

GENOME SPACE



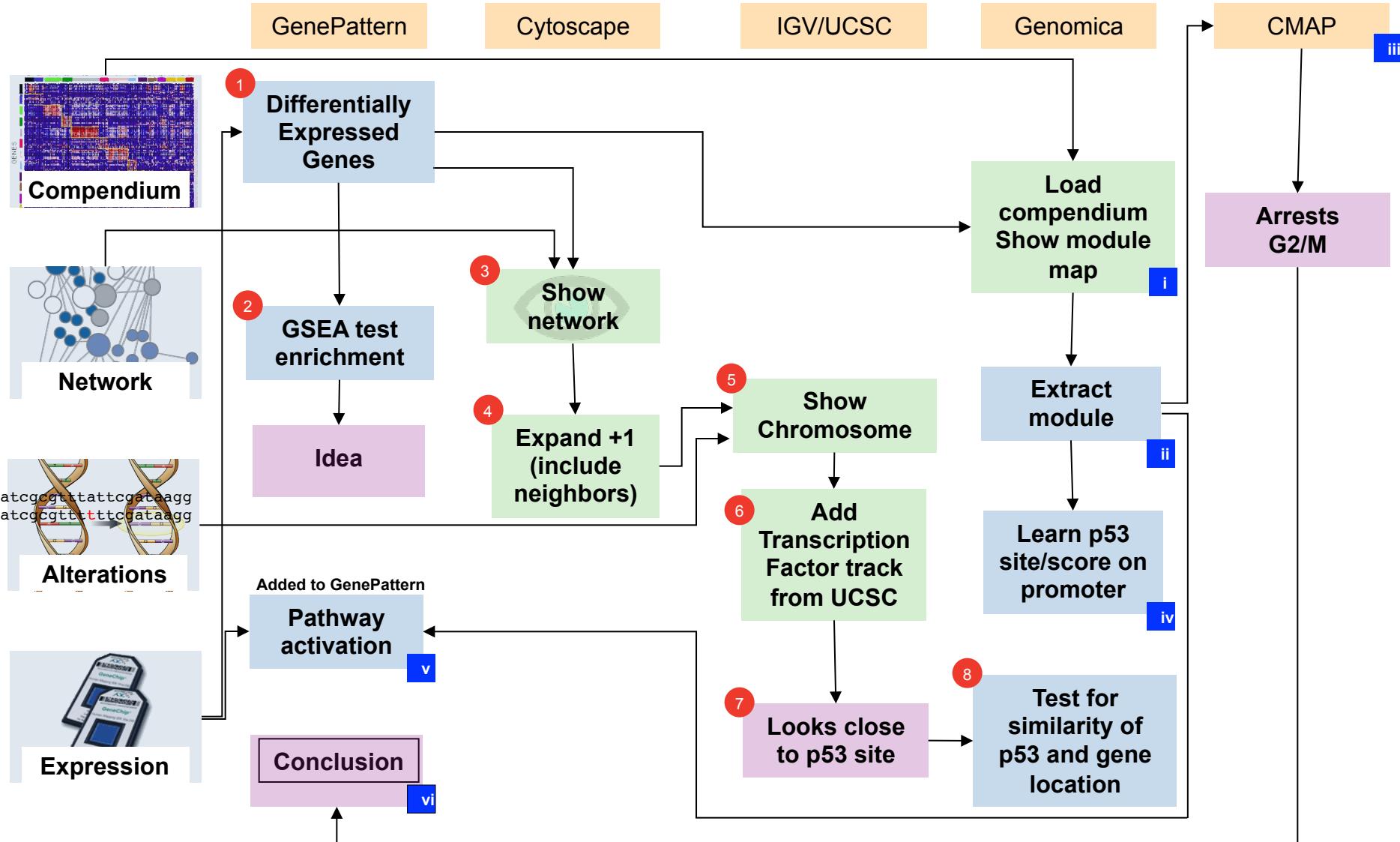
GENOME SPACE



Outline

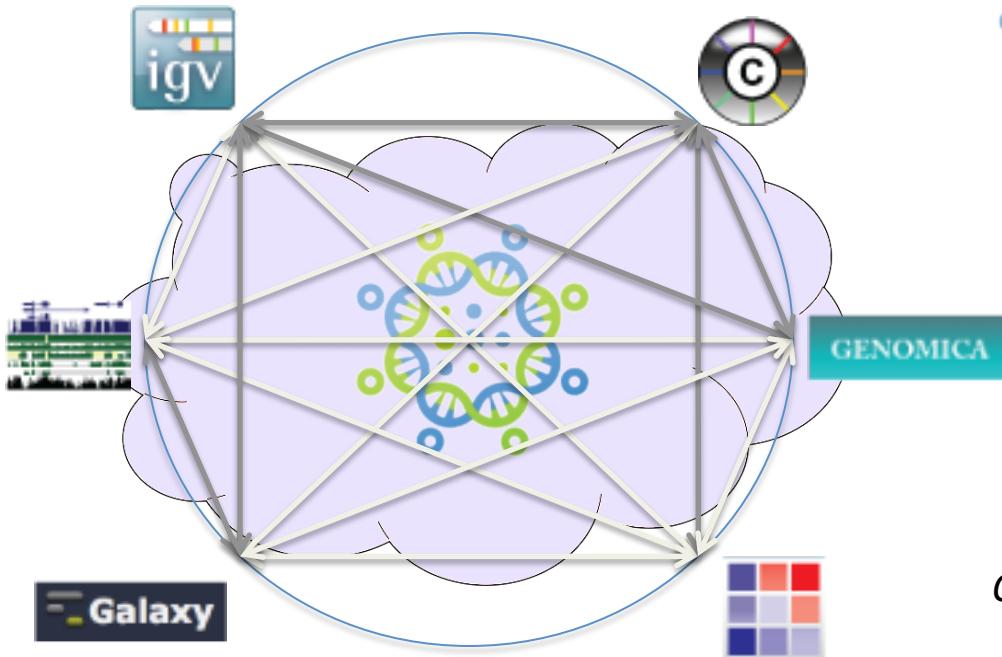
- Introduction to GenomeSpace
- GenomeSpace Tools and Recipes
- GenomeSpace User Interface
- Integrative analysis exercise
- Other GenomeSpace Tools
- GenomeSpace development
- Q and A

The vision: Integrative Translational Genomics



Online community to share diverse computational tools

GENOME SPACE



Seed Tools

Cytoscape
Galaxy
GenePattern
Genomica
IGV
UCSC Browser

Outreach to new tools



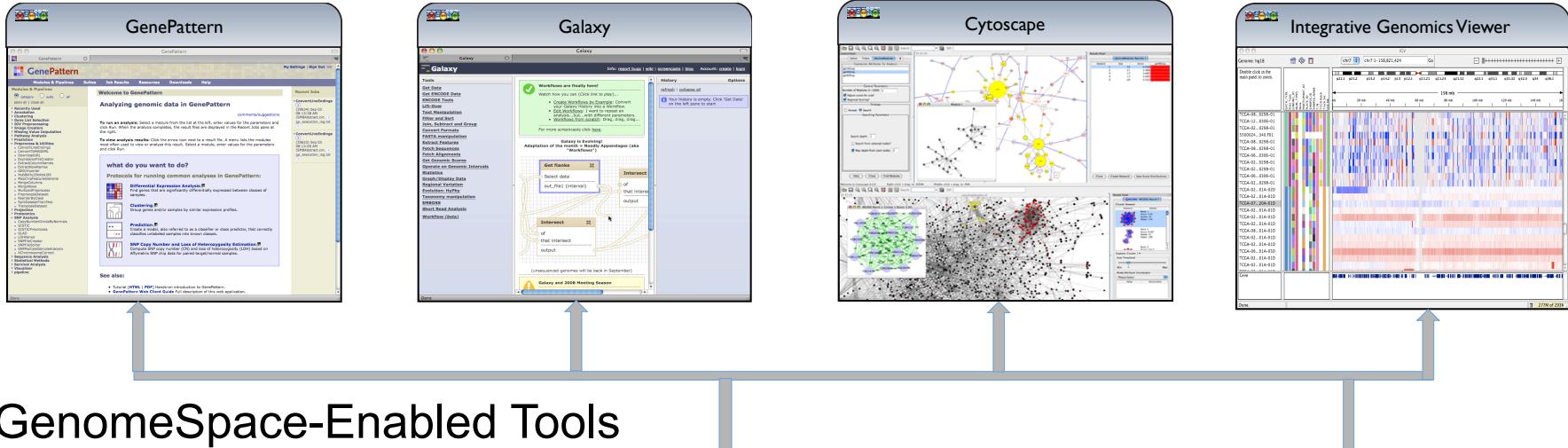
Driving Biological Projects

lincRNAs
Cancer stem cells
Patient Stratification
Outreach to new DBPs

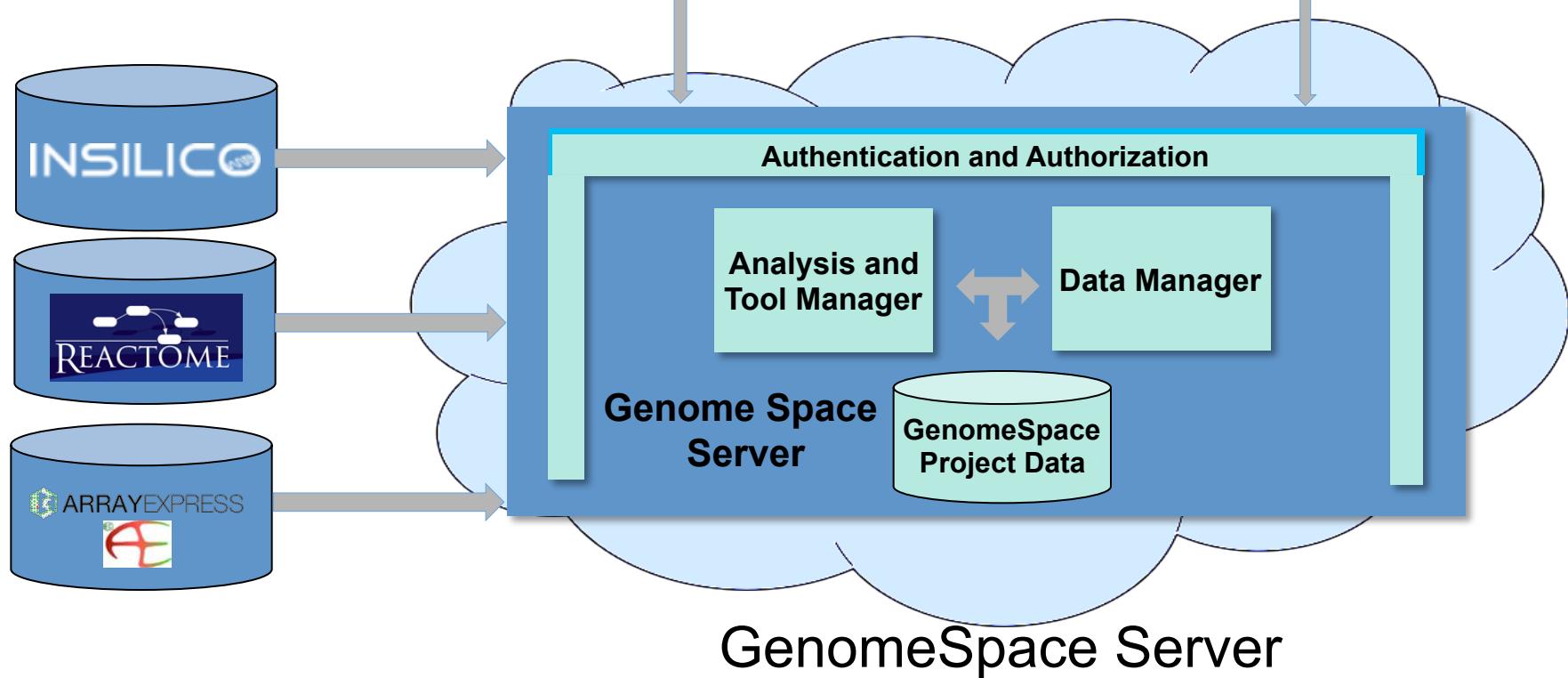
GenomeSpace: a connection layer between integrative analysis tools

- 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



GenomeSpace-Enabled Tools



Register

The screenshot shows the GenomeSpace website homepage. At the top, there is a navigation bar with links like 'Most Visited', 'Getting Started', 'Latest Headlines', 'GenomeSpace: Home', 'Welcome to Gen...', 'GS admin', 'ICBP Data & Ana...', 'Module Develop...', 'Rails 4 requires ...', and 'Bookmarks'. Below the navigation bar is the 'User Login' section, which includes a logo of a cloud with nodes, the text 'GENOMESPACE', and a search bar. The main content area features a large banner with the text 'GENOME SPACE BETA' and 'Frictionless connection of bioinformatics tools'. A red circle highlights the 'Register' button, which is located on the left side of the banner. To the right of the 'Register' button is a green 'User Login' button. To the right of the banner is a collage of bioinformatics-related images, including a DNA sequence viewer, a chromatogram, and a complex network graph. Below the banner, there is a 'STATUS' indicator showing '03.18.13 07:01AM' and 'All systems are operating normally'. To the right of the status are social media icons for Twitter, Facebook, RSS, and email. On the left, there is a 'WHAT'S NEW' section with tabs for 'News Highlights' and 'GenomeSpace Team Blog', and a link to the 'GenomeSpace at BOSC-Broad Interoperability Hackathon'. On the right, there is a sidebar with a tweet from '@genomespace' about the VIZBI workshop and a mention of the Bioinformatics Interoperability Hackathon.

www.genomespace.org

Register

The screenshot shows a web browser window for the GenomeSpace Register page at www.genomespace.org/register. The page has a light blue header with the GenomeSpace logo (a stylized green and blue cloud-like icon) and the word "GENOME SPACE". On the right side of the header is a "User Login" link. Below the header is a navigation bar with links for "What is GenomeSpace?", "Tools", "Documentation", "Developers", "Support", and "About". A search bar is also present. The main content area is titled "Register for a GenomeSpace account". It contains instructions: "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." Below these instructions is a registration form enclosed in a red oval. The form fields are: "Username" (text input), "Email" (text input), "Password" (text input), and "Password confirm" (text input). At the bottom of the form is a "Register" button.

Register

GenomeSpace User Registration

https://identity.genomespace.org/identityServer/registration.html?openid.return_to='https://gsui.genomespace.org/j'

Most Visited Getting Started Latest Headlines GenomeSpace: H... Welcome to Gen... GS admin ICBP Data & Ana... Module Develop... Rails 4 requires ... Bookmarks

GenomeSpace User Registration

 GENOME SPACE

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.

OK

Register

A screenshot of a Gmail inbox interface. The main window displays an email from "genomespacenotifier@gmail.com" to "michaelr" with the subject "GenomeSpace user registration". The email body contains a message and a redacted URL link. A red circle highlights the URL link. To the right of the email, there is a "People (2)" panel for "genomespacenotifier" with options like "Add to circles" and "Show details". The left sidebar shows various Google services and a list of saved messages.

genomespacenotifier@gmail.com
to michaelr 4:14 PM (3 minutes ago)

Hello. To complete the registration of the GenomeSpace account for your G+ profile, 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.

Click here to [Reply](#), [Reply to all](#), or [Forward](#)

20% full
Using 5.1 GB of your 25 GB

©2013 Google - [Terms of Service](#) - [Privacy Policy](#) - [Program Policies](#)
Powered by Google

Last account activity: 1 minute ago
Open in 1 other location [Details](#)

Search people...
Catherine Stein

Login

GenomeSpace Message

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

Most Visited Getting Started Latest Headlines GenomeSpace: H... Welcome to Gen... GS admin ICBP Data & Ana... Module Develop... Rails 4 requires ... Wikipedia (en) Bookmarks

14 18 Q M GenomeSpace Message

GENOME SPACE

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' ★

Most Visited Getting Started Latest Headlines GenomeSpace: H... Welcome to Gen... GS admin ICBP Data & Ana... Module Develop... Rails 4 requires ... Bookmarks

GenomeSpace OpenId Login

OpenID Login

USERNAME:

PASSWORD:

Sign In **Cancel**

[Register new GenomeSpace user](#)
[Forgot your password?](#)

GenomeSpace UI

Welcome to GenomeSpace

<https://gsui.genomespace.org/gsui/gsui.html>

Most Visited Getting Started Latest Headlines GenomeSpace: H... Welcome to Gen... GS admin ICBP Data & Ana... Module Develop... Rails 4 requires ... Bookmarks

14 18 Q M G S Welcome to GenomeSpace + mmr

GENOME SPACE BETA mmr

File | Launch | View | Manage | Recipes | Help

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

Directories

	Filename	Owner	Size	Last Modified
<input type="checkbox"/>	mmr	mmr		
<input type="checkbox"/>	Shared to mmr	System		
<input type="checkbox"/>	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



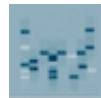
Cistrome



Cytoscape



GenePattern



Genomica



ISACreator



geWorkbench



Gitools



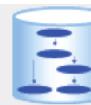
IGV



InSilicoDB



UCSC Table Browser

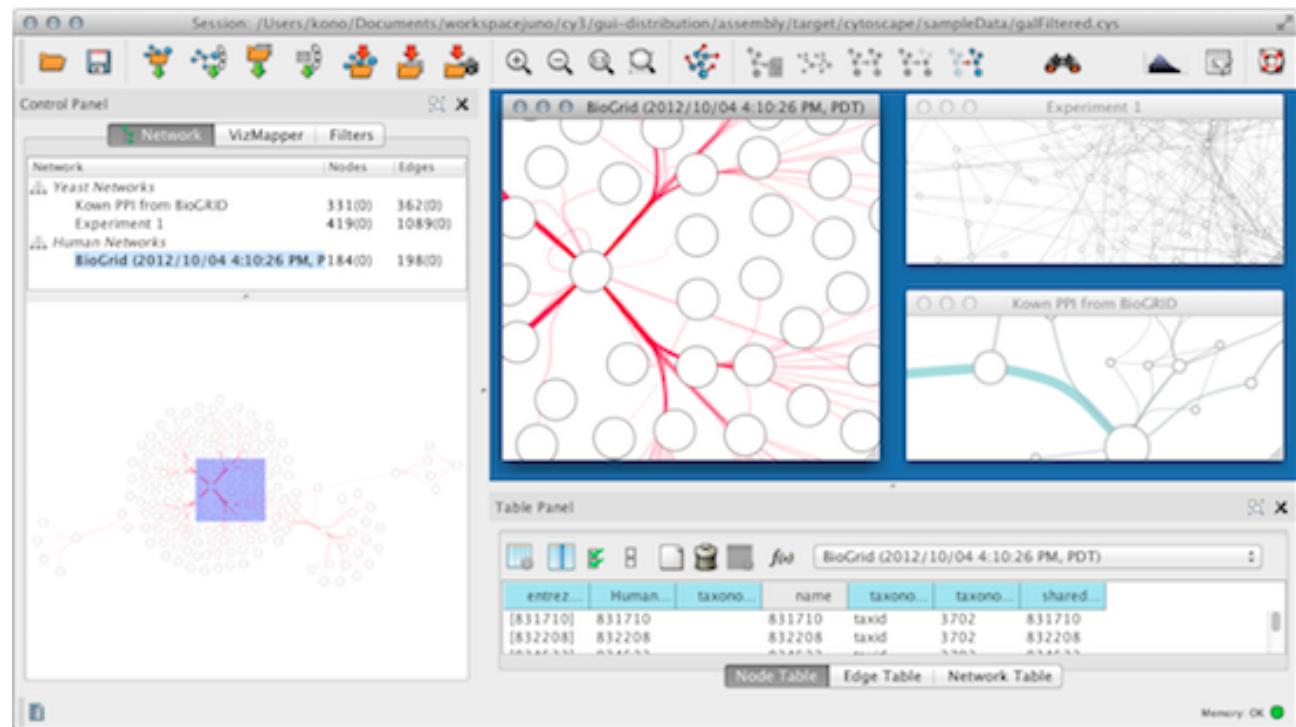


MSigDB



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.





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.

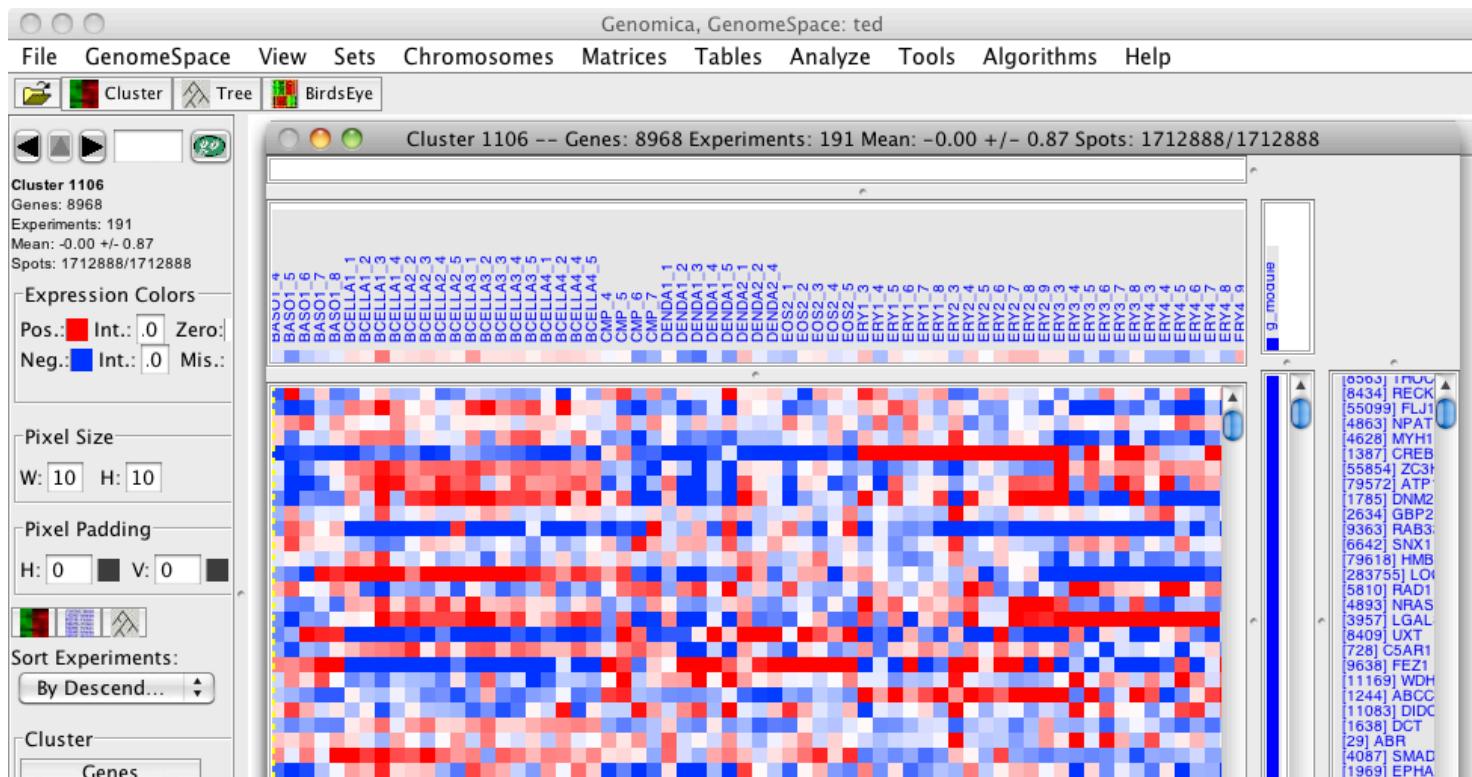
The screenshot shows the Galaxy web interface. At the top, there's a navigation bar with tabs for "Galaxy", "Analyze Data", "Workflow", "Shared Data", "Visualization", "Cloud", "Help", and "User". A message in the center says "Galaxy will be down on Thursday, March 14, to relocate hardware to a new server room. All jobs running at the time of the shutdown will be terminated." On the left, a sidebar titled "Tools" lists various genomic analysis tools. In the center, there's a large banner for the "Galaxy OSLO conference" featuring the text "Galaxy workflow", "transparent", "reproducible", "combinatory", "data", "omics", "research", "science", "informatics", "OSLO", and "Galaxy is hiring". Below the banner, there's a section titled "Live Quickies" with four cards: "Uploading Data using FTP", "Managing account histories", "454 Mapping: Single End", and "Get Genomic Scores". On the right, there's a "History" panel listing several imported datasets and genome space imports, each with a preview thumbnail and download link.

Job ID	Description	Status
imported: Exons vs Repeats	416.4 MB	
10: GenomeSpace importer	on on CCLE Expression Entrez 2012-04-06.res	
9: GenomeSpace importer	on on SNPs.bed.bed	
8: GenomeSpace importer	on 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.





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.

The screenshot shows the GenePattern web application running in a browser window. The title bar reads "GenePattern". The address bar shows the URL "genepattern.broadinstitute.org/gp/pages/index.jsf". The top menu bar includes "Modules & Pipelines", "Suites", "Job Results", "Resources", "Downloads", "Help", and "GenomeSpace". On the right side of the header, there are links for "My Settings" and "Sign out".

The main content area features a "Welcome to GenePattern" message and a section titled "Analyzing genomic data in GenePattern". Below this, a box titled "what do you want to do?" lists two bullet points: "Click a protocol to run an analysis. GenePattern guides you step by step." and "Click Quick Start for instructions on how to run any module in GenePattern.". Another section titled "Protocols for running common analyses in GenePattern:" lists "Run an Analysis in GenePattern" (with a brief description) and "Differential Expression Analysis" (with a brief description).

A sidebar on the right contains sections for "Recent Jobs" and "Uploads". Under "Recent Jobs", there are three entries: "ConvertLineEndings" (Mar 01 01:15), "PreprocessDataset" (Feb 07 12:21), and "ExtractColumnName" (Feb 01 02:46). Each entry includes a link to the job details and download links for "stderr.txt" and "stdout.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.

The screenshot shows the ArrayExpress homepage with a dark header bar. The header includes the EMBL-EBI logo, a search bar with examples like "E-MEXP-31, cancer, p53, Geuvadis", and links for "Services", "Research", "Training", "Industry", and "About us". Below the header, there's a main navigation bar with links for "Home", "Experiments", "Arrays", "Submit", "Help", and "About ArrayExpress". A "Feedback" and "Login" button are also present. The main content area features a large title "ArrayExpress - functional genomics data" and a brief description of the database's purpose. To the right, a sidebar titled "Data Content" displays statistics: "Updated today at 06:00", "36287 experiments", "1050421 assays", and "13.52 TB of archived data".

ArrayExpress < EMBL-EBI

www.ebi.ac.uk/arrayexpress/

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

EMBL-EBI

Services Research Training Industry About us

ArrayExpress

Examples: [E-MEXP-31](#), [cancer](#), [p53](#), [Geuvadis](#)

Search Advanced

Home Experiments Arrays Submit Help About ArrayExpress

Feedback Login

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

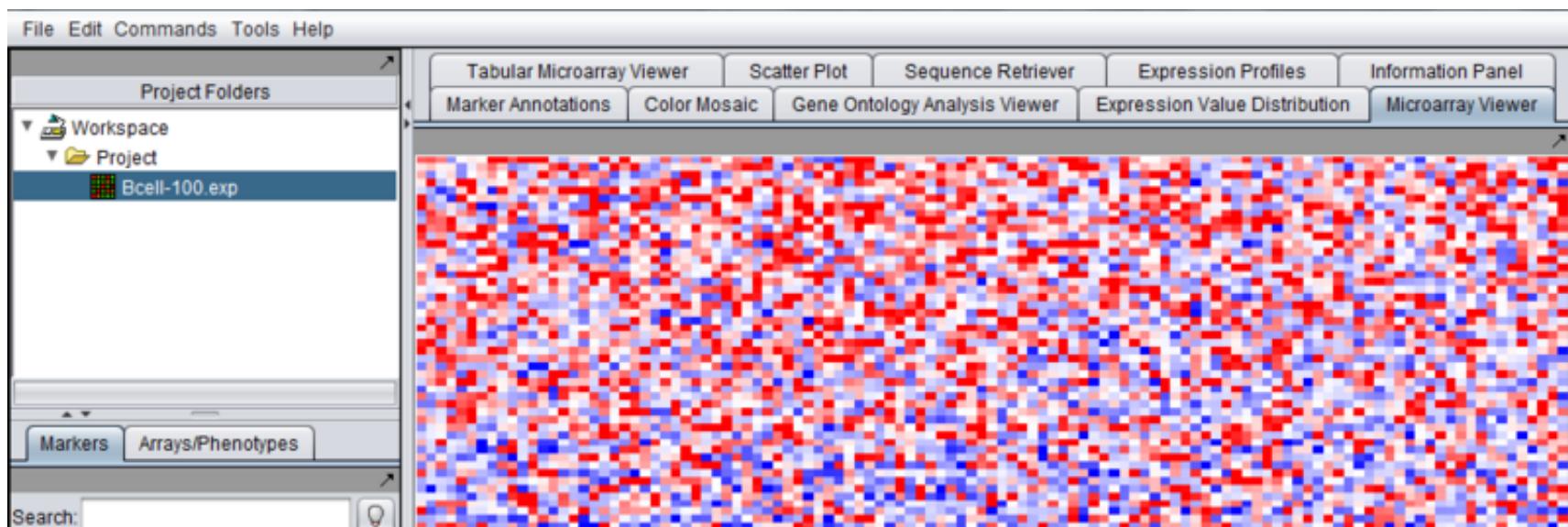
Updated today at 06:00

- 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

cistrome.org/ap/root

Galaxy / Cistrome

Analyze Data Workflow Shared Data Lab Visualization Help User

Using 0 bytes

Tools

search tools

CISTROME TOOLBOX

[Import Data](#)

[Data Preprocessing](#)

[Gene Expression](#)

[Integrative Analysis](#)

[Liftover/Others](#)

GALAXY TOOLBOX

Get Data

History

Unnamed history 0 bytes

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

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!

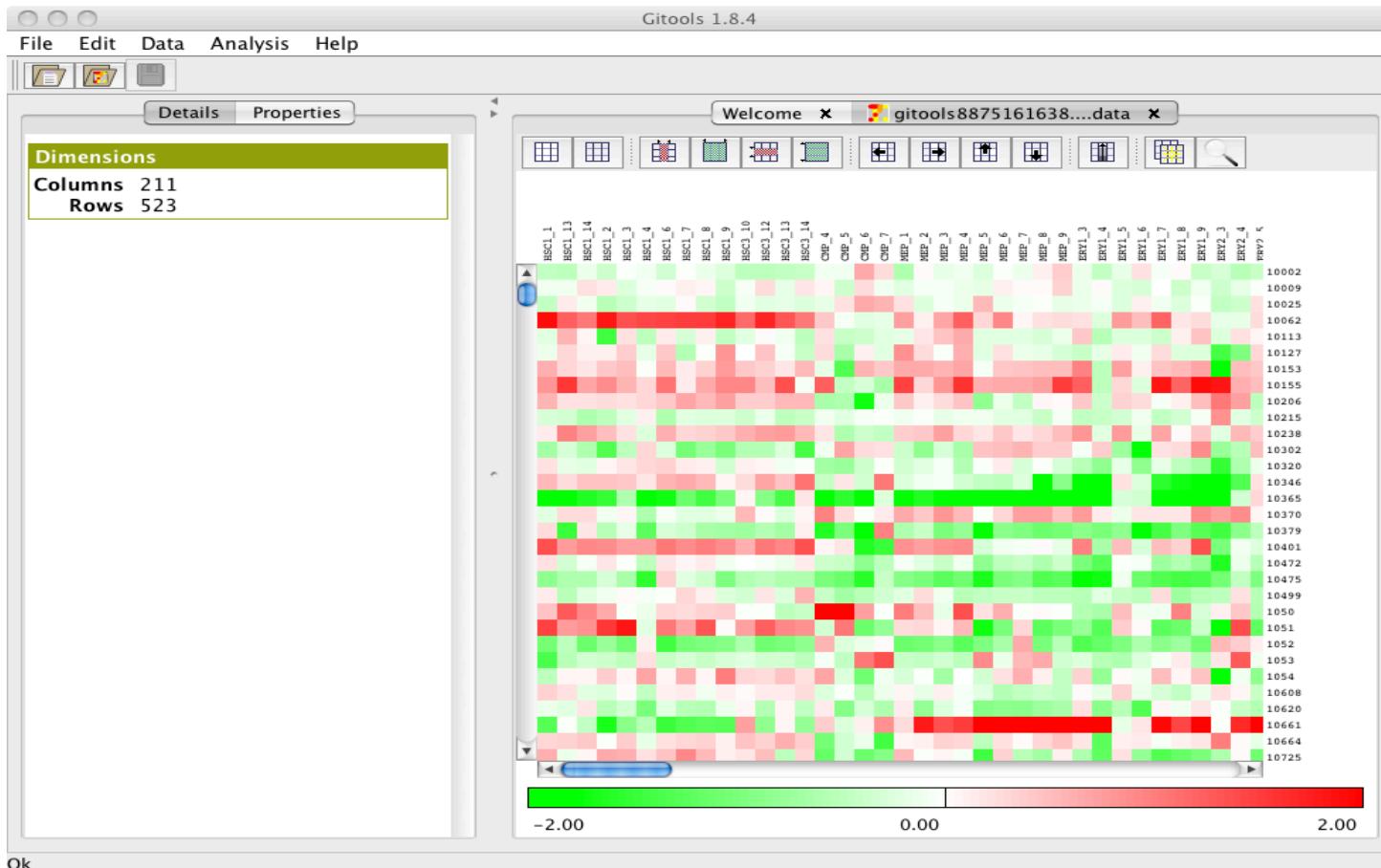
UPDATE: We believe that we found the cause of the problem. Everything should be back and functional. ****Please report the bugs if you are still experiencing problems****

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.



Gitools

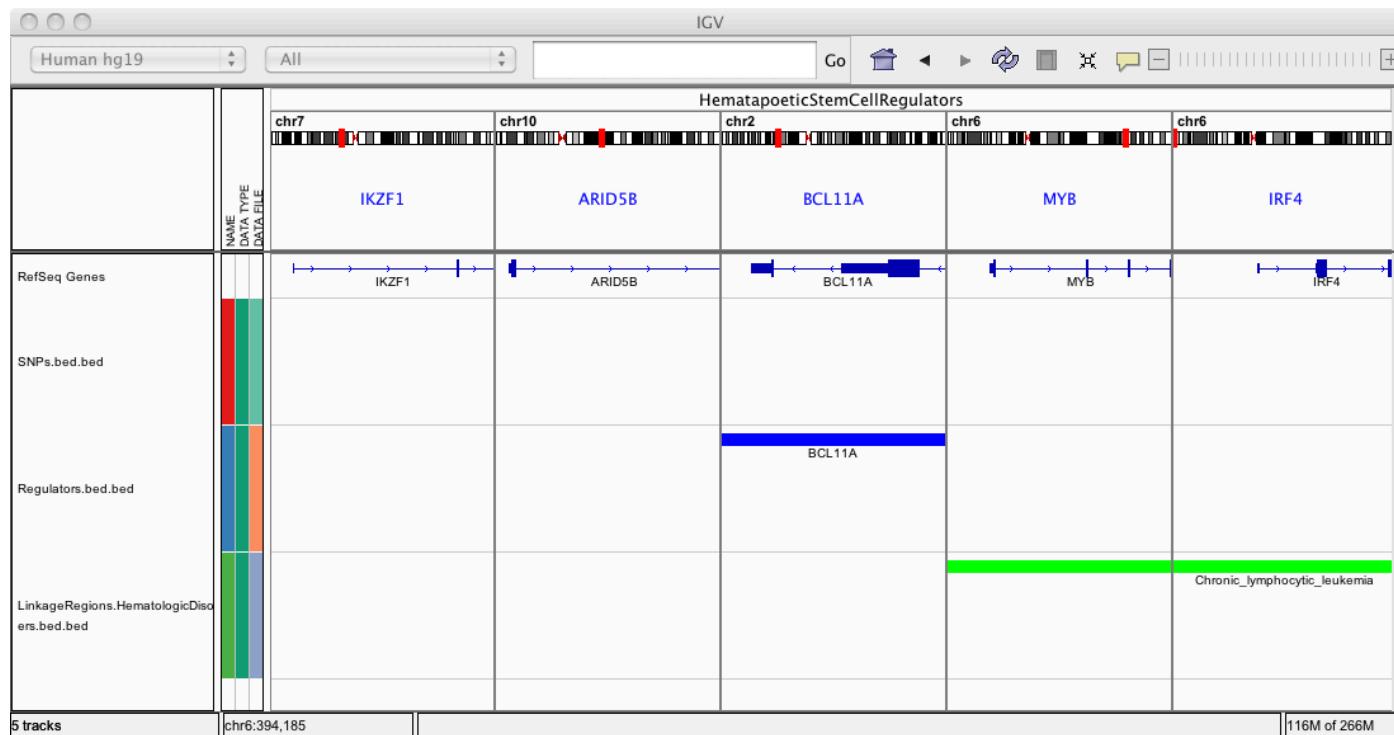
- Gitools is a framework for analysis and visualization of genomic data using interactive heatmaps.





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.

The screenshot shows the InSilicoDB web interface. At the top, there is a navigation bar with the InSilico logo, a search bar containing "gse14990", and user information like "Samples basket" and "liefeld@broadinstitute.org". Below the navigation bar, the main content area displays the dataset "GSE14990 MYC regulation of a 'poor prognosis' metastatic cancer cell state". On the left side, there are several filter panels:

- DataSets source:** Includes checkboxes for "My safe" (selected) and "Public" (6/6).
- Curation:** Includes a checkbox for "Manually curated" (selected) and an unchecked checkbox for "To curate".
- Platforms:** Includes checkboxes for "MicroArray" (selected) [13/13] and "High Throughput Sequencing" [13/13].
- Measurement type:** Includes checkboxes for "MicroArray" [1/1] and "High Throughput Sequencing" [3/3].

At the bottom of the dataset view, there are buttons for "Edit/Show clinical annotation", "Download", "Export" (with options for CSV, Excel, and JSON), and social sharing links for Twitter and Email. The dataset details show it is "Public", "100%", and has "15" samples.



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.

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help

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 add custom tracks track hubs

table: knownGene describe table schema

region: genome position chr21:33,031,597-33,041,570 lookup define regions

identifiers (names/accessions): paste list upload list

filter: create

intersection: create

correlation: create

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

get output summary/statistics

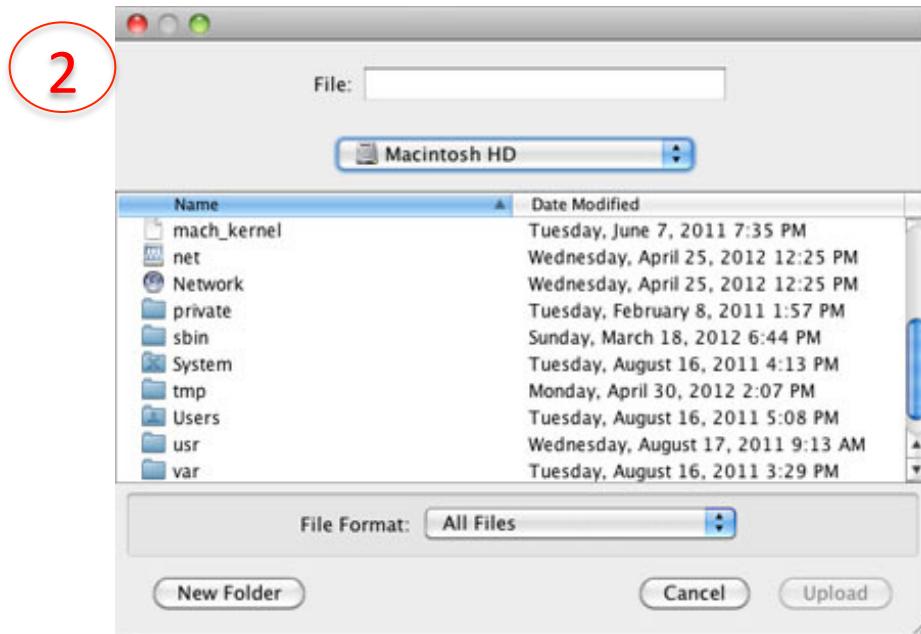
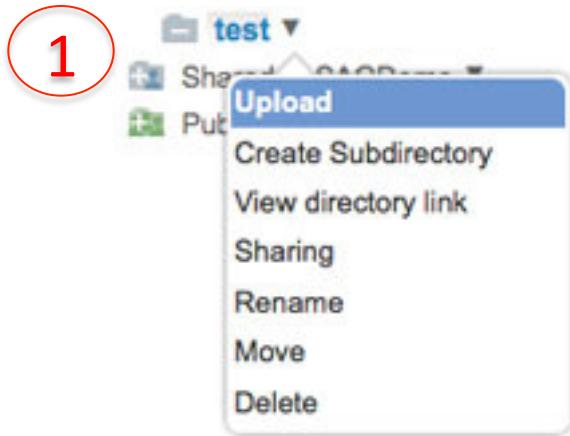
To reset all user cart settings (including custom tracks), [click here](#).



Basic GenomeSpace recipes

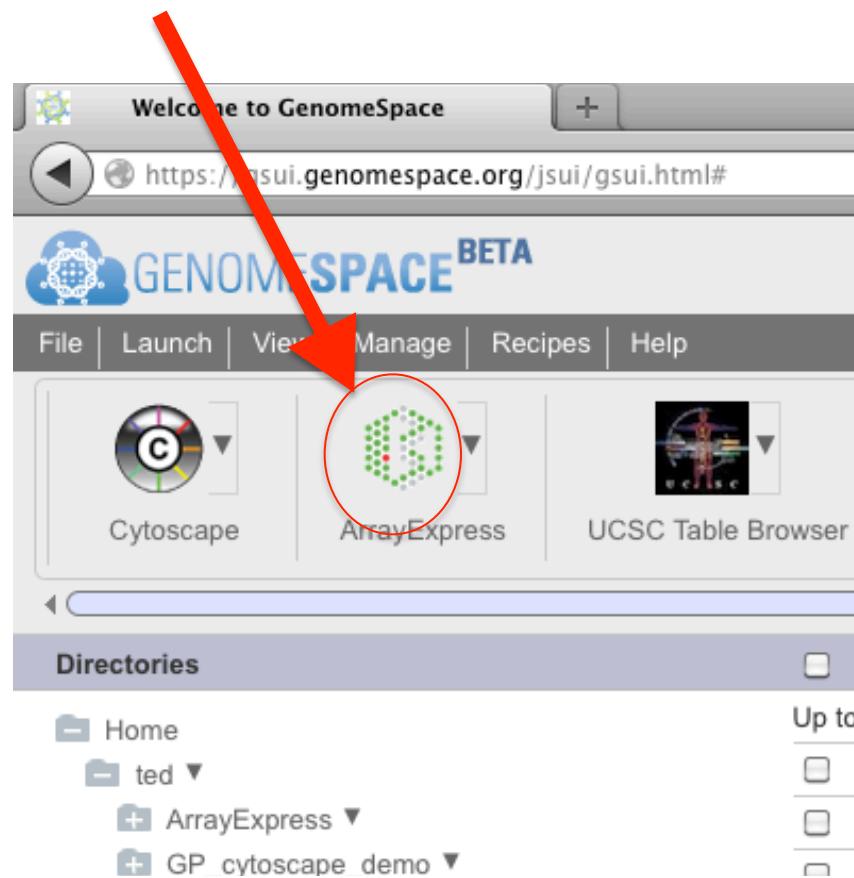
- Uploading data
- Launching tools
- Transitioning across tools

Uploading Data



Launching tools

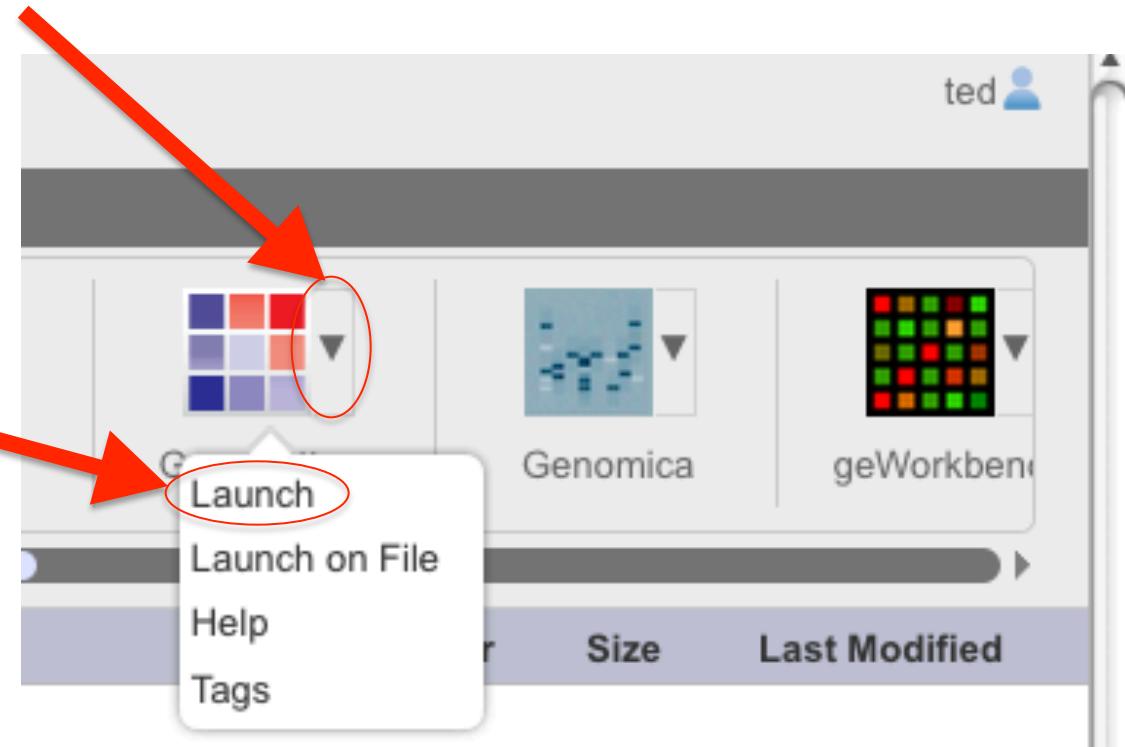
Click on the tool's icon



Launching tools

Open the tool's context menu

Then click on Launch
(or Launch on File)

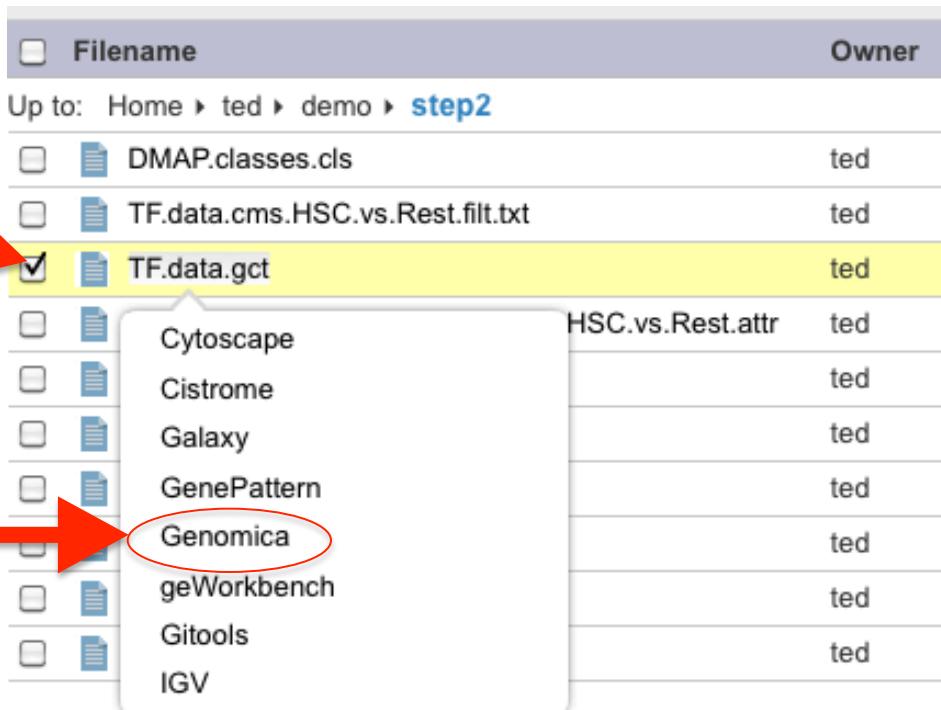


Launching Tools

Click the checkbox for one (or more) files



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



Launching tools

The screenshot shows the GenomeSpace web interface. At the top, there is a toolbar with several tool icons: Browser, Cistrome, Galaxy, GenePattern, Genomica, and geWorkbench. A red arrow points from the left towards the GenePattern icon. Below the toolbar is a file browser window. The current directory is 'Up to: Home > ted > demo > step2'. Inside this directory, there are several files listed:

Filename	Owner	Size	Last Modified
DMAP.classes.cls	ted	1.31 Mb	12 months ago
TF.data.cms.HSC.Js.Rest.filt.tx	ted	1.83 Mb	12 months ago
TF.data.gct	ted	433 bytes	5 months ago
TF.data.genesymbol.comp.marker.attr	ted	2.02 Mb	12 months ago
TF.data.tab	ted	1.31 Mb	12 months ago
all_aml_test.res	ted	1 Kb	12 months ago
all_aml_test.slice.gos.res	ted	134 Kb	12 months ago
all_aml_train.comp.marker.attr	ted	1.31 Mb	12 months ago

At the bottom of the interface, there is a navigation bar with links for 'and Harvard' and 'Funded by the National Human Genome Research Institute (NHGRI)'. There is also a link to 'GSE14990GPL570_RNA_FRMAGENE_16835'.

Click and drag a
File onto a tool
icon

The screenshot shows a 'Launch GenePattern' dialog box. It displays the file 'TF.data.gct' under 'Currently Selected Files'. The dialog has two buttons at the bottom: 'Launch' (highlighted with a red arrow) and 'Close'.

Below the dialog is a list of files in the same directory as the previous screenshot:

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
ted	1.83 Mb	12 months ago
ted	433 bytes	5 months ago
ted	2.02 Mb	12 months ago

At the bottom of the interface, there is a navigation bar with links for '©2012 The Broad Institute of MIT and Harvard' and 'Funded by the National Human Genome Research Institute (NHGRI)'.

Then click the
Launch button

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

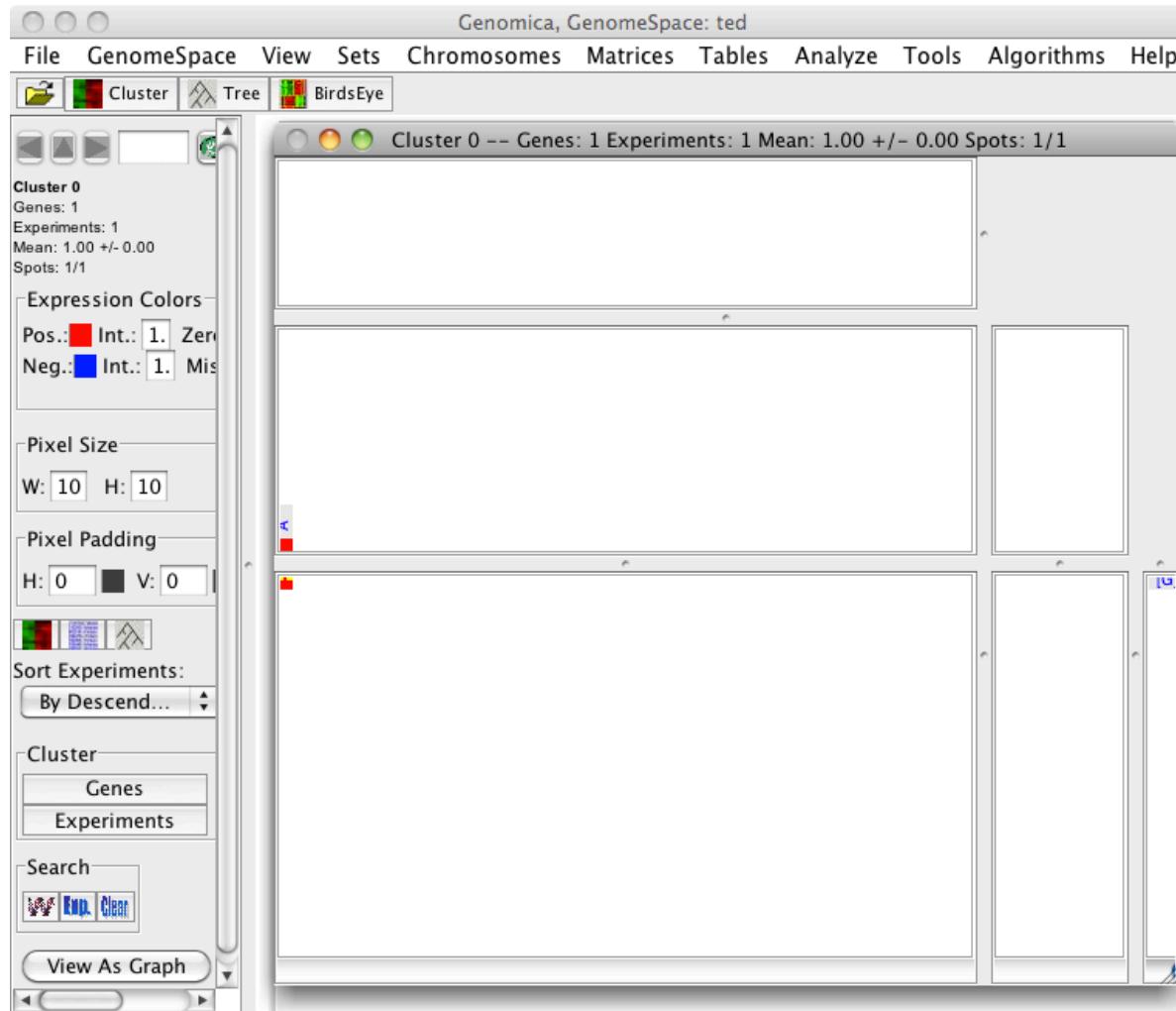
Home ▶ Public ▶ SharedData ▶ Demos ▶ Scenario ▶ step3 ▶ **80_module.gxp**

Or

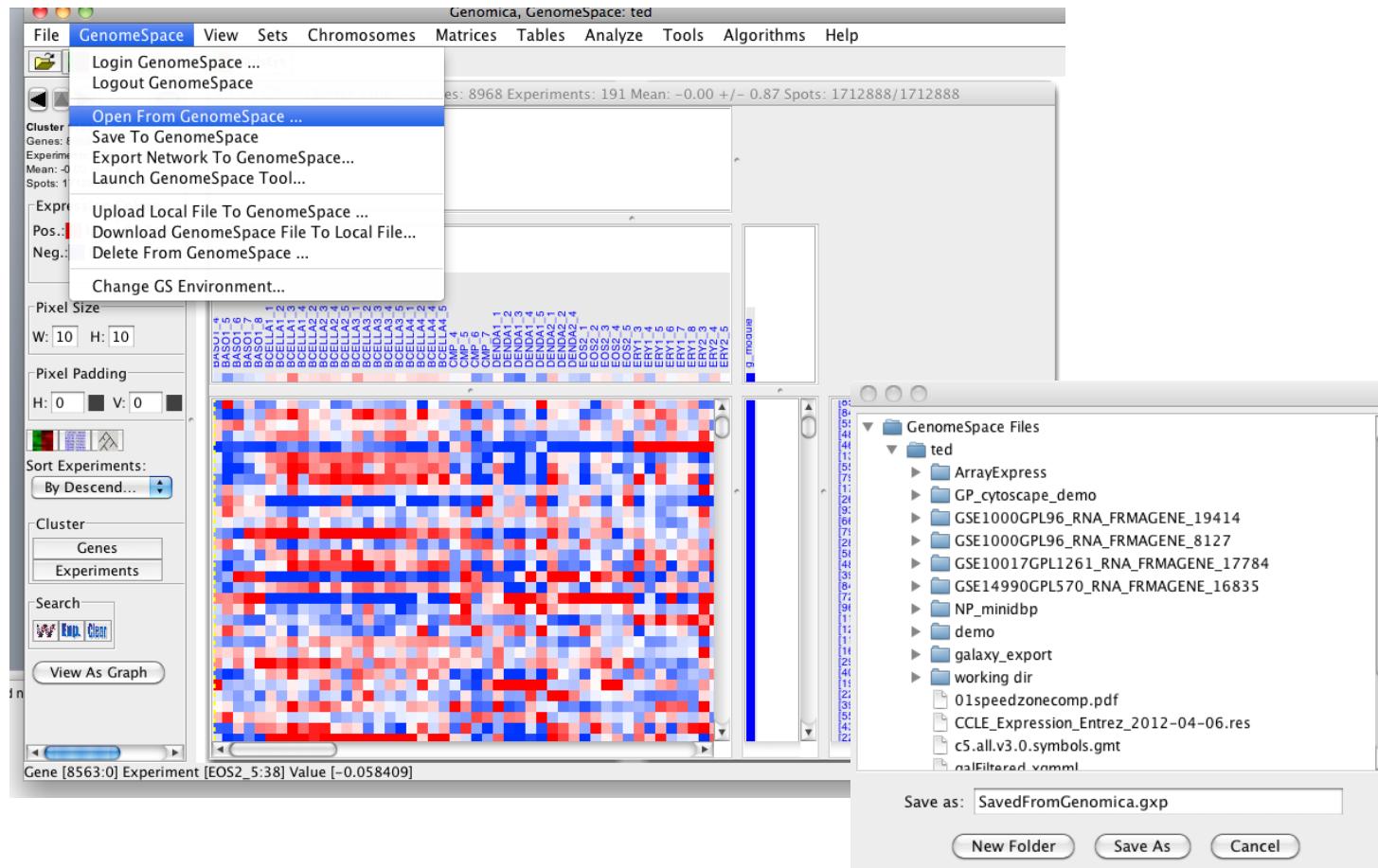
Home ▶ Shared to <your id> ▶ mmr ▶ FGED ▶ **80_module.gxp**

Loading into Genomica

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



Saving Back to GenomeSpace



Launching GenePattern

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

The screenshot shows a web browser window titled "Receive GenomeSpace File". The address bar displays the URL "genepattern.broadinstitute.org/gp/pages/genomespace/receiveFile.jsp?files=https%3A%2F%2Fdm.genomespace.org%2Fc". The main content area features the GenePattern logo and a banner with a colorful grid pattern. Below the banner is a navigation bar with links: Modules & Pipelines, Suites, Job Results, Resources, Downloads, Help, and GenomeSpace. The "GenomeSpace" link is highlighted with a blue background. The main message on the page reads "GenomeSpace has sent you a file" and lists a file named "SaveFromGenomica.gxp Converted to gct". It includes a "Save a Copy:" dropdown set to "Upload Directory" and a "Submit" button. At the bottom, there is a "Send to Module:" dropdown set to "PreprocessDataset" and another "Submit" button.

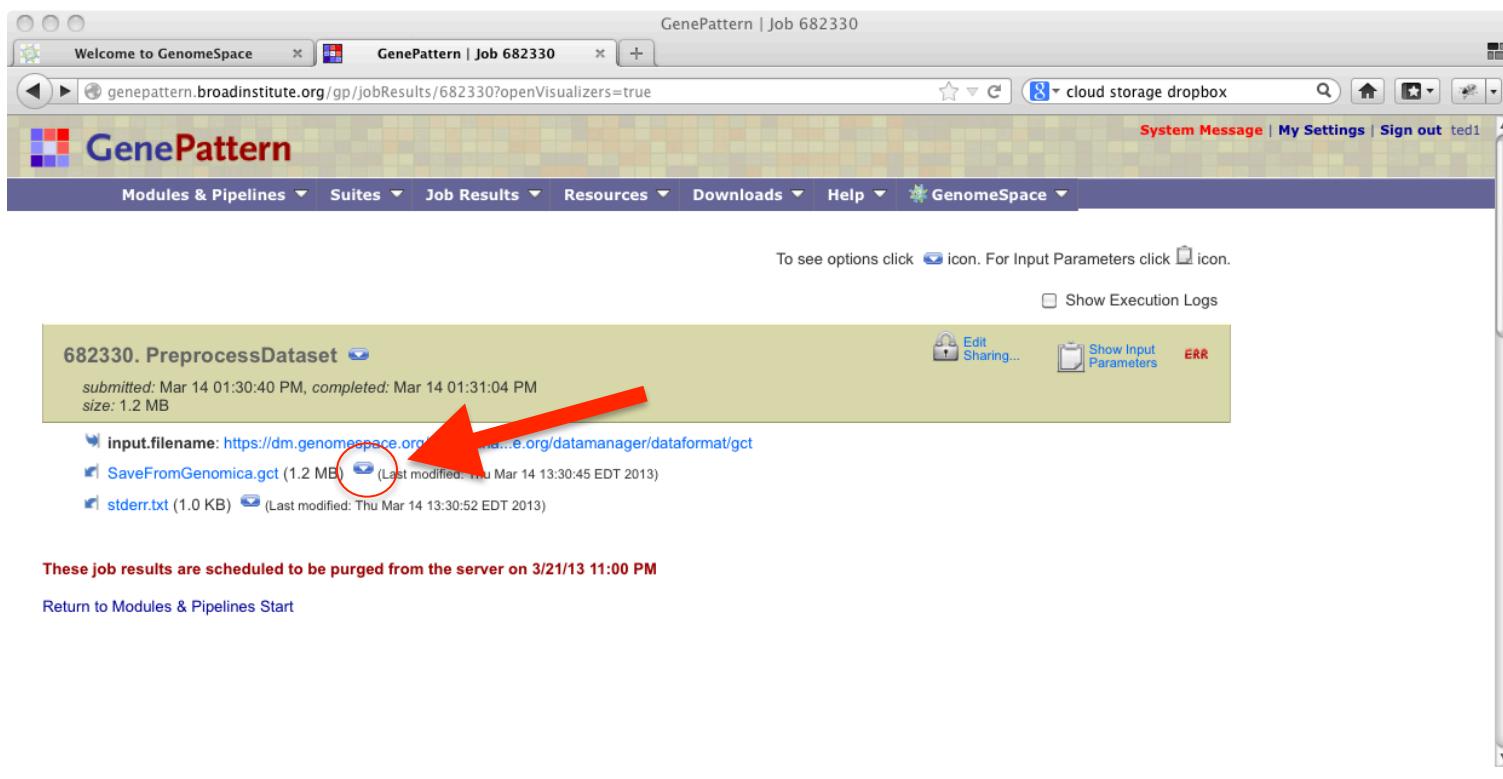
Process the data

- Run PreprocessDataset with default parameters

The screenshot shows the GenePattern software interface for running a PreprocessDataset. The window title is "GenePattern - PreprocessDataset". The main menu bar includes "Welcome to GenomeSpace", "GenePattern - PreprocessDataset", "Modules & Pipelines", "Suites", "Job Results", "Resources", "Downloads", "Help", and "GenomeSpace". The left sidebar lists various modules and pipelines, with "Preprocess & Utilities" expanded. The central panel displays the "PreprocessDataset" configuration form, version 4. It includes fields for "input filename*", "threshold and filter" (set to "yes"), "floor" (20), "ceiling" (20000), "min fold change" (3), and "min delta" (100). A "Run" button is at the top right. To the right, a "Recent Jobs" panel shows completed tasks: "PreprocessDataset" (Mar 14 01:27:43 PM), "ConvertLineEndings" (Mar 01 01:15:18 PM), "PreprocessDataset" (Feb 07 12:21:20 PM), and "ExtractColumnNames" (Feb 01 02:49:57 PM).

Save the result

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



Save the result

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

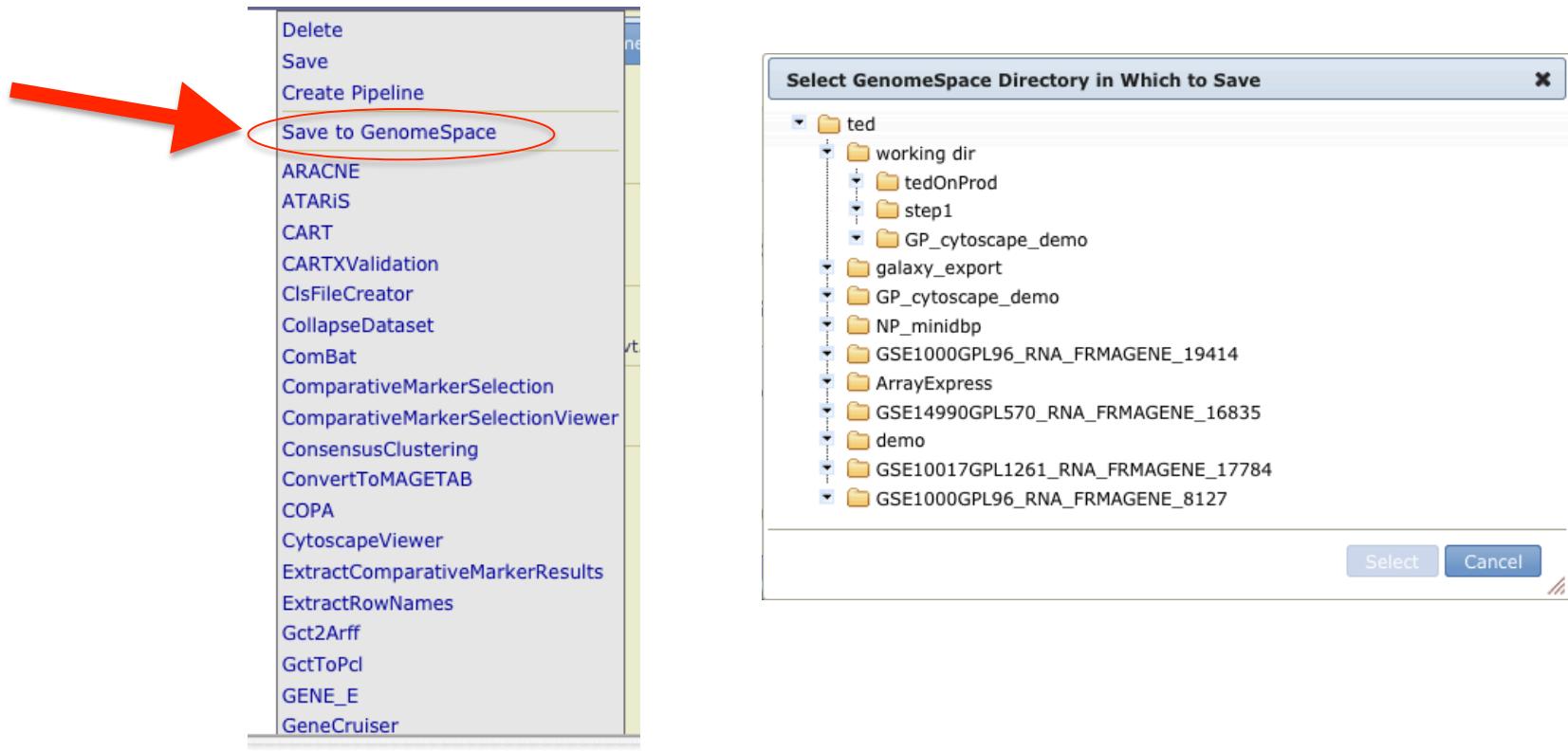
The screenshot shows the GenePattern web interface. On the left, there's a sidebar with a search bar and a list of recently used modules: Recently Used, Annotation, caBIG, Clustering, Data Format Conversion, Flow Cytometry, Gene List Selection, GENE-E, IGV, Image Creators, Missing Value Imputation, MIT_701X, Pathway Analysis, Prediction, Preprocess & Utilities, Projection, Proteomics, RNA-seq, RNAi, Sequence Analysis, SNP Analysis, Statistical Methods, Survival Analysis, Uncategorized, Visualizer, and pipeline. The main content area has a green banner at the top stating "3/8/13 GSEALeadingEdgeViewer updated for compatibility with Java 7." Below it is a "Welcome to GenePattern" header and a "Analyzing genomic data in GenePattern" section. This section includes a "what do you want to do?" list with two bullet points: "Click a protocol to run an analysis. GenePattern guides you step by step." and "Click Quick Start for instructions on how to run any module in GenePattern." It also lists "Protocols for running common analyses in GenePattern:" with icons for "Run an Analysis in GenePattern" (a blue book icon) and "Differential Expression Analysis" (a red and blue heatmap icon). To the right, there's a "Recent Jobs" panel with a list of completed tasks:

Job Type	Date	File Name	Status
PreprocessDataset	(682330)Mar 14 01:31:04 PM	SaveFromGenomica.gct	Success
PreprocessDataset	(682327)Mar 14 01:27:43 PM	SaveFromGenomica.gct	Success
ConvertLineEndings	(676469)Mar 01 01:15:18 PM	BHMutant_HCC1500_4uM24h.cvt.txt	Success
PreprocessDataset	(666128)Feb 07 12:21:20 PM	stderr.txt	Success

A red arrow points to the first "PreprocessDataset" entry in the list.

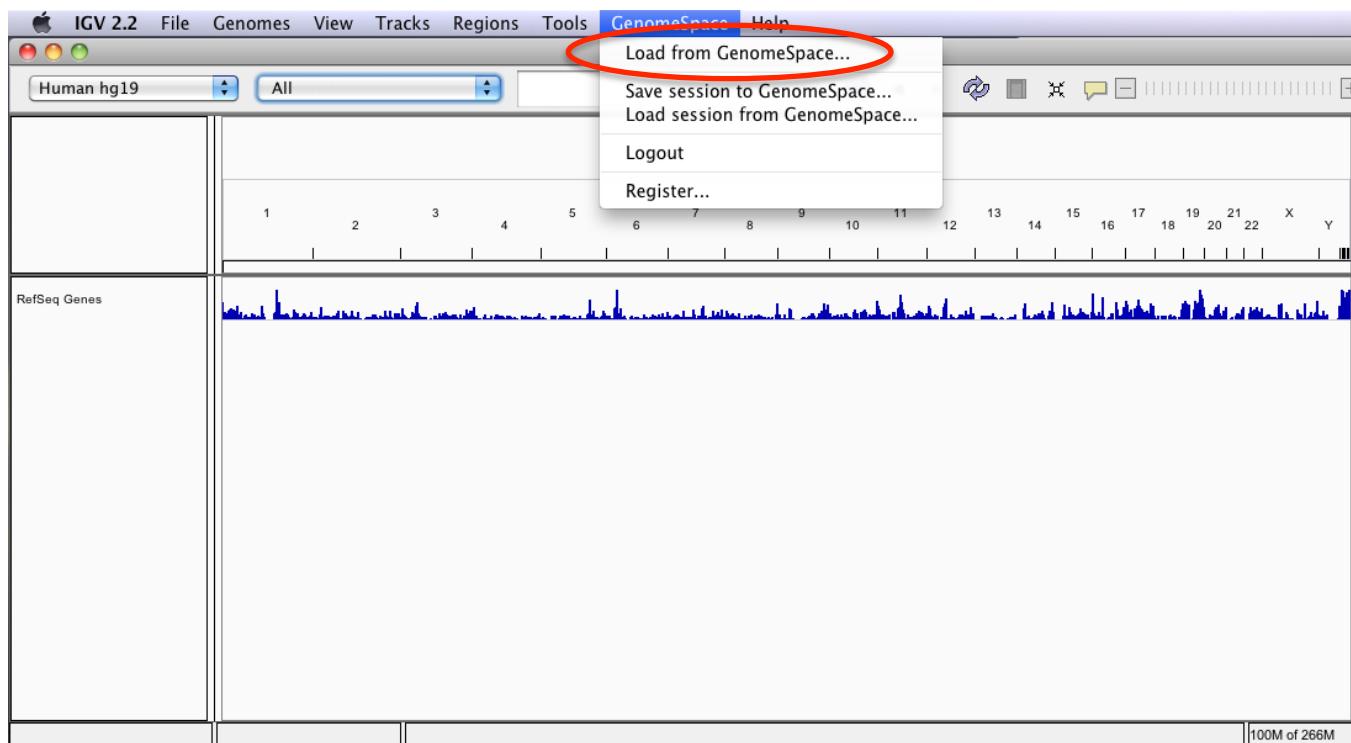
Saving to GenomeSpace

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

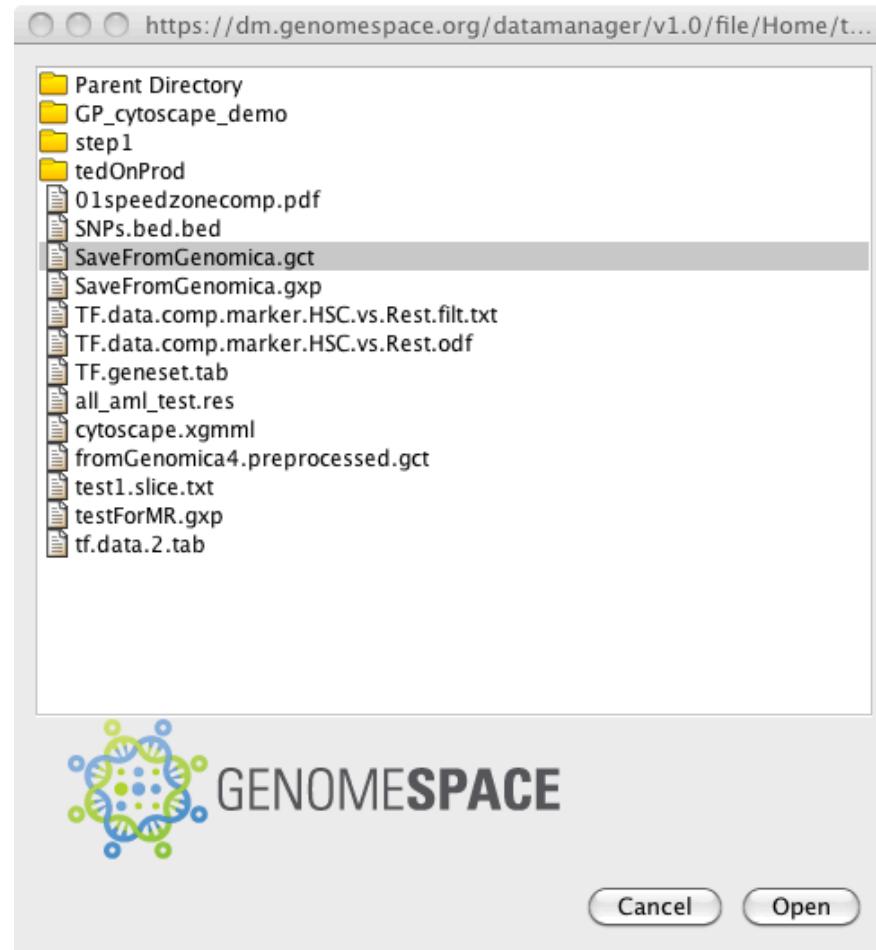


Send to IGV

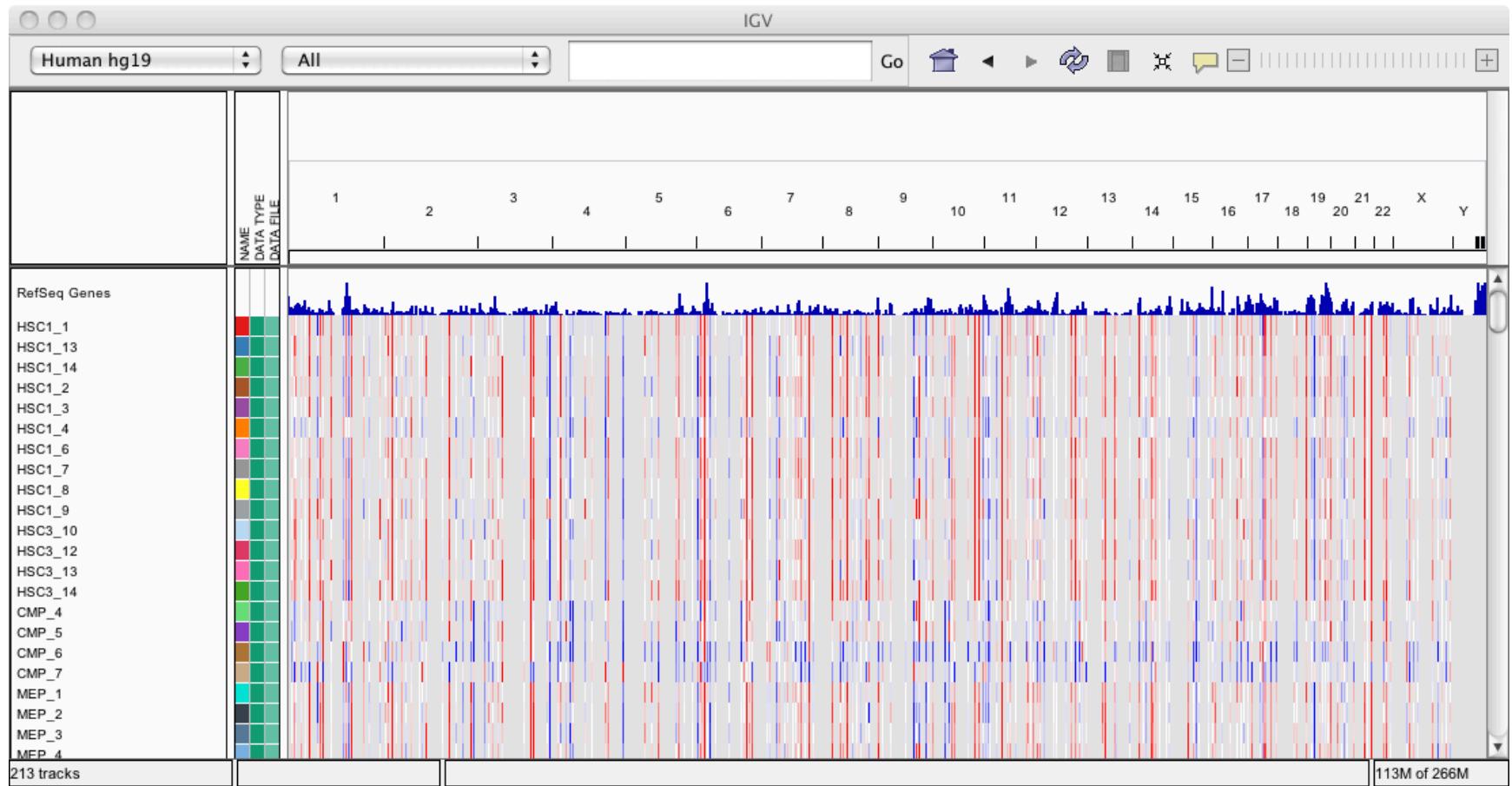
- In the GenomeSpace interface, launch IGV
 - Open the ‘GenomeSpace’ menu and ‘Load from GenomeSpace’



Select your file (from GenePattern)



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

File Preview dialog showing the 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 0.1110710
```

The file contains gene expression data for 523 genes across 211 samples, with columns for sample names and descriptions followed by 10 HSC1 samples and one HSC3 sample.

File Preview options:

- Extract rows / cols
- Convert
- Download
- View file link
- Sharing
- Rename
- Move
- Delete

Recent files:

- all_aml.res
- cytoscape.xgmmml
- fromGenomica4.preprocessed.gct
- test1.slice.txt
- testForMR.gxp
- tf.data.2.tab

Extracting Rows and/or Columns

The screenshot shows the GenomeSpace web interface. At the top, there are tabs for "Welcome to GenomeSpace" and "GenePattern". The URL in the address bar is <https://gsui.genomespace.org/gsui/gsui.html#>. A sidebar on the left contains links for "Upload", "Create Subdirectory", "Preview", "Extract rows and columns" (which is highlighted in blue), "Convert", "Download", "View Link URL", "Sharing", "Rename", "Move", and "Delete". Below this is a tree view of a directory structure under "demo": "demo", "workingDir" (selected), "GP_cytoscape_demo", "step1", "tedOnProd", "Shared to ted", and "Public". The main content area shows a file list titled "Extract rows and columns" with columns for "Filename", "Owner", "Size", and "Last Modified". The file "SaveFromGenomica.gct" is selected, indicated by a checked checkbox and highlighted in yellow. Other files listed include "SaveFromGenomica.gxp", "TF.data.comp.marker.HSC.vs.Rest.filt.txt", "TF.data.comp.marker.HSC.vs.Rest.ofd", "TF.geneset.tab", and "all_aml_test.res".

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
<input checked="" type="checkbox"/> 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.ofd	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

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

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’

Welcome to GenomeSpace

GenePattern

https://gsui.genomespace.org/jgui/gsui.html#

GENOME SPACE BETA

ted

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

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	-0.49470					
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	-0.21882					
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	1.71958					
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	0.03669					
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	0.47363					

Showing up to the first 10 lines of SaveFromGenomica.gct

Save Close

all_am1_test.res ted 1.83 Mb 6 minutes ago

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

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

The screenshot shows the GENOME SPACE BETA application window. At the top, there's a menu bar with File, Launch, View, Manage, Recipes, and Help. Below the menu is a toolbar with icons for Cytoscape, ArrayExpress, and UCSC Table Brow. A dropdown menu from the 'File' menu is open, showing options like Refresh Current Directory, Recent Uploads, and Customize toolbar... The 'Customize toolbar...' option is highlighted. A 'Customize toolbar' dialog box is displayed, containing a list of GenomeSpace Tools. The list includes Cytoscape, ArrayExpress, UCSC Genome Browser, Cistrome, Galaxy, GenePattern, Genomica, geWorkbench, Gitools, IGV, and InSilicoDB. The 'UCSC Genome Browser' entry has an unchecked checkbox next to it, which is circled in red. A large red arrow points from the text 'Drag and drop to reorder GenomeSpace tools.' in the dialog to the list of tools.

Uncheck the tool
To remove it from
The toolbar

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

Revert Save Close

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.

The screenshot shows the ArrayExpress homepage with the following details:

- Header:** The title "ArrayExpress < EMBL-EBI" is at the top left. The URL "www.ebi.ac.uk/arrayexpress/" is in the address bar, with a "Google" search result visible to its right.
- Cookie Notice:** A dark overlay box displays a message about cookies, a "Dismiss this notice" button, and links to "Cookie and Privacy statements".
- Navigation Bar:** Includes links for "EMBL-EBI", "Services", "Research", "Training", "Industry", and "About us".
- Main Content Area:** The title "ArrayExpress" is prominently displayed. Below it is a search bar with examples like "E-MEXP-31, cancer, p53, Geuvadis" and buttons for "Search" and "Advanced".
- Footer:** A navigation menu with links "Home", "Experiments", "Arrays", "Submit", "Help", "About ArrayExpress", "Feedback", and "Login".
- Side Panel:** A sidebar titled "Data Content" shows statistics: "Updated today at 06:00", "36287 experiments", "1050421 assays", and "13.52 TB of archived data".
- Bottom Content:** A main text block explains the database's purpose: "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."



Cistrome

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

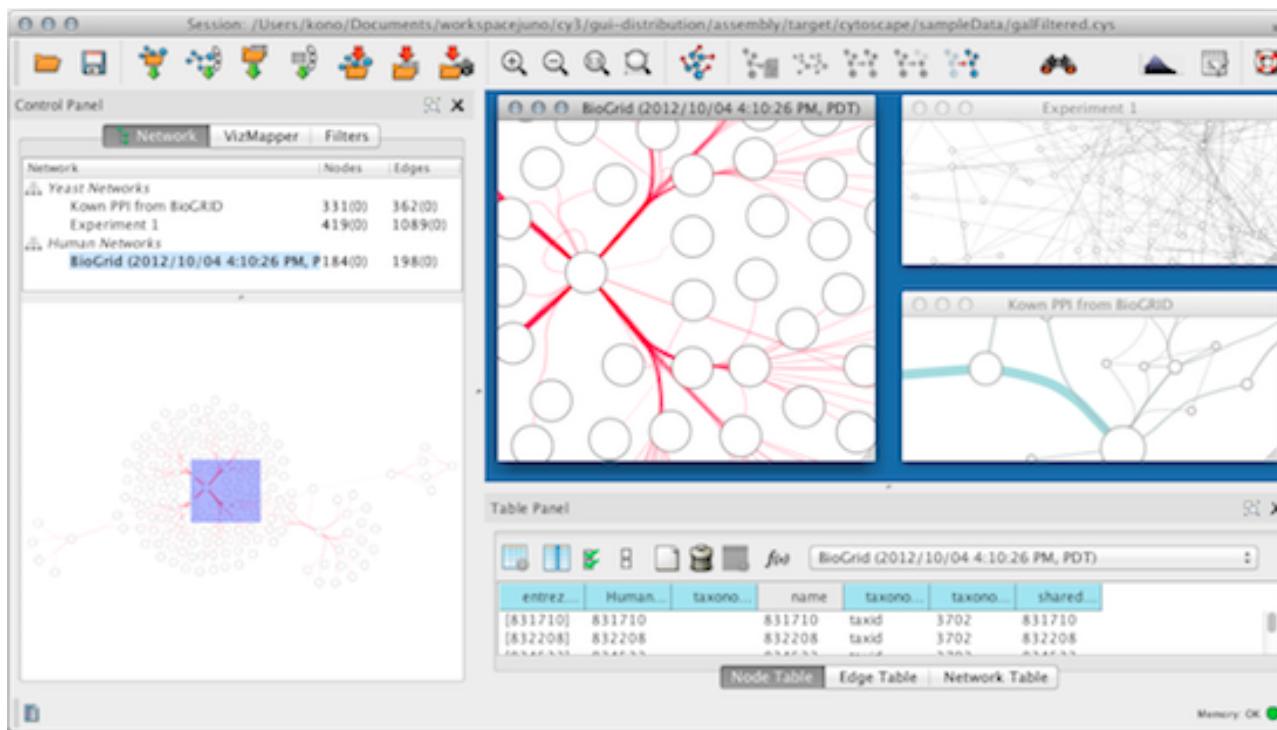
- Preliminary peak calling
- Correlation analyses
- Downstream genome feature association
- Gene expression analyses
- Motif discovery

The screenshot shows a web browser window for the Galaxy / Cistrome interface. The title bar reads "Galaxy / Cistrome". The address bar shows the URL "cistrome.org/ap/". The top navigation bar includes links for "Analyze Data", "Workflow", "Shared Data", "Lab", "Visualization", "Help", and "User". On the left, there's a "Tools" sidebar with a search bar and sections for "CISTROME TOOLBOX" (Import Data, Data Preprocessing, Gene Expression, Integrative Analysis, Liftover/Others) and "GALAXY TOOLBOX" (Get Data, Text Manipulation, Filter and Sort, Join, Subtract and Group, Convert Formats, Extract Features). The main content area displays the "Cistrome Analysis Pipeline (AP) Module". It describes the module's purpose of organizing analysis tools into a pipeline for users. It includes a "Full list of deployed Cistrome tools" link and a "Quick Links to Demonstrations" section. Below this is a link to "A demonstration of ChIP-seq analysis with expression data.". A "Recent News" section at the bottom shows the date "Mar 13, 2012" and the note "Genome Space supported". On the right, there's a "History" panel showing an "Unnamed history" entry for "1: GenomeSpace import on TF.data.gct" (1.3 MB).



Cytoscape

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





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.

The screenshot shows the Galaxy web interface. The top navigation bar includes tabs for "Galaxy", "Analyze Data", "Workflow", "Shared Data", "Visualization", "Cloud", "Help", and "User". A message at the top states: "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." The left sidebar contains a "Tools" section with a search bar and a list of tools: Get Data, Send Data, ENCODE Tools, Lift-Over, Text Manipulation, Convert Formats, FASTA manipulation, Filter and Sort, Join, Subtract and Group, Extract Features, Fetch Sequences, Fetch Alignments, Get Genomic Scores, and Operate on Genomic Intervals. The main content area features a large banner for the "Galaxy OSLO Conference 2012" with the text "Galaxy is hiring". Below the banner is a "Live Quickies" section with four cards: "Uploading Data using FTP", "Managing account histories", "454 Mapping: Single End", and "Get fastQ population:". The right sidebar displays a "History" list of imported datasets:

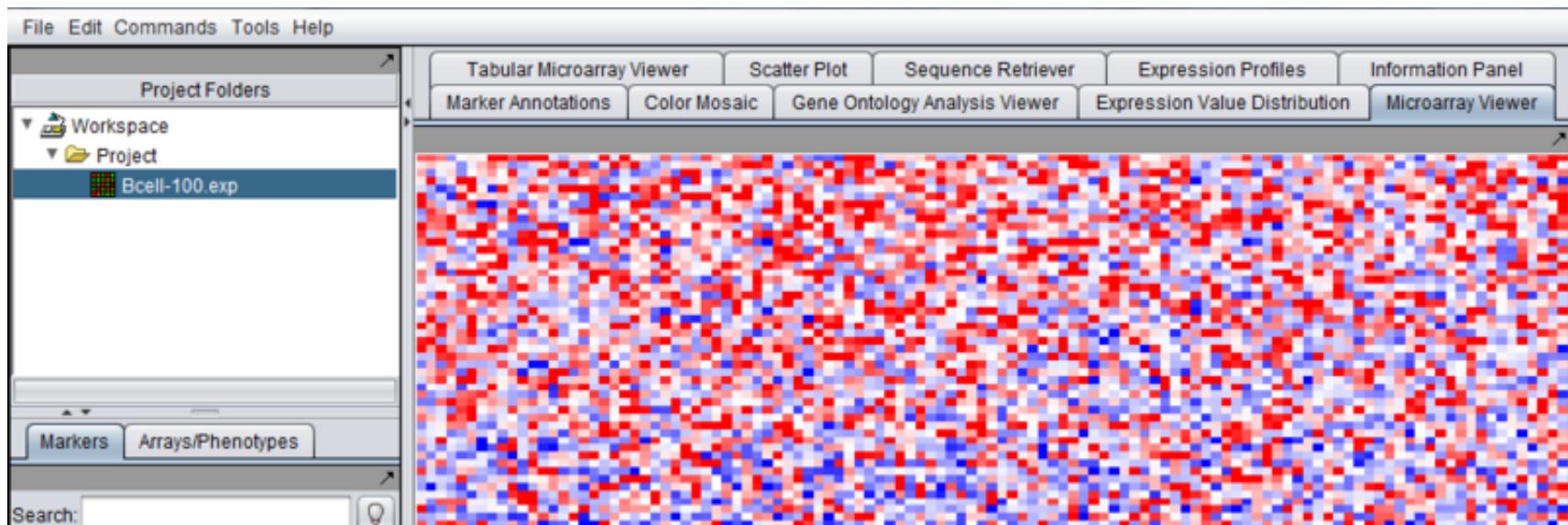
- imported: Exons vs Repeats (416.4 MB)
- 10: GenomeSpace importer on CCLE Expression Entrez 2012-04-06.res
- 9: GenomeSpace importer on SNPs.bed.bed
- 8: GenomeSpace importer on all aml test.res
- 7: top 5 exons
- 2: UCSC Main on Human: rmask (genome)
- 1: UCSC Main on Human



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)





InSilicoDB

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

The screenshot shows the InSilicoDB web interface. At the top, there is a navigation bar with the InSilico logo, a search bar containing "gse14990", and user information like "Samples basket" and "liefeld@broadinstitute.org". Below the navigation bar is a search results table.

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

On the left side, there is a sidebar titled "Filters" with several sections:

- DataSets source:** Includes checkboxes for "My safe" (checked) and "Public" (checked, [6/6]).
- Curation:** Includes checkboxes for "Manually curated" (checked) and "To curate" (unchecked).
- Platforms:** Includes checkboxes for "MicroArray" (checked, [13/13]) and "High Throughput Sequencing" (checked, [13/13]).
- Measurement type:** Includes checkboxes for "MicroArray" (checked, [1/1]) and "High Throughput Sequencing" (checked, [3/3]).

At the bottom of the interface, there are buttons for "Edit/Show clinical annotation", "Download", "Export", and social sharing links for Twitter and Email.



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 add custom tracks track hubs

table: knownGene describe table schema

region: genome position chr21:33,031,597-33,041,570 lookup define regions

identifiers (names/accessions): paste list upload list

filter: create

intersection: create

correlation: create

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

get output summary/statistics

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

The screenshot shows the MSigDB v4.0 homepage. At the top, there's a navigation bar with links for 'Most Visited', 'GP Web', 'jQuery UI', 'GP Prod', 'Bookmarks', 'login', and 'register'. The 'Molecular Signatures Database' tab is highlighted. Below the navigation is a banner for 'Gene Set Enrichment Analysis' (GSEA) with a blue mountain graphic. The main content area has a light blue header with the 'MSigDB' logo and 'Molecular Signatures Database'. The page is divided into several sections: 'Overview' (describing the database as a collection of annotated gene sets for GSEA), 'Collections' (listing 7 major collections: c1, c2, c3, c4, and c5), and 'Registration' (link for users to register). The 'Collections' section contains detailed descriptions of each type of gene set:

- 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