Mar 14 01:31:04 PM
size: 1.2 MB

input.filename: https://dm.genomespace.org ... a...e.org/datamanager/dataformat/gct

SaveFromGenomica.gct (1.2 MB) (Last modified: Thu Mar 14 13:30:45 EDT 2013)

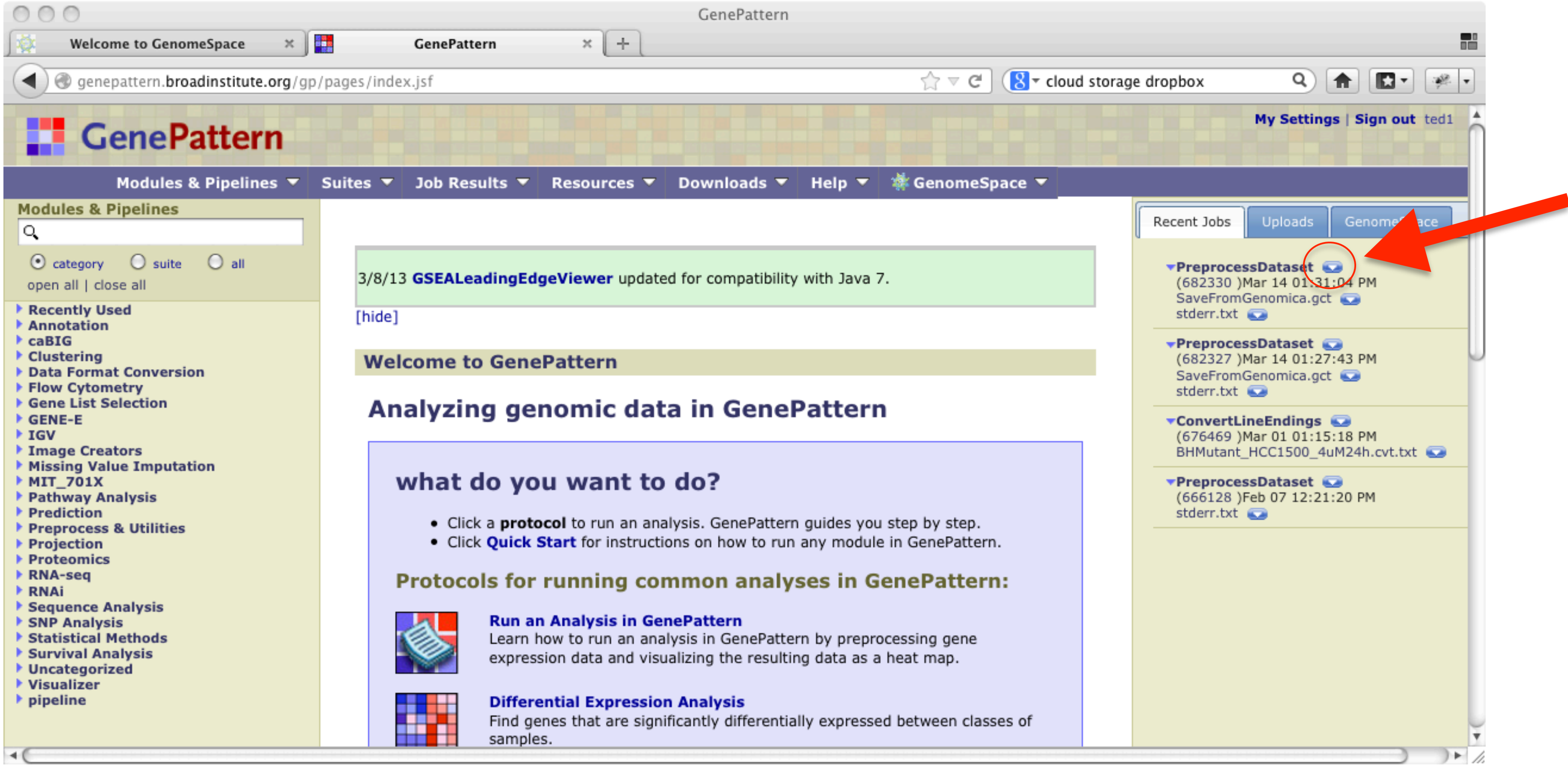
stderr.txt (1.0 KB) (Last modified: Thu Mar 14 13:30:52 EDT 2013)

These job results are scheduled to be purged from the server on 3/21/13 11:00 PM

Return to Modules & Pipelines Start

Save the result

…or the context menu for the file on the GenePattern home page.



3/8/13 **GSEALeadingEdgeViewer** updated for compatibility with Java 7.

[hide]

Welcome to GenePattern

Analyzing genomic data in GenePattern

what do you want to do?

- Click a **protocol** to run an analysis. GenePattern guides you step by step.
- Click **Quick Start** for instructions on how to run any module in GenePattern.

Protocols for running common analyses in GenePattern:

Run an Analysis in GenePattern
Learn how to run an analysis in GenePattern by preprocessing gene expression data and visualizing the resulting data as a heat map.

Differential Expression Analysis
Find genes that are significantly differentially expressed between classes of samples.

PreprocessDataset
(682330)Mar 14 01:31:04 PM
SaveFromGenomica.gct
stderr.txt

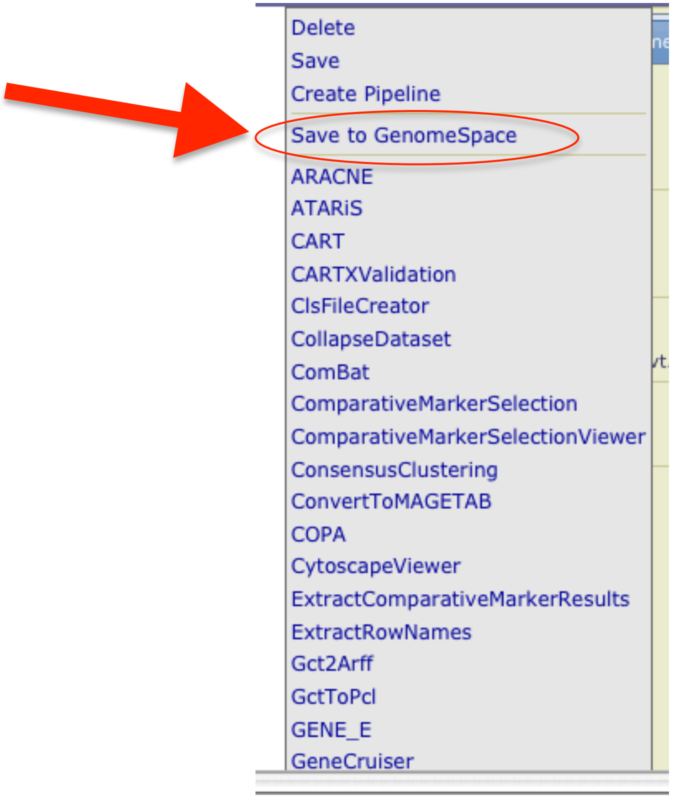
PreprocessDataset
(682327)Mar 14 01:27:43 PM
SaveFromGenomica.gct
stderr.txt

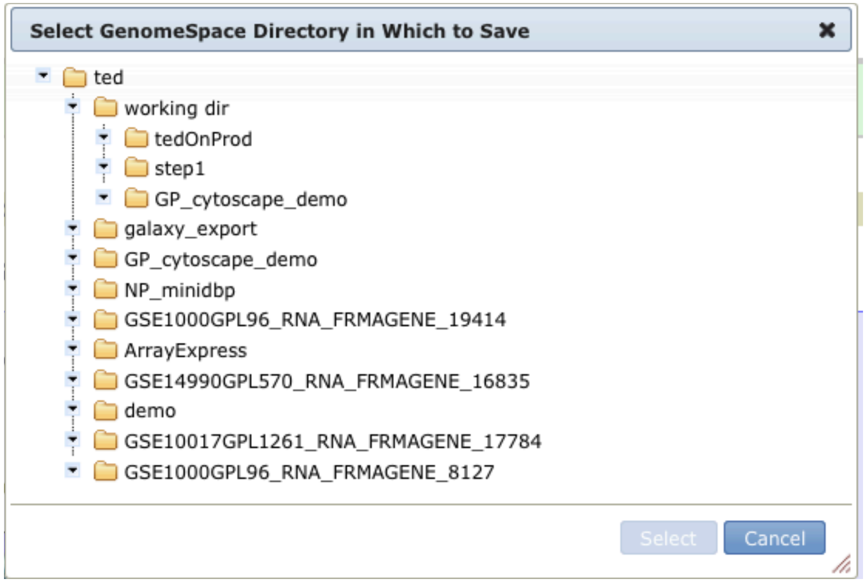
ConvertLineEndings
(676469)Mar 01 01:15:18 PM
BHMutant_HCC1500_4uM24h.cvt.txt

PreprocessDataset
(666128)Feb 07 12:21:20 PM
stderr.txt

Saving to GenomeSpace

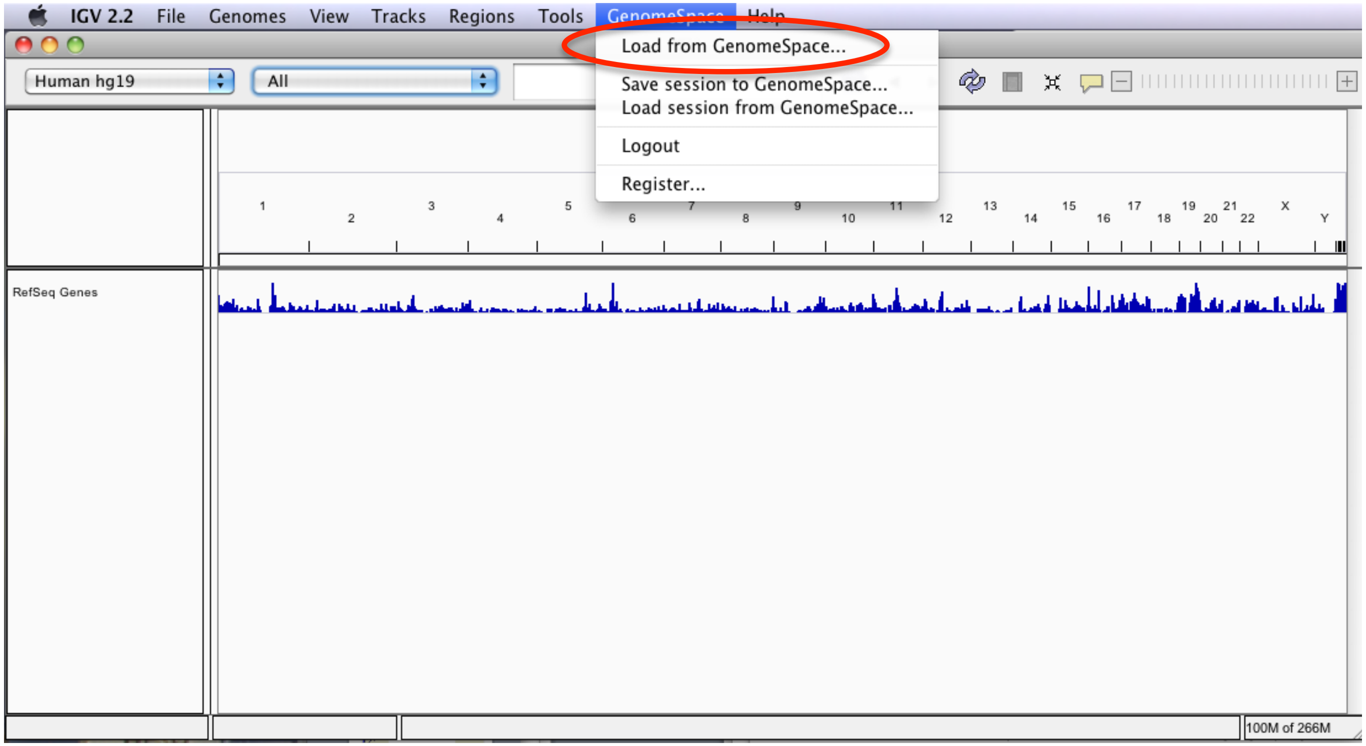
Click “Save to GenomeSpace” from the context menu and then select a targetdirectory



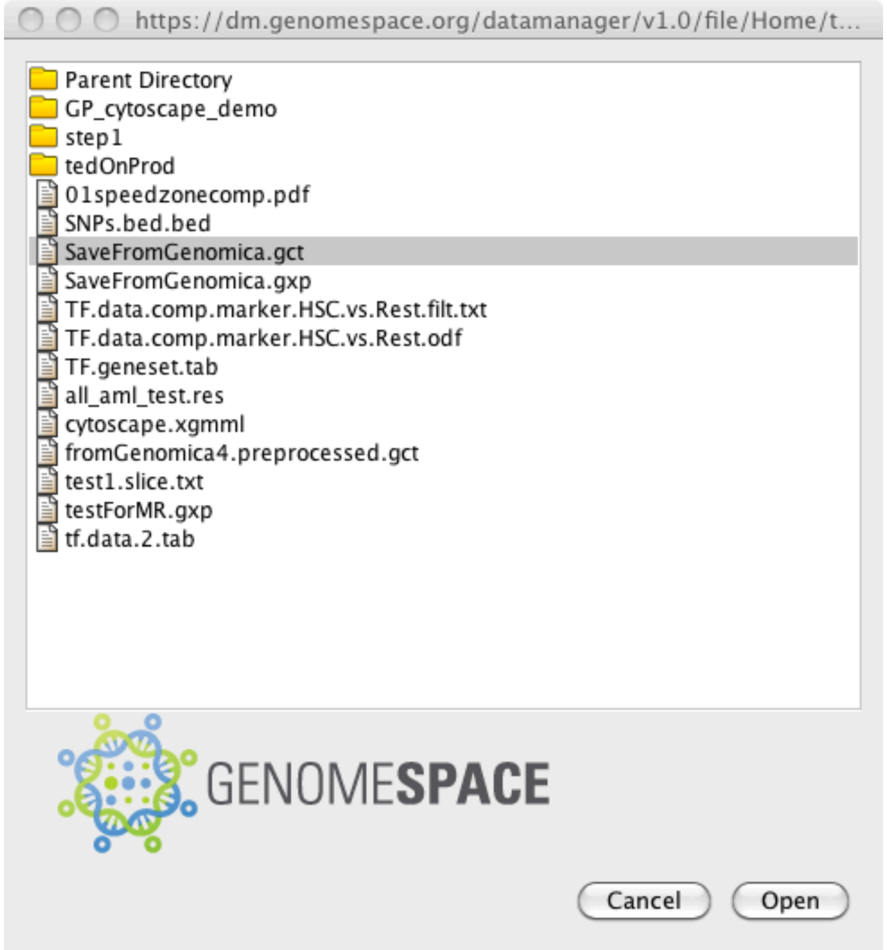


Send to IGV

- In the GenomeSpace interface, launch IGV
  - Open the 'GenomeSpace' menu and 'Load from GenomeSpace'



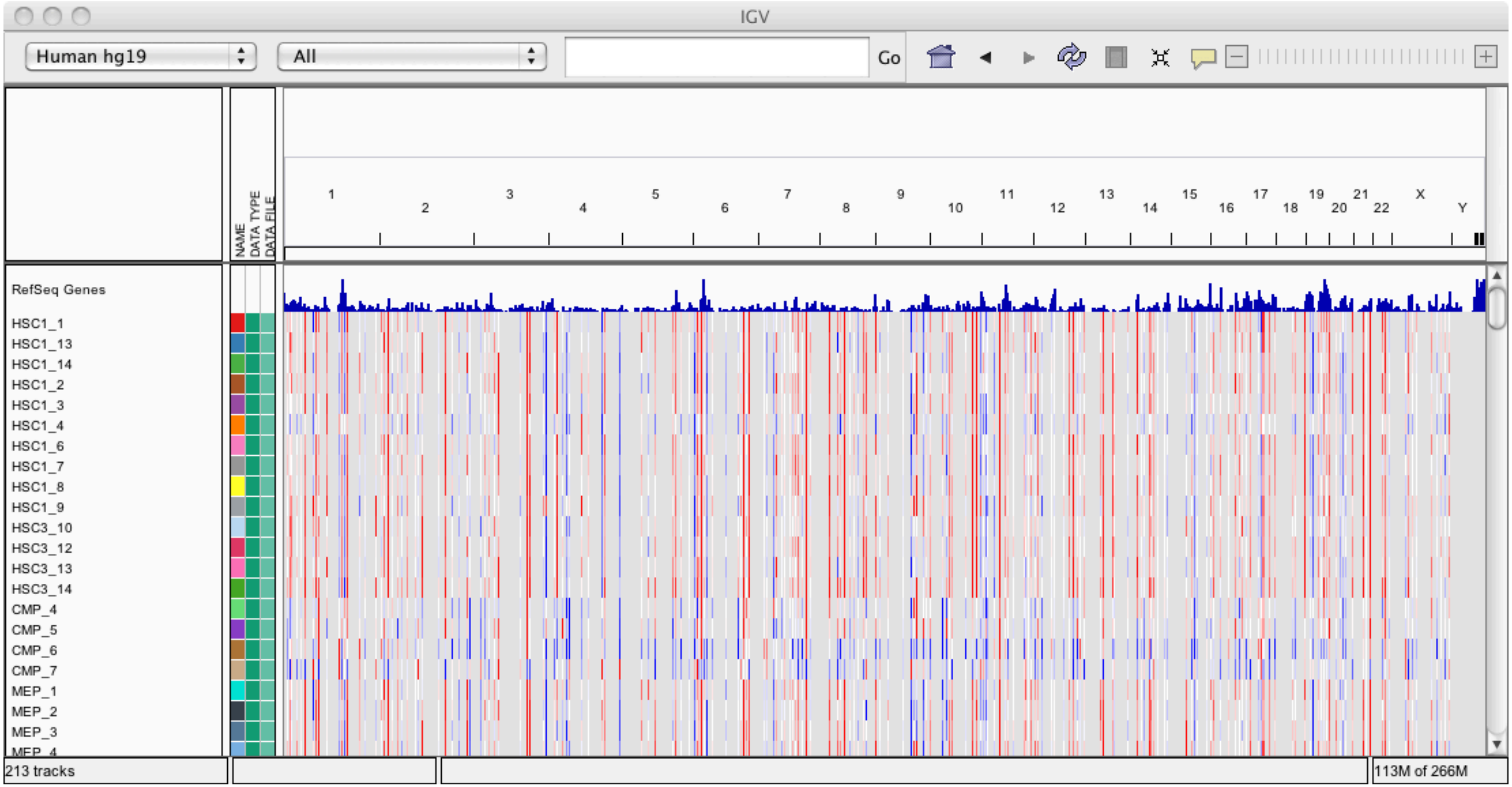
Select your file (from GenePattern)



https://dm.genomespace.org/datamanager/v1.0/file/Home/t...

- Parent Directory
- GP_cytoscape_demo
- step1
- tedOnProd
- 01speedzonecomp.pdf
- SNPs.bed.bed
- SaveFromGenomica.gct
- SaveFromGenomica.gxp
- TF.data.comp.marker.HSC.vs.Rest.filt.txt
- TF.data.comp.marker.HSC.vs.Rest.odf
- TF.geneset.tab
- all_aml_test.res
- cytoscape.xgmml
- fromGenomica4.preprocessed.gct
- test1.slice.txt
- testForMR.gxp
- tf.data.2.tab

Visualize in IGV



GenomeSpace UI

A detailed tour of the GenomeSpace User Interface

Agenda

- File Management
- File operations
- Sharing with others
- Organizing your tools

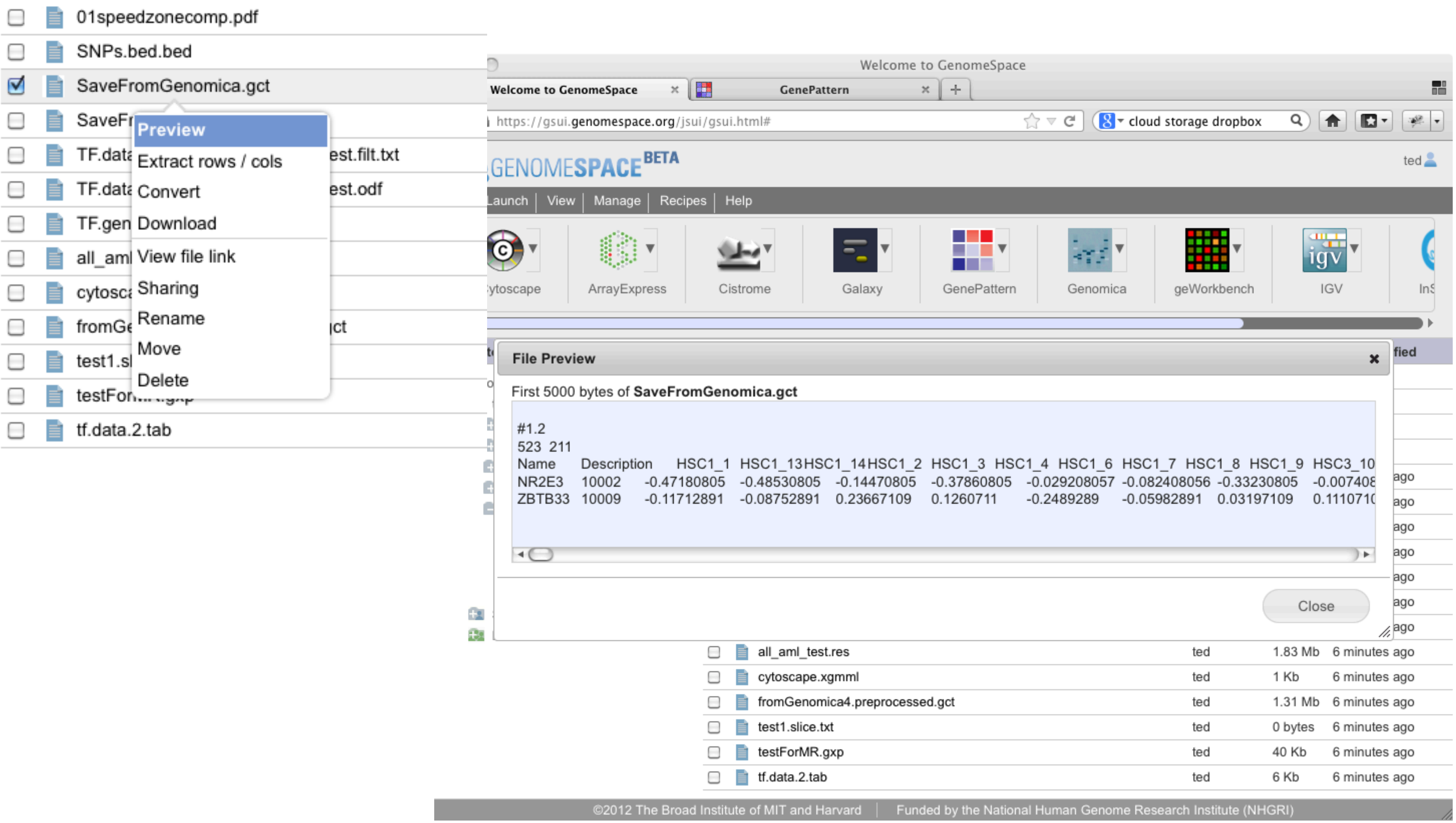
File Management

- Move a file or directory
- Copy …
- Deleting …
- Creating subdirectories
- Recent uploads

File Operations

- Previewing a file
- Extracting rows and/or columns
- Format conversion

File Preview



☐ 01speedzonecomp.pdf
☐ SNPs.bed.bed
☑ SaveFromGenomica.gct
☐ SaveFr
☐ TF.data … est.filt.txt
☐ TF.data … est.odf
☐ TF.gen
☐ all_aml
☐ cytosca
☐ fromGe … ct
☐ test1.sl
☐ testFor
☐ tf.data.2.tab

File Preview

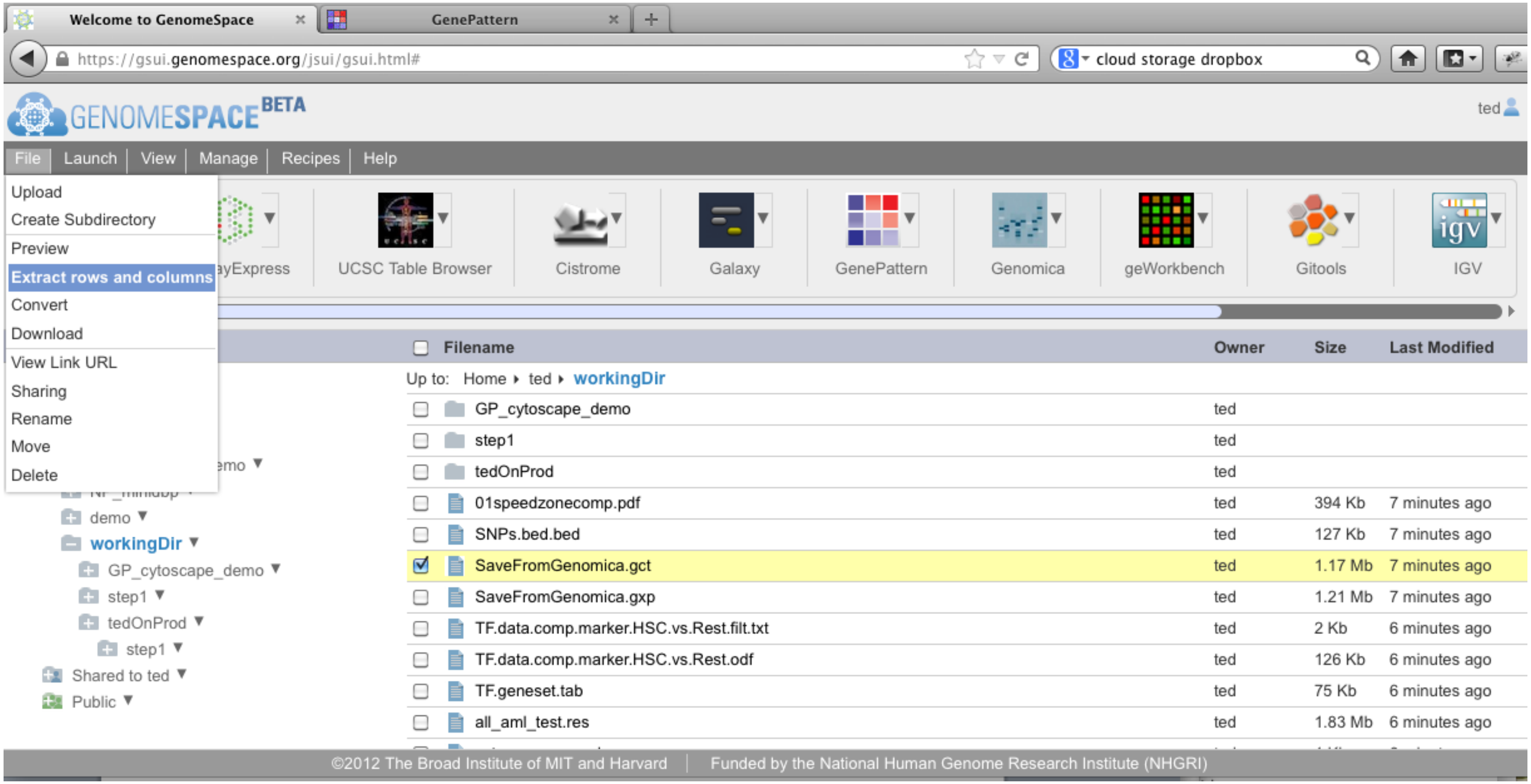
First 5000 bytes of **SaveFromGenomica.gct**

```
#1.2
523  211
Name    Description    HSC1_1   HSC1_13HSC1_14HSC1_2  HSC1_3   HSC1_4   HSC1_6    HSC1_7    HSC1_8   HSC1_9   HSC3_10
NR2E3   10002    -0.47180805  -0.48530805  -0.14470805  -0.37860805  -0.029208057 -0.082408056 -0.33230805  -0.007408
ZBTB33  10009    -0.11712891  -0.08752891  0.23667109   0.1260711    -0.2489289   -0.05982891  0.03197109   0.111071
```

☐	all_aml_test.res	ted	1.83 Mb	6 minutes ago
☐	cytoscape.xgmml	ted	1 Kb	6 minutes ago
☐	fromGenomica4.preprocessed.gct	ted	1.31 Mb	6 minutes ago
☐	test1.slice.txt	ted	0 bytes	6 minutes ago
☐	testForMR.gxp	ted	40 Kb	6 minutes ago
☐	tf.data.2.tab	ted	6 Kb	6 minutes ago

©2012 The Broad Institute of MIT and Harvard | Funded by the National Human Genome Research Institute (NHGRI)

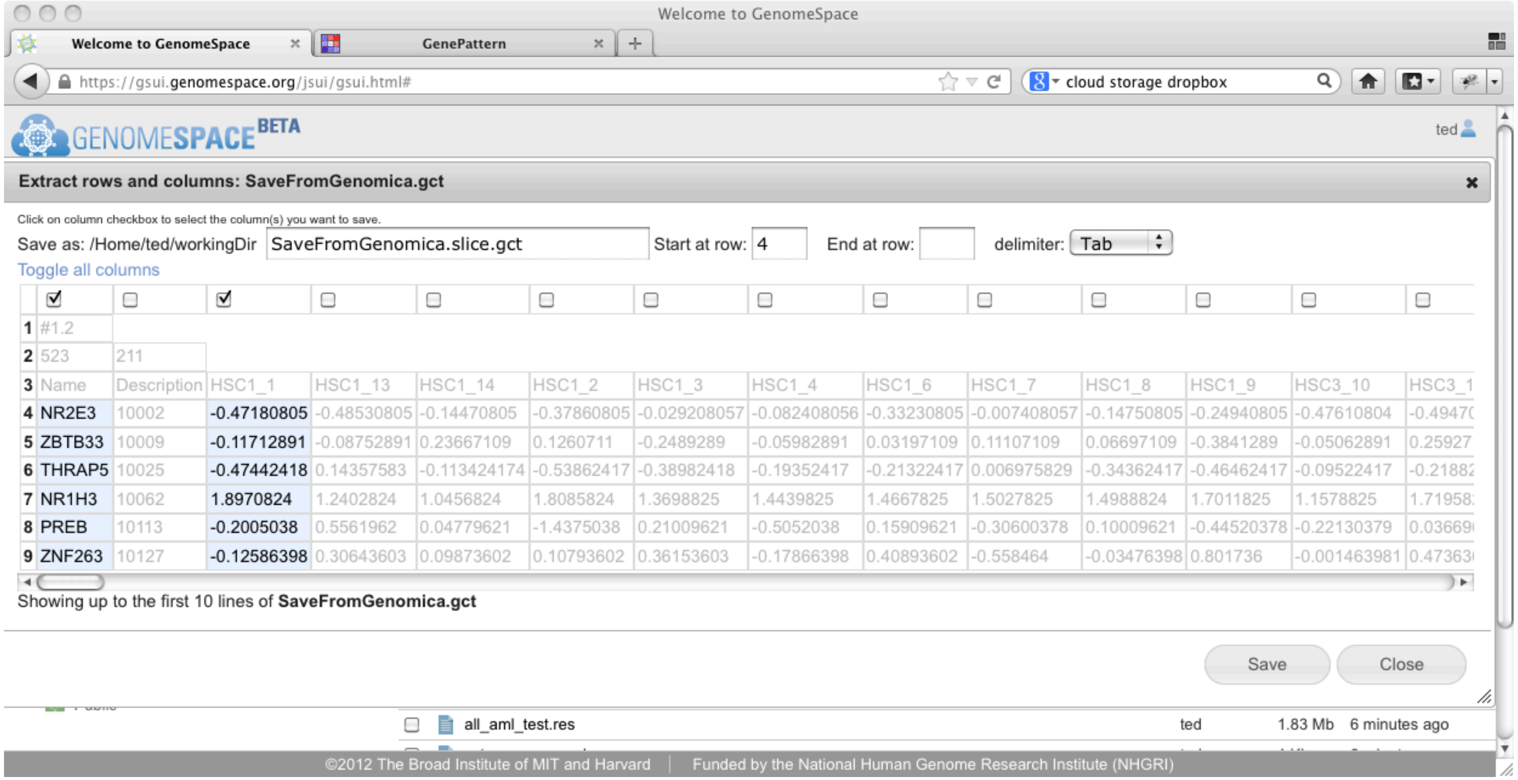
Extracting Rows and/or Columns



☐	Filename	Owner	Size	Last Modified
☐	GP_cytoscape_demo	ted		
☐	step1	ted		
☐	tedOnProd	ted		
☐	01speedzonecomp.pdf	ted	394 Kb	7 minutes ago
☐	SNPs.bed.bed	ted	127 Kb	7 minutes ago
☑	SaveFromGenomica.gct	ted	1.17 Mb	7 minutes ago
☐	SaveFromGenomica.gxp	ted	1.21 Mb	7 minutes ago
☐	TF.data.comp.marker.HSC.vs.Rest.filt.txt	ted	2 Kb	6 minutes ago
☐	TF.data.comp.marker.HSC.vs.Rest.odf	ted	126 Kb	6 minutes ago
☐	TF.geneset.tab	ted	75 Kb	6 minutes ago
☐	all_aml_test.res	ted	1.83 Mb	6 minutes ago

Extracting rows and/or columns

- Check the columns you want to include
- Provide a first (and optionally last) row index to include
- Edit the file name and ‘Save’



Extract rows and columns: SaveFromGenomica.gct

Click on column checkbox to select the column(s) you want to save.

Save as: /Home/ted/workingDir SaveFromGenomica.slice.gct Start at row: 4 End at row: delimiter: Tab

Toggle all columns

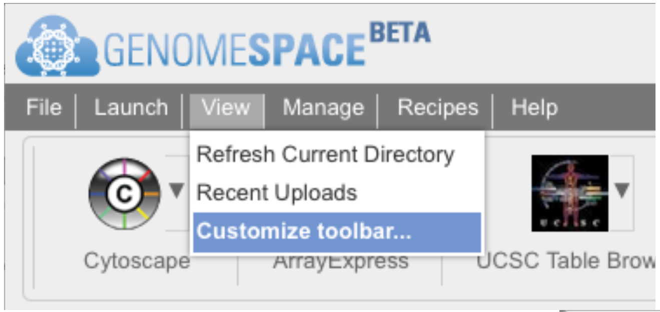
	☑	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
1	#1.2													
2	523	211												
3	Name	Description	HSC1_1	HSC1_13	HSC1_14	HSC1_2	HSC1_3	HSC1_4	HSC1_6	HSC1_7	HSC1_8	HSC1_9	HSC3_10	HSC3_1
4	NR2E3	10002	-0.47180805	-0.48530805	-0.14470805	-0.37860805	-0.029208057	-0.082408056	-0.33230805	-0.007408057	-0.14750805	-0.24940805	-0.47610804	-0.4947
5	ZBTB33	10009	-0.11712891	-0.08752891	0.23667109	0.1260711	-0.2489289	-0.05982891	0.03197109	0.11107109	0.06697109	-0.3841289	-0.05062891	0.25927
6	THRAP5	10025	-0.47442418	0.14357583	-0.113424174	-0.53862417	-0.38982418	-0.19352417	-0.21322417	0.006975829	-0.34362417	-0.46462417	-0.09522417	-0.2188
7	NR1H3	10062	1.8970824	1.2402824	1.0456824	1.8085824	1.3698825	1.4439825	1.4667825	1.5027825	1.4988824	1.7011825	1.1578825	1.7195
8	PREB	10113	-0.2005038	0.5561962	0.04779621	-1.4375038	0.21009621	-0.5052038	0.15909621	-0.30600378	0.10009621	-0.44520378	-0.22130379	0.0366
9	ZNF263	10127	-0.12586398	0.30643603	0.09873602	0.10793602	0.36153603	-0.17866398	0.40893602	-0.558464	-0.03476398	0.801736	-0.001463981	0.4736

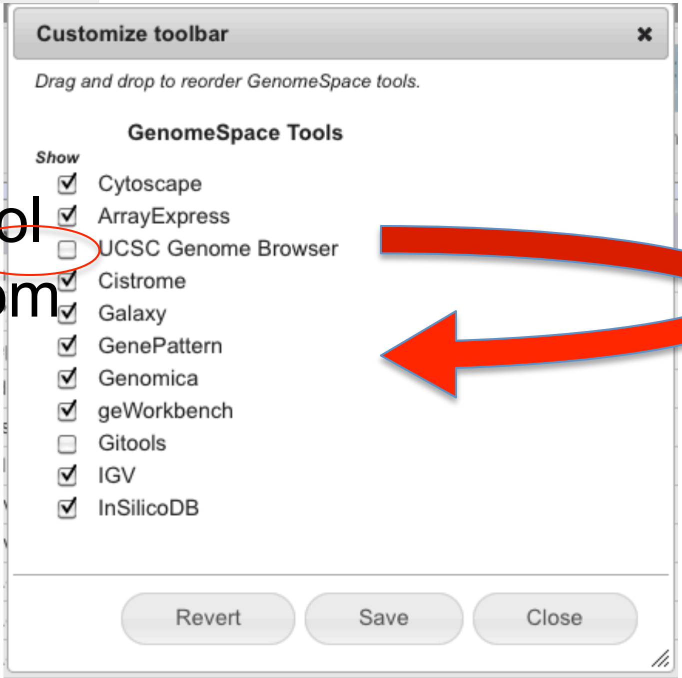
Showing up to the first 10 lines of **SaveFromGenomica.gct**

Sharing with others

- Sharing files with
  - Individuals, groups
    - Creating groups for sharing
- Sharing links
  - With other GenomeSpace users
  - To people without GenomeSpace accounts

Organizing tools




Customize toolbar

Drag and drop to reorder GenomeSpace tools.

GenomeSpace Tools

Show

- ☑ Cytoscape
- ☑ ArrayExpress
- ☐ UCSC Genome Browser
- ☑ Cistrome
- ☑ Galaxy
- ☑ GenePattern
- ☑ Genomica
- ☑ geWorkbench
- ☐ Gitools
- ☑ IGV
- ☑ InSilicoDB

Uncheck the tool
To remove it from
The toolbar

Drag and drop
tools in the list
reorder them

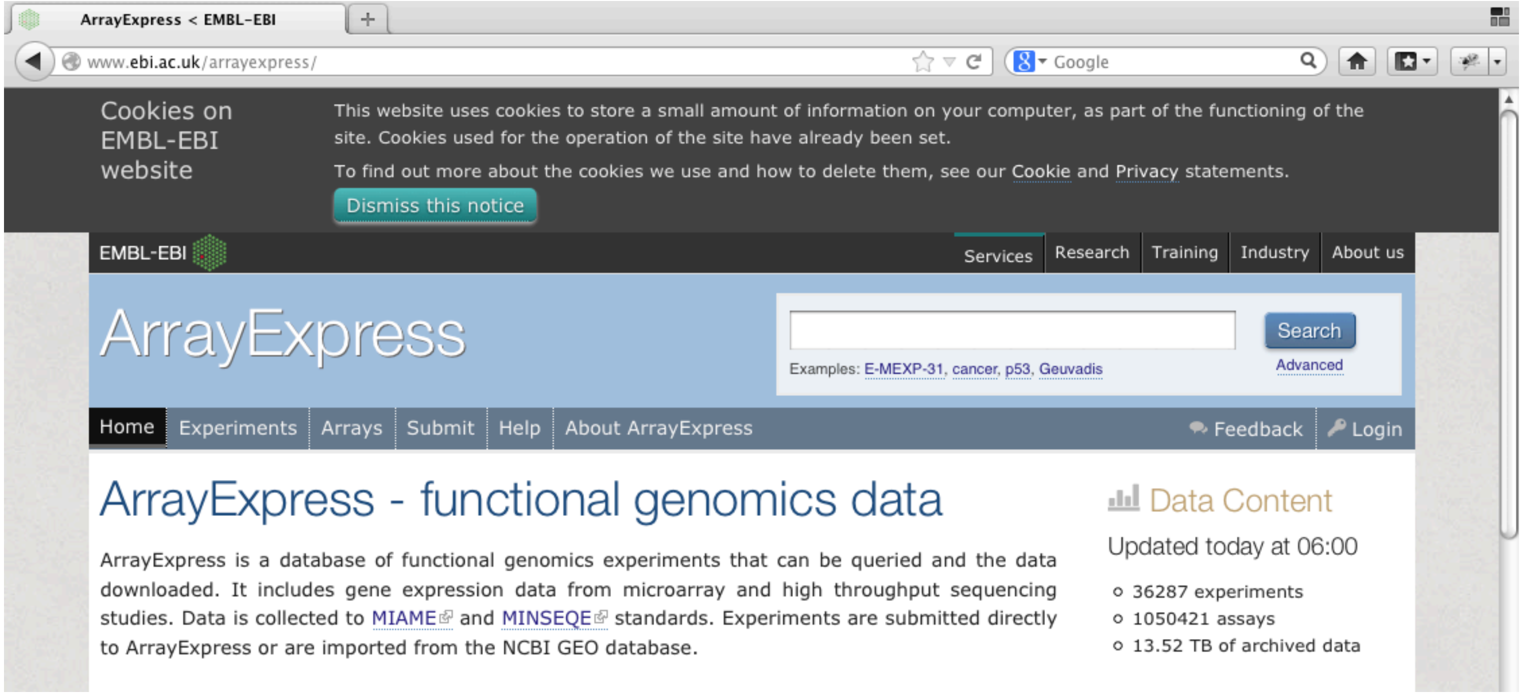
Other GenomeSpace Tools

- ArrayExpress
- Galaxy
- Cistrome
- Cytoscape
- GenePattern
- Genomica
- ISAcreator

- geWorkbench
- Gitools
- IGV
- InSilicoDB
- UCSC Table Browser
- MSigDB

ArrayExpress

- Repository of over 30,000 gene expression and other functional genomics experiments comprising nearly 1 million assays.
- Query and retrieve data in a number of different formats including MIAME and MINSEQE.



Cookies on EMBL-EBI website

This website uses cookies to store a small amount of information on your computer, as part of the functioning of the site. Cookies used for the operation of the site have already been set.

To find out more about the cookies we use and how to delete them, see our Cookie and Privacy statements.

Examples: E-MEXP-31, cancer, p53, Geuvadis

ArrayExpress - functional genomics data

ArrayExpress is a database of functional genomics experiments that can be queried and the data downloaded. It includes gene expression data from microarray and high throughput sequencing studies. Data is collected to MIAME and MINSEQE standards. Experiments are submitted directly to ArrayExpress or are imported from the NCBI GEO database.

Data Content

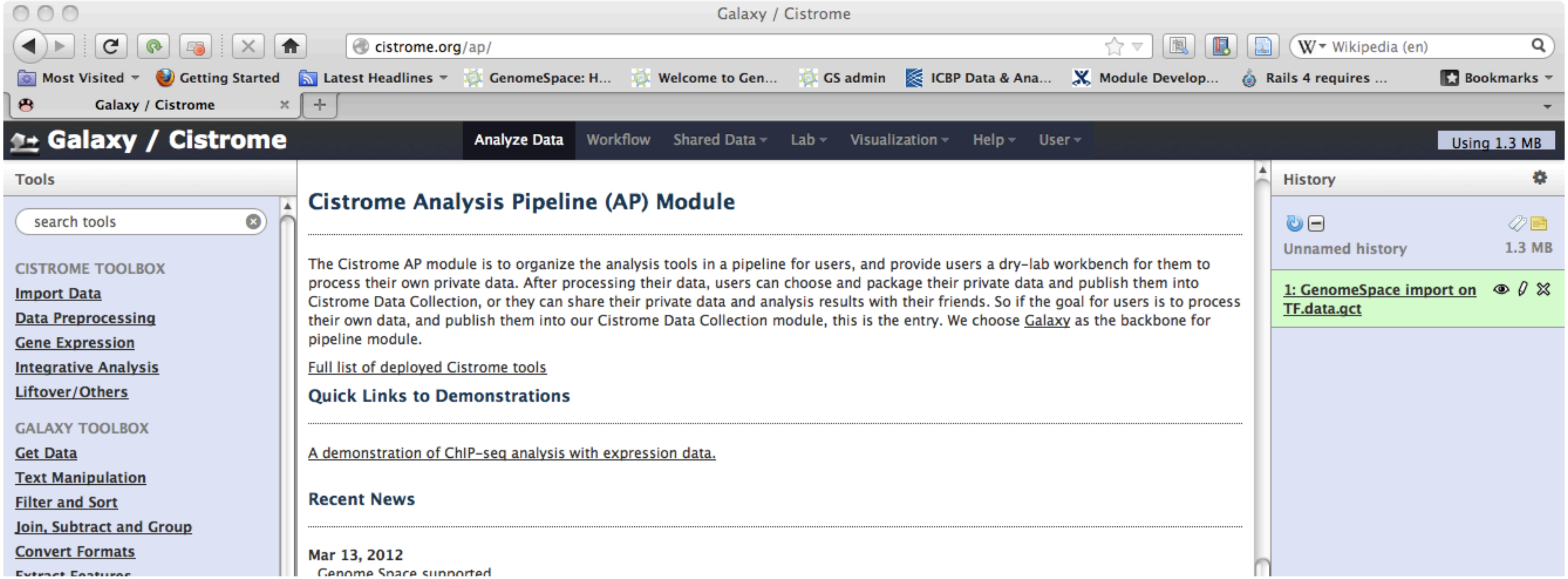
Updated today at 06:00

- 36287 experiments
- 1050421 assays
- 13.52 TB of archived data

Cistrome

29 ChIP-chip and ChIP-seq tools, including:

- Preliminary peak calling
- Correlation analyses
- Downstream genome feature association
- Gene expression analyses
- Motif discovery



Cistrome Analysis Pipeline (AP) Module

The Cistrome AP module is to organize the analysis tools in a pipeline for users, and provide users a dry-lab workbench for them to process their own private data. After processing their data, users can choose and package their private data and publish them into Cistrome Data Collection, or they can share their private data and analysis results with their friends. So if the goal for users is to process their own data, and publish them into our Cistrome Data Collection module, this is the entry. We choose Galaxy as the backbone for pipeline module.

Full list of deployed Cistrome tools

Quick Links to Demonstrations

A demonstration of ChIP-seq analysis with expression data.

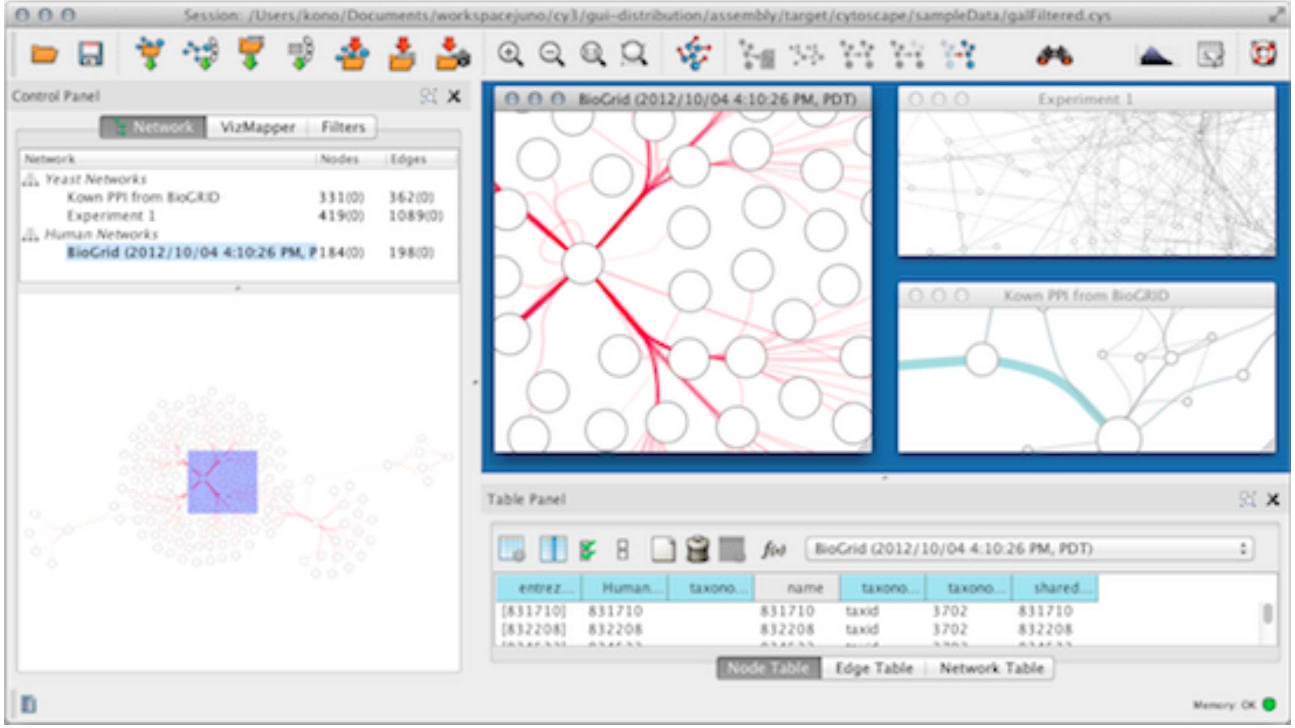
Recent News

Mar 13, 2012

Genome Space supported

Cytoscape

- Visualize molecular interaction networks and biological pathways
- Integrate networks with annotations, gene expression profiles, and other data



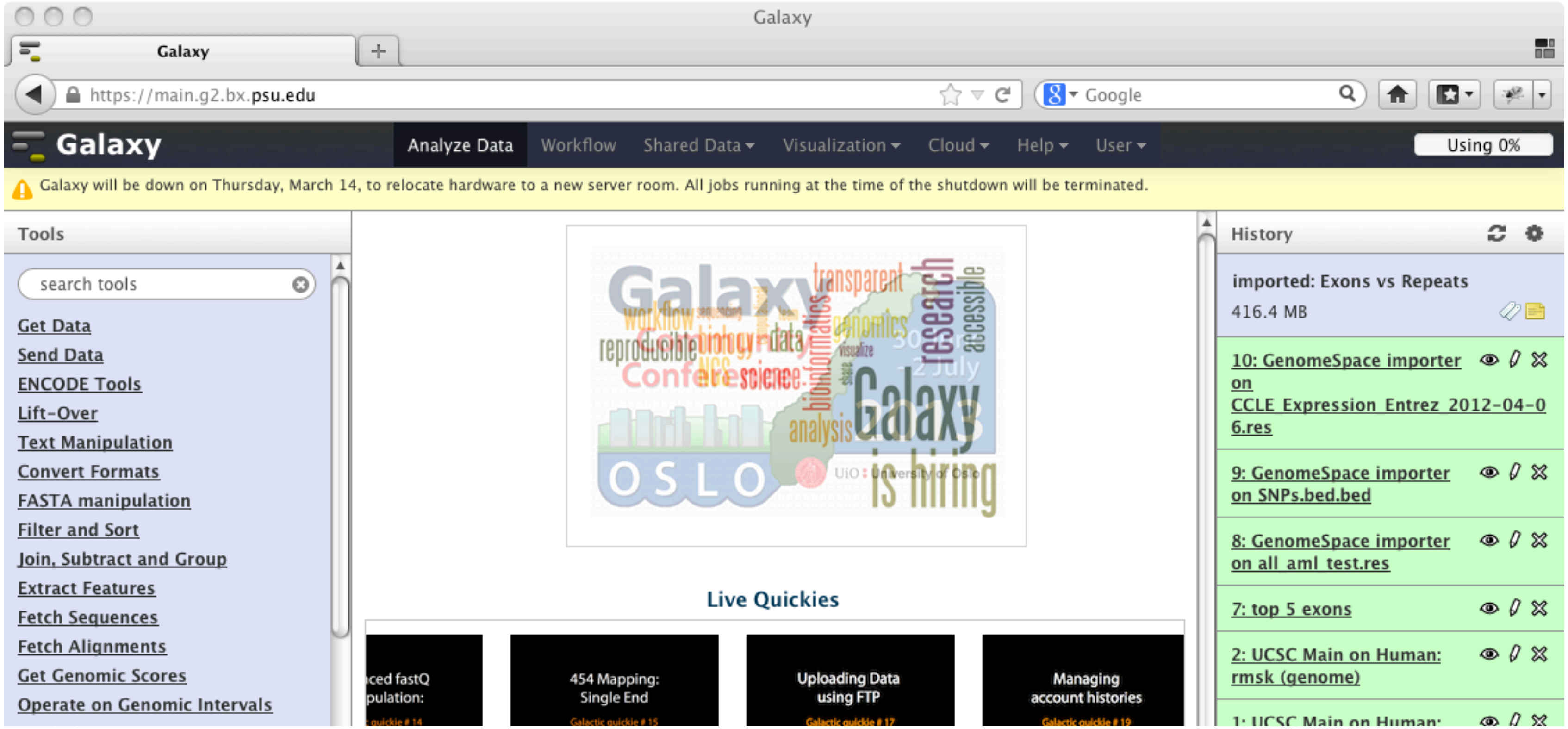
Network	Nodes	Edges
Yeast Networks		
Kown PPI from BioGRID	331(0)	362(0)
Experiment 1	419(0)	1089(0)
Human Networks		
BioGrid (2012/10/04 4:10:26 PM, P	184(0)	198(0)

BioGrid (2012/10/04 4:10:26 PM, PDT)

entrez...	Human...	taxono...	name	taxono...	taxono...	shared...
[831710]	831710		831710	taxid	3702	831710
[832208]	832208		832208	taxid	3702	832208

Galaxy

Galaxy is an open-source, scalable framework for tool integration that allows users to analyze multiple alignments, compare genomic annotations, and profile metagenomic samples, among many possible analyses; workflows allow the linking together of analyses.

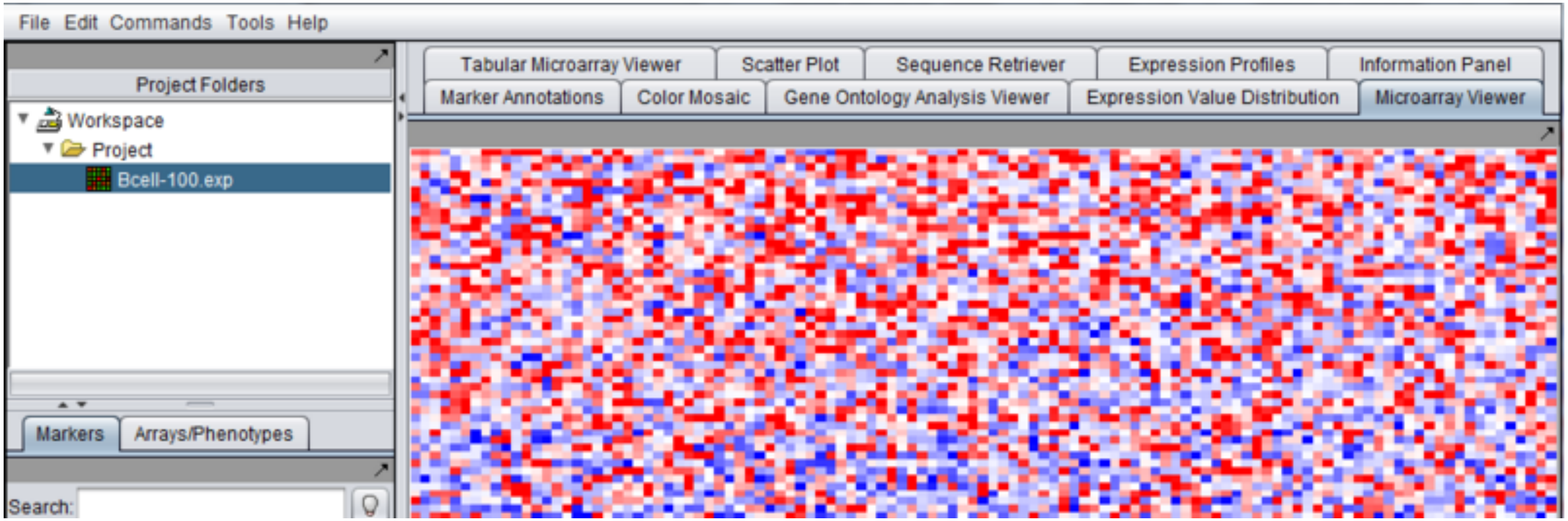


Galaxy will be down on Thursday, March 14, to relocate hardware to a new server room. All jobs running at the time of the shutdown will be terminated.

geWorkbench

Analysis, visualization, and annotation of biomedical data, including:

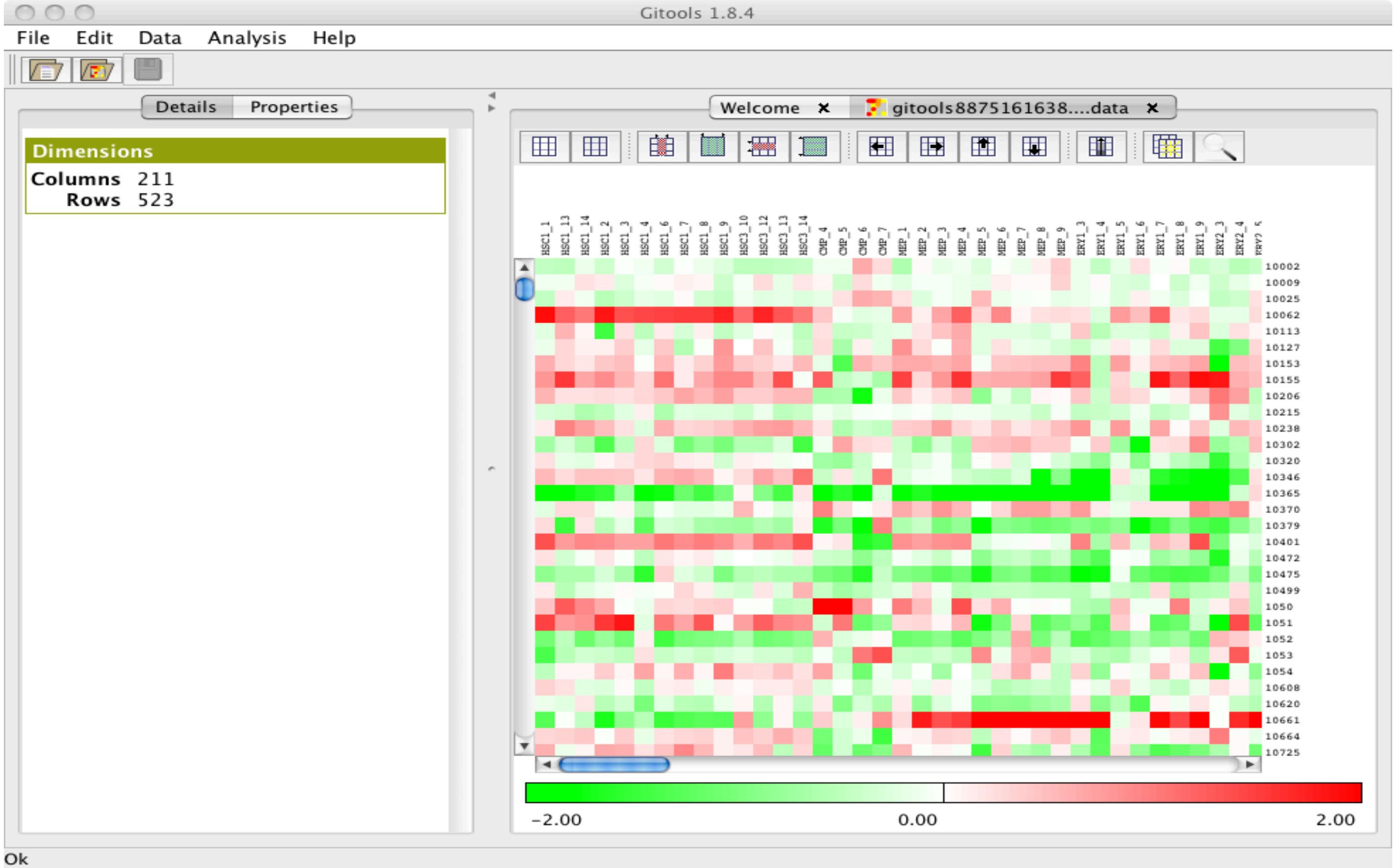
- Microarray filtering, normalization, clustering, network reverse engineering
- Basic and advanced statistical methods
- Regulator analysis
- Common visualization tools
- Links to databases



Gitools

Analysis and visualization of genomic data, including:

- Interactive heatmaps
- Enrichment analysis (e.g. of Gene Ontology terms)
- Import from Web-based data sources (IntOgen, BioMart)



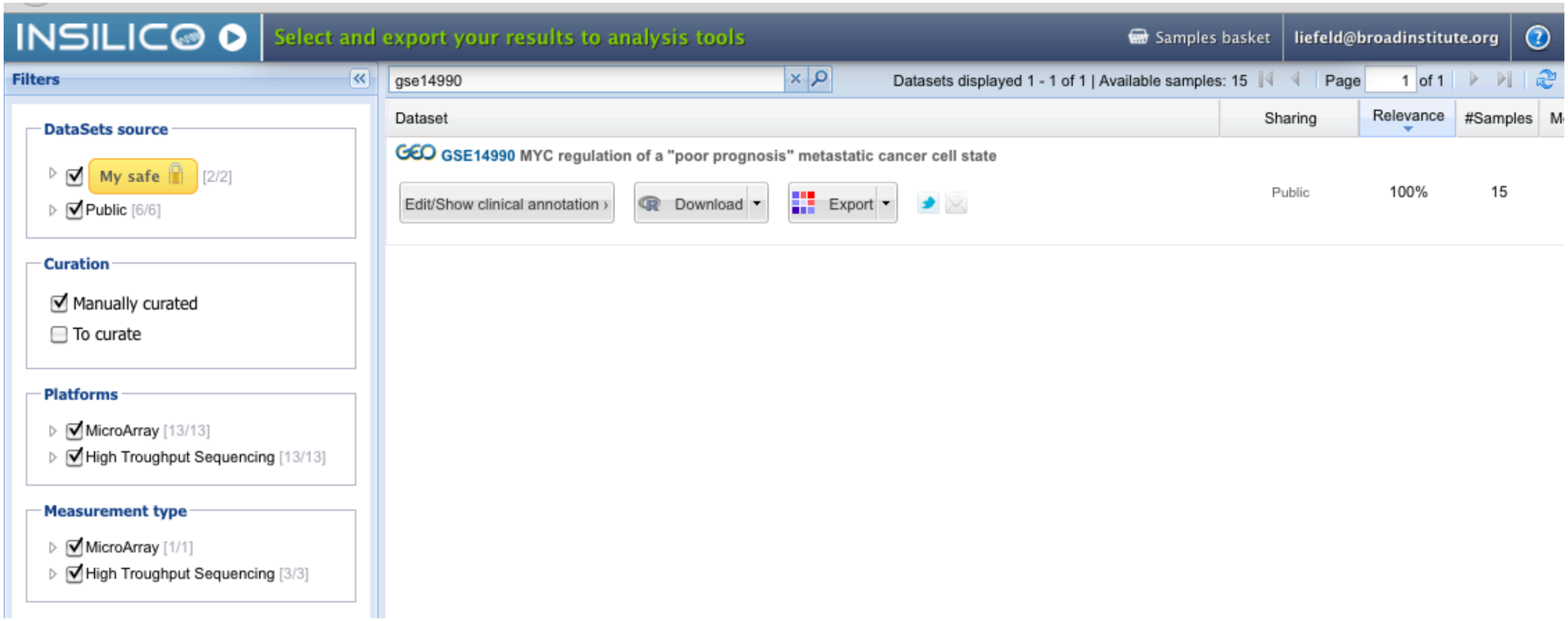
Dimensions

Columns 211

Rows 523

InSilicoDB

Web-based genomics data portal containing thousands of curated public datasets, including all of the Gene Expression Omnibus (GEO).



Dataset	Sharing	Relevance	#Samples
GEO GSE14990 MYC regulation of a "poor prognosis" metastatic cancer cell state	Public	100%	15

UCSC Table Browser

- Query and retrieve genomic sequence data in text format
- Send data to GenomeSpace and other analysis and visualization tools
- Calculate intersections between genome tracks

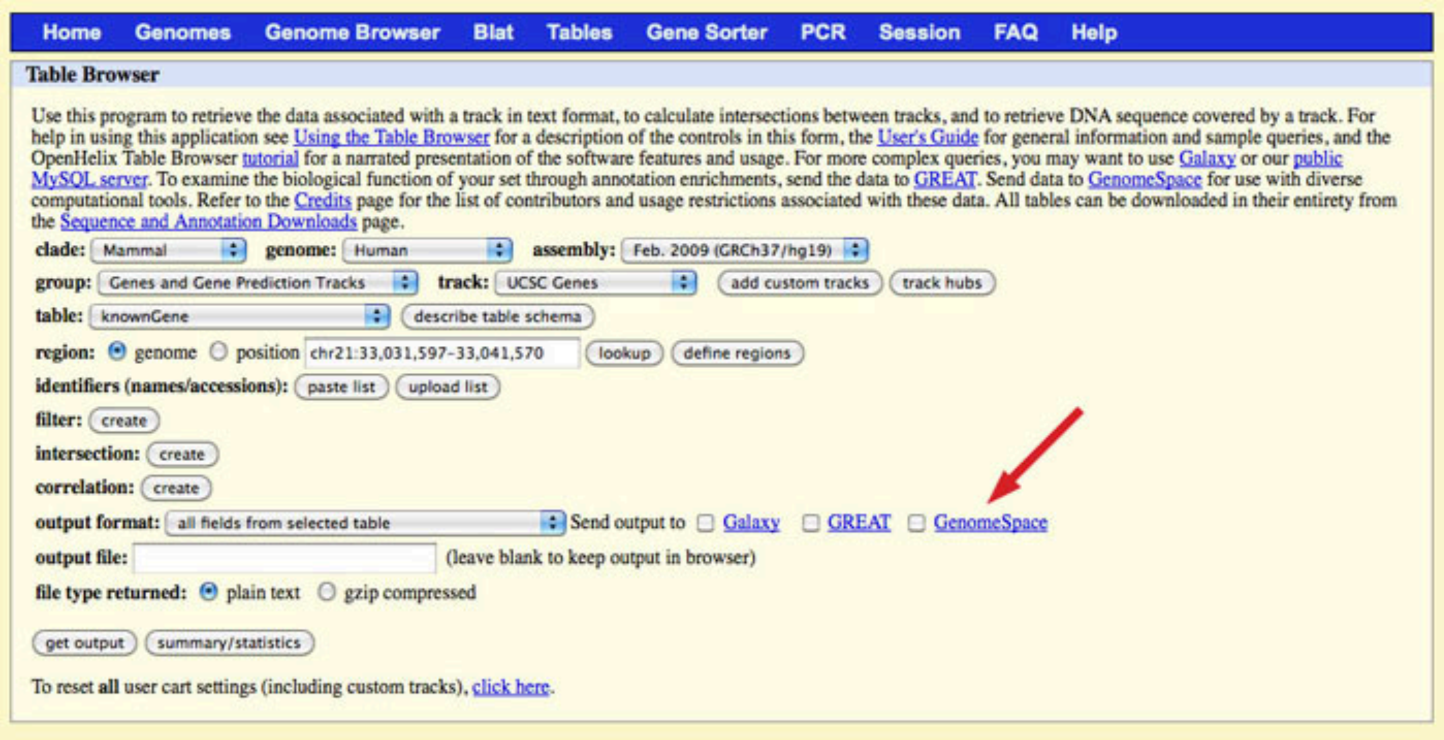


Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see Using the Table Browser for a description of the controls in this form, the User's Guide for general information and sample queries, and the OpenHelix Table Browser tutorial for a narrated presentation of the software features and usage. For more complex queries, you may want to use Galaxy or our public MySQL server. To examine the biological function of your set through annotation enrichments, send the data to GREAT. Send data to GenomeSpace for use with diverse computational tools. Refer to the Credits page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the Sequence and Annotation Downloads page.

clade: Mammal **genome:** Human **assembly:** Feb. 2009 (GRCh37/hg19)

group: Genes and Gene Prediction Tracks **track:** UCSC Genes

table: knownGene

region: ● genome ○ position chr21:33,031,597-33,041,570

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: all fields from selected table Send output to ☐ Galaxy ☐ GREAT ☐ GenomeSpace

output file: (leave blank to keep output in browser)

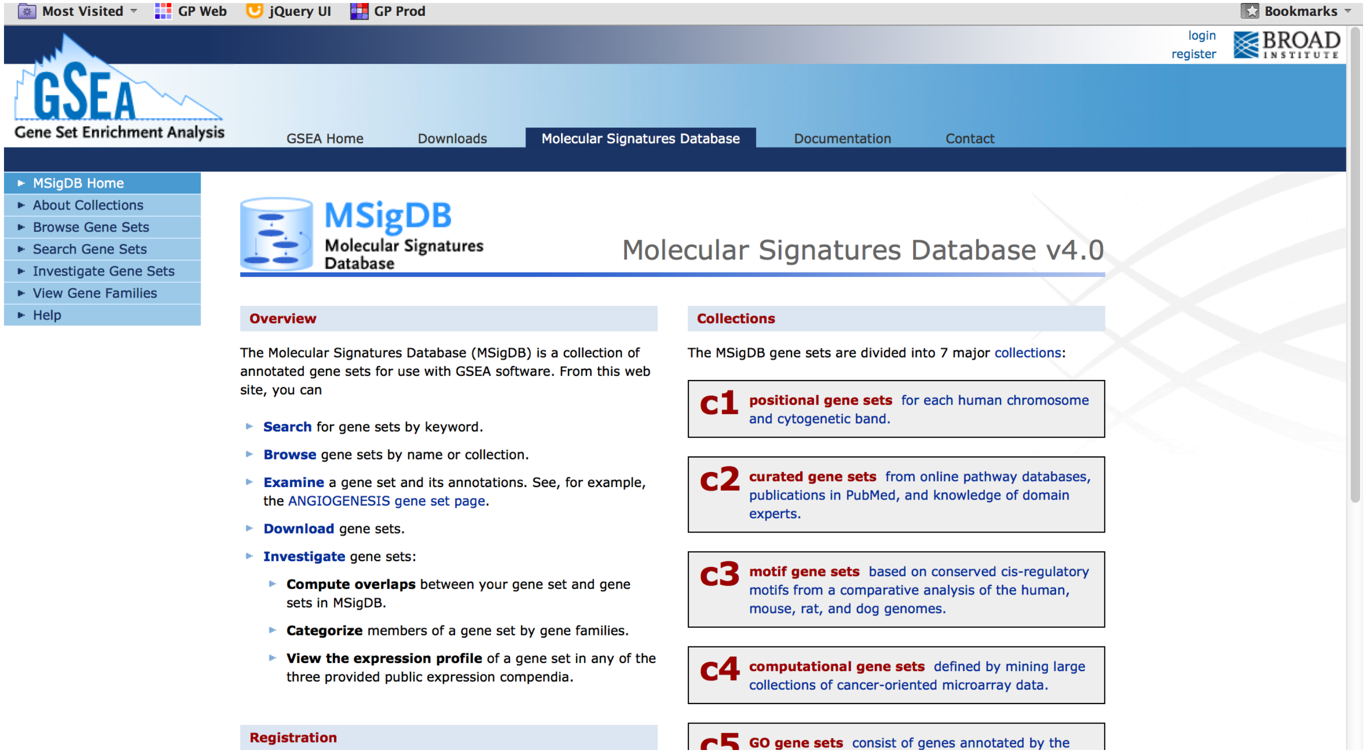
file type returned: ● plain text ○ gzip compressed

To reset **all** user cart settings (including custom tracks), click here.

MSigDB

Molecular Signatures Database

- Query and retrieve a large compendium of gene sets, including regulatory, metabolic, and genomic pathways, genomic position-based gene sets, etc.
- Send data to GenomeSpace and other analysis and visualization tools
- Calculate overlap statistics between gene sets



Molecular Signatures Database v4.0

Overview

The Molecular Signatures Database (MSigDB) is a collection of annotated gene sets for use with GSEA software. From this web site, you can

- **Search** for gene sets by keyword.
- **Browse** gene sets by name or collection.
- **Examine** a gene set and its annotations. See, for example, the ANGIOGENESIS gene set page.
- **Download** gene sets.
- **Investigate** gene sets:
  - **Compute overlaps** between your gene set and gene sets in MSigDB.
  - **Categorize** members of a gene set by gene families.
  - **View the expression profile** of a gene set in any of the three provided public expression compendia.

Registration

Collections

The MSigDB gene sets are divided into 7 major collections:

c1 **positional gene sets** for each human chromosome and cytogenetic band.

c2 **curated gene sets** from online pathway databases, publications in PubMed, and knowledge of domain experts.

c3 **motif gene sets** based on conserved cis-regulatory motifs from a comparative analysis of the human, mouse, rat, and dog genomes.

c4 **computational gene sets** defined by mining large collections of cancer-oriented microarray data.

c5 **GO gene sets** consist of genes annotated by the