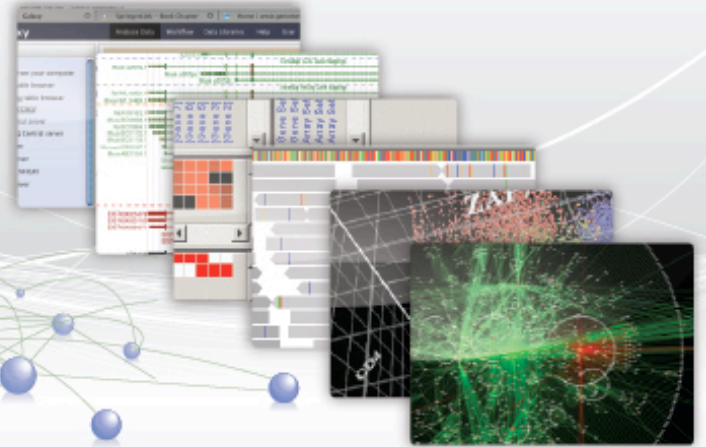


GENOMESPACE



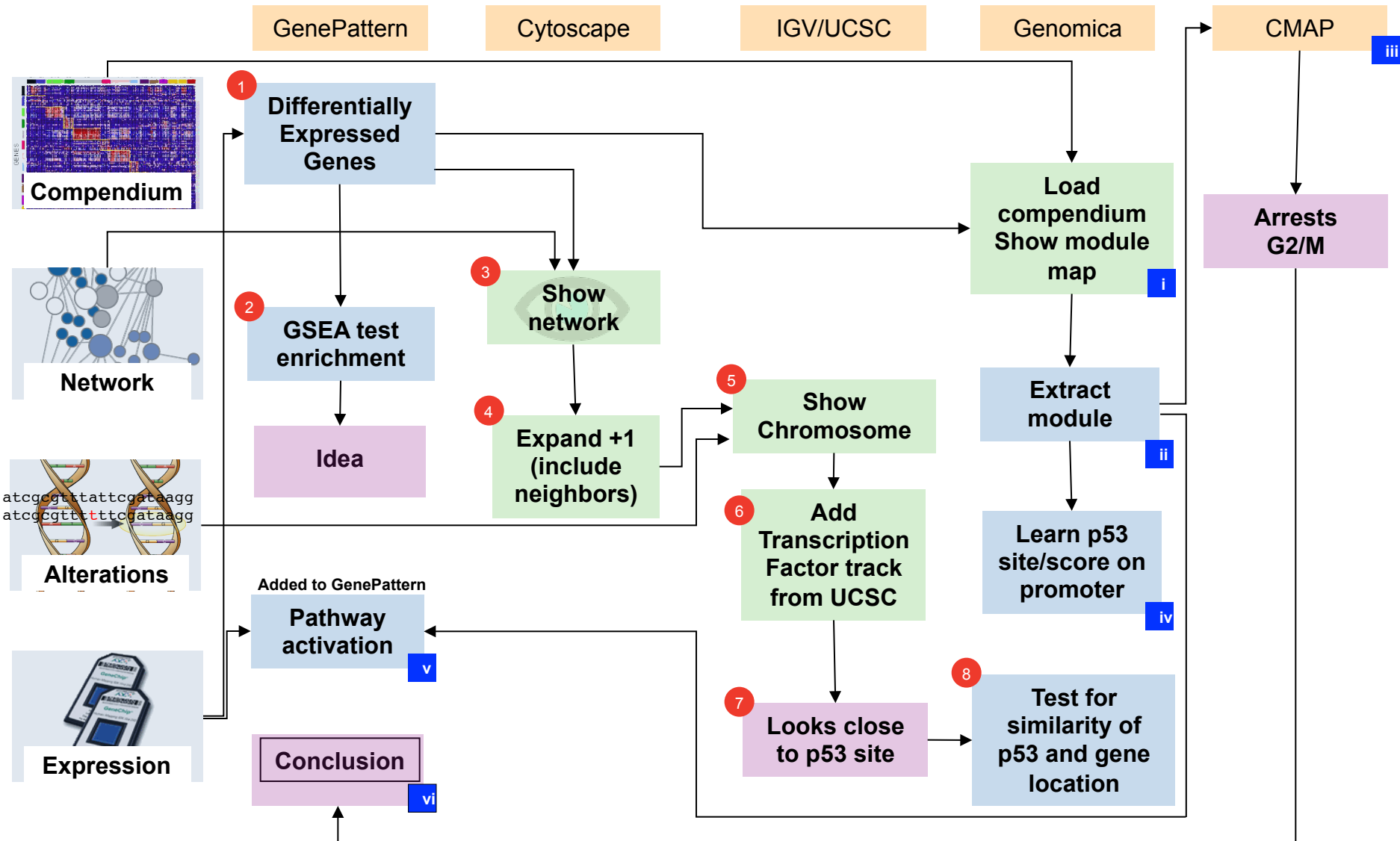
GENOMESPACE



Outline

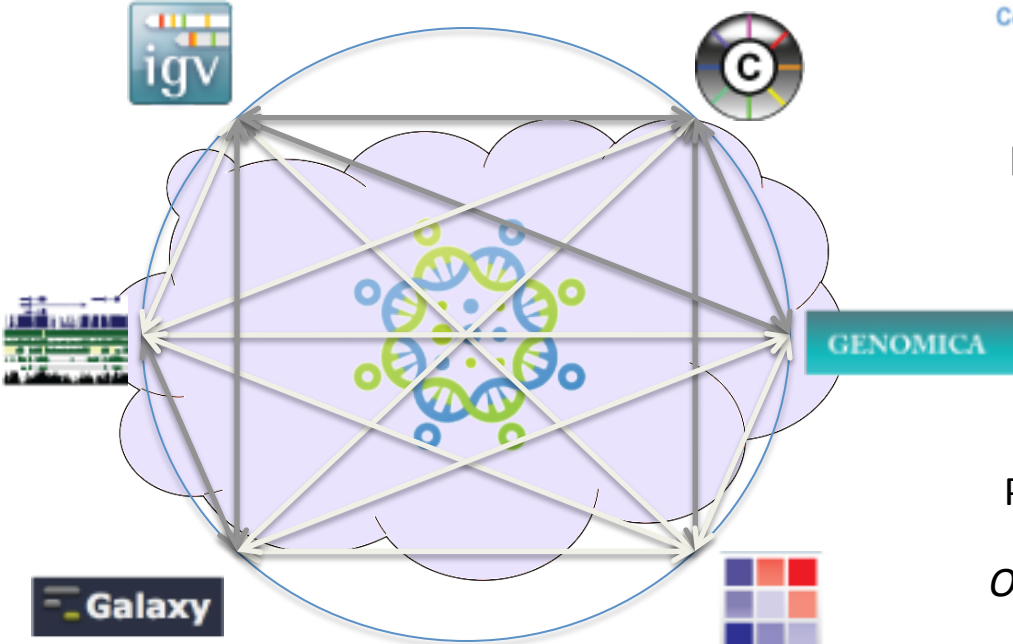
- Introduction to GenomeSpace
- GenomeSpace Tools and Recipes
- GenomeSpace User Interface
- Integrative analysis exercise
- Other GenomeSpace Tools
- GenomeSpace development
- Q and A

The vision: Integrative Translational Genomics



Online community to share diverse computational tools

GENOMESPACE



Seed Tools

- Cytoscape
- Galaxy
- GenePattern
- Genomica
- IGV
- UCSC Browser

Outreach to new tools

Driving Biological Projects

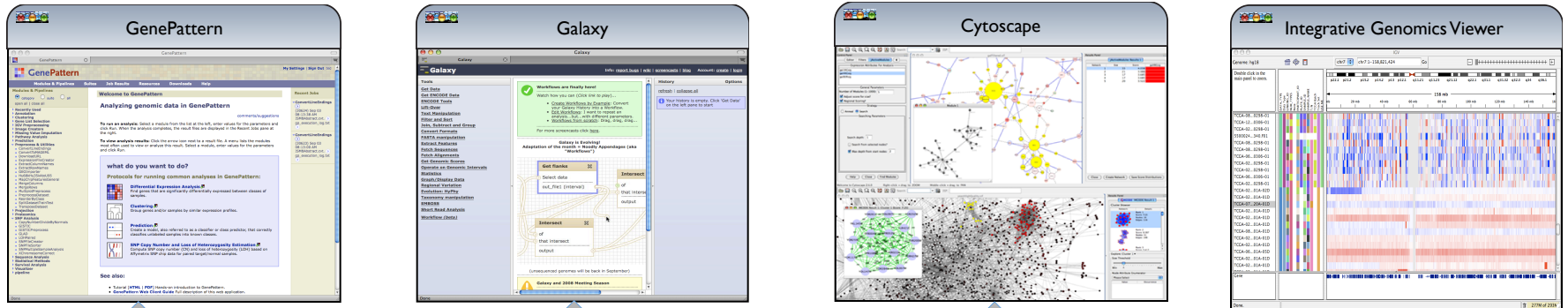
- lincRNAs
- Cancer stem cells
- Patient Stratification
- Outreach to new DBPs*

www.genomespace.org

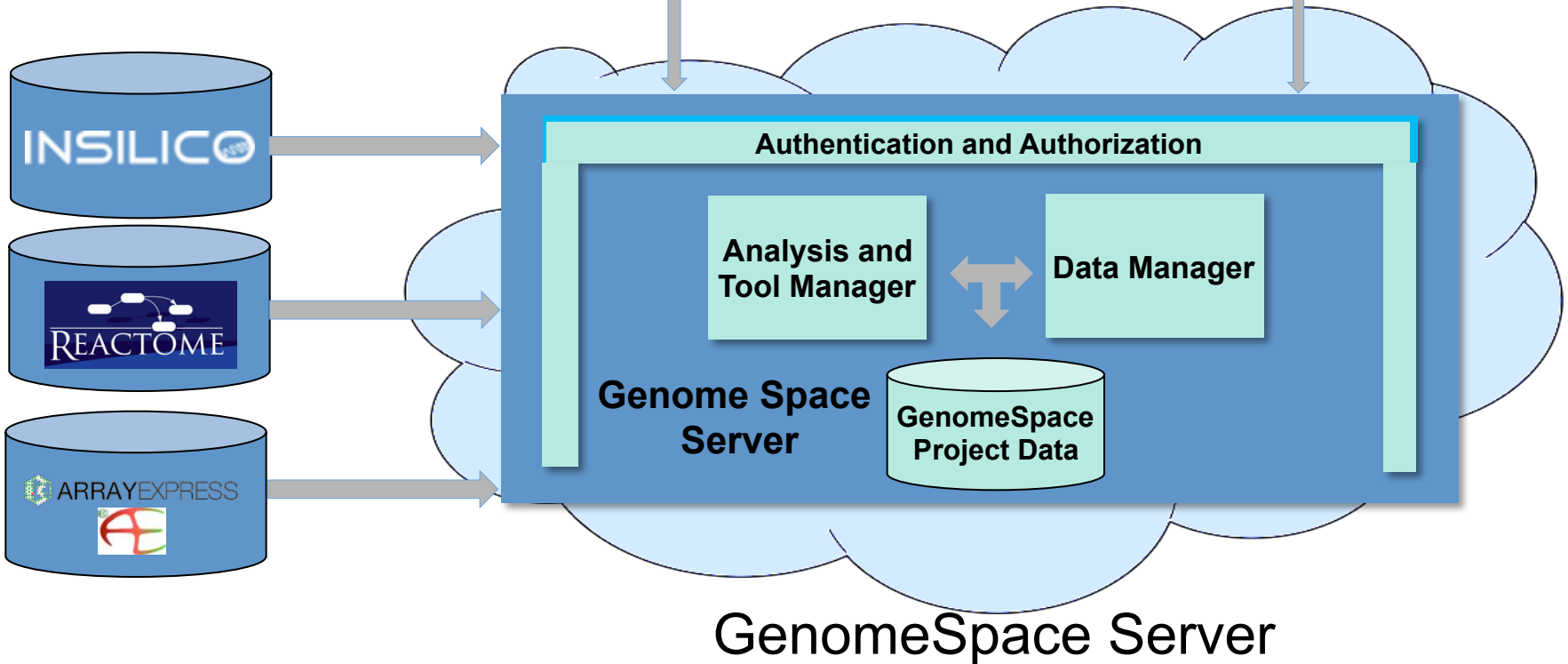
GenomeSpace: a connection layer between integrative analysis tools

- 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 image shows a screenshot of the GenomeSpace website homepage. The browser's address bar displays "www.genomespace.org". The page features a navigation menu with links for "What is GenomeSpace?", "Tools", "Documentation", "Developers", "Support", and "About". A search bar is located to the right of the navigation menu. The main content area has a large banner with the text "GENOMESPACE BETA" and "Frictionless connection of bioinformatics tools". Below this banner are two buttons: "Register" (highlighted with a red circle) and "User Login". The "Register" button is blue with white text, and the "User Login" button is green with white text. Below the banner, there is a "STATUS" section indicating that all systems are operating normally. To the right of the status section are social media icons for Twitter, Facebook, RSS, and Email. Below the status section is a "WHAT'S NEW" section with tabs for "News Highlights" and "GenomeSpace Team Blog". The "News Highlights" tab is selected, showing a news item titled "GenomeSpace at BOSC-Broad Interoperability Hackathon". To the right of the "WHAT'S NEW" section is a "Mentions" section for "@genomespace", showing a tweet from genomespace about the VIZBI workshop.

www.genomespace.org

Register

www.genomespace.org/register

Wikipedia (en)

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

GenomeSpace: Register

GENOMESPACE

User Login

What is GenomeSpace? Tools Documentation Developers Support About

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

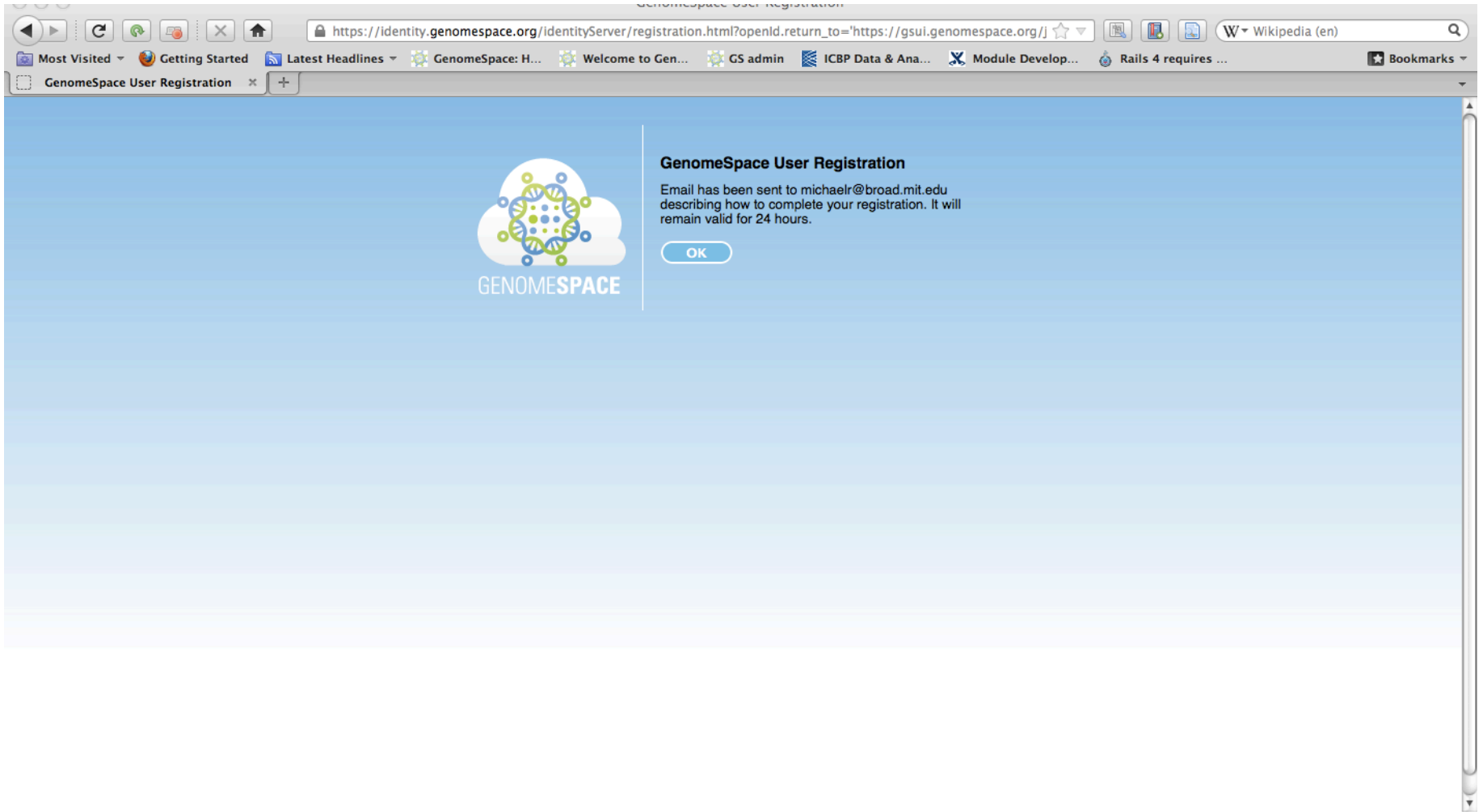
Email

Password

Password confirm

Register

Register



The screenshot shows a web browser window with the following elements:

- Address Bar:** `https://identity.genomespace.org/identityServer/registration.html?openId.return_to=https://gsui.genomespace.org/j`
- Browser Tabs:** "GenomeSpace User Registration" (active), "Getting Started", "Latest Headlines", "GenomeSpace: H...", "Welcome to Gen...", "GS admin", "ICBP Data & Ana...", "Module Develop...", "Rails 4 requires ..."
- Page Content:**
 - Logo:** The GenomeSpace logo, featuring a stylized DNA double helix with blue and green nodes, set against a white cloud shape. Below it is the text "GENOMESPACE".
 - Section Header:** "GenomeSpace User Registration"
 - Text:** "Email has been sent to michaelr@broad.mit.edu describing how to complete your registration. It will remain valid for 24 hours."
 - Button:** A white button with rounded corners and a thin blue border containing the text "OK".

Register

The screenshot shows a Gmail interface with a navigation bar at the top containing 'Broad Institute' and 'michaelr@broadinstitute.org'. The main content area displays an email titled 'GenomSpace user registration' from 'genomspacenotifier@gmail.com' received at 4:14 PM. The email body contains the following text:

Hello. To complete the registration of the GenomSpace account for **GenomSpace**, please follow this link:

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

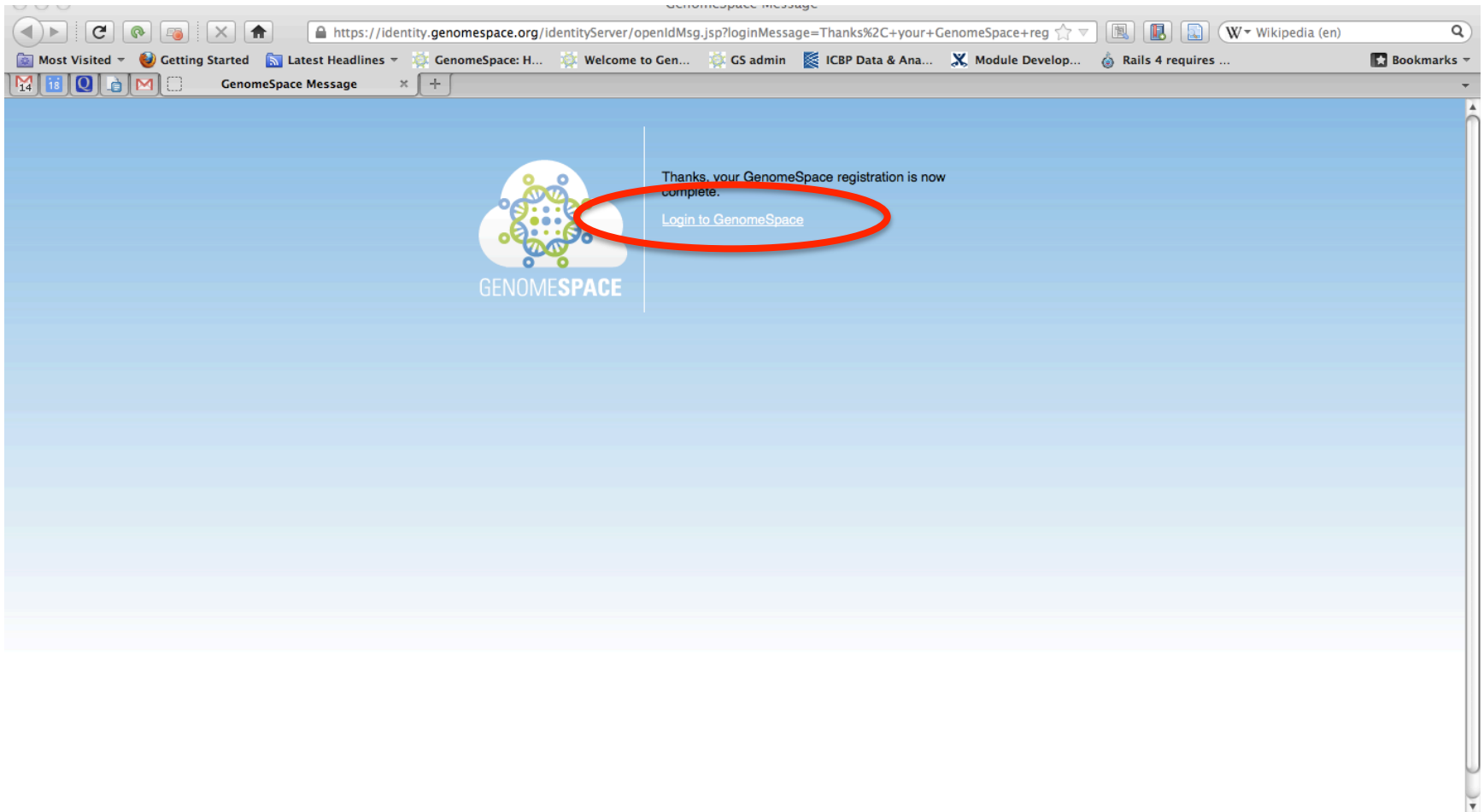
You can also copy and paste it into your browser's address bar.

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

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

At the bottom of the email, there are links for 'Reply', 'Reply to all', and 'Forward'. The footer of the email includes '©2013 Google - Terms of Service - Privacy Policy - Program Policies' and 'Powered by Google'. The Gmail interface also shows a left sidebar with 'Inbox (14)', 'Drafts (45)', and 'Events', and a right sidebar with 'People (2)' and 'genomspacenotifier'.

Login



The screenshot shows a web browser window with the following elements:


- Address Bar:** `https://identity.genomespace.org/identityServer/openIdMsg.jsp?loginMessage=Thanks%2C+your+GenomeSpace+reg`
- Browser Tabs:** Most Visited, Getting Started, Latest Headlines, GenomeSpace: H..., Welcome to Gen..., GS admin, ICBP Data & Ana..., Module Develop..., Rails 4 requires ...
- Page Content:**
 - Logo:** GENOMESPACE logo featuring a stylized DNA double helix and a cloud.
 - Text:** "Thanks, your GenomeSpace registration is now complete."
 - Link:** "Login to GenomeSpace" (highlighted with a red circle).

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:

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

GenomeSpace UI

The screenshot displays the GenomeSpace user interface within a web browser. The browser's address bar shows the URL `https://gsui.genomespace.org/jsui/gsui.html`. The page header includes the GenomeSpace logo and the text "GENOMESPACE BETA". A navigation menu contains "File", "Launch", "View", "Manage", "Recipes", and "Help". Below this is a horizontal toolbar with icons for various tools: Cistrome, Cytoscape, Galaxy, GenePattern, Genomica, geWorkbench, IGV, InSilicoDB, UCSC Table Browser, ArrayExpress, and Gitools. The main content area is divided into two sections. On the left, under "Directories", there is a "Home" section with sub-items: "mmr", "Shared to mmr", and "Public". On the right, a table lists files and directories:

<input type="checkbox"/>	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



geWorkbench



Galaxy



Gitoools



Cistrome



IGV



Cytoscape



InSilicoDB



GenePattern



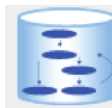
Genomica



UCSC Table Browser



ISAcreeator



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.

The screenshot displays the Cytoscape software interface. The main window shows a network visualization with nodes and edges. The interface is divided into several panels:

- Control Panel:** Contains a 'Network' tab with a table listing networks and their node/edge counts.
- Network Visualization:** A large central window showing a network graph with nodes and edges. A smaller window on the right shows a zoomed-in view of a specific network.
- Table Panel:** A table at the bottom showing data for the selected network.

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)

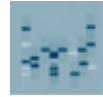
entrez...	Human...	taxono...	name	taxono...	taxono...	shared...
[831710]	831710		831710	taxid	3702	831710
[832208]	832208		832208	taxid	3702	832208



Galaxy

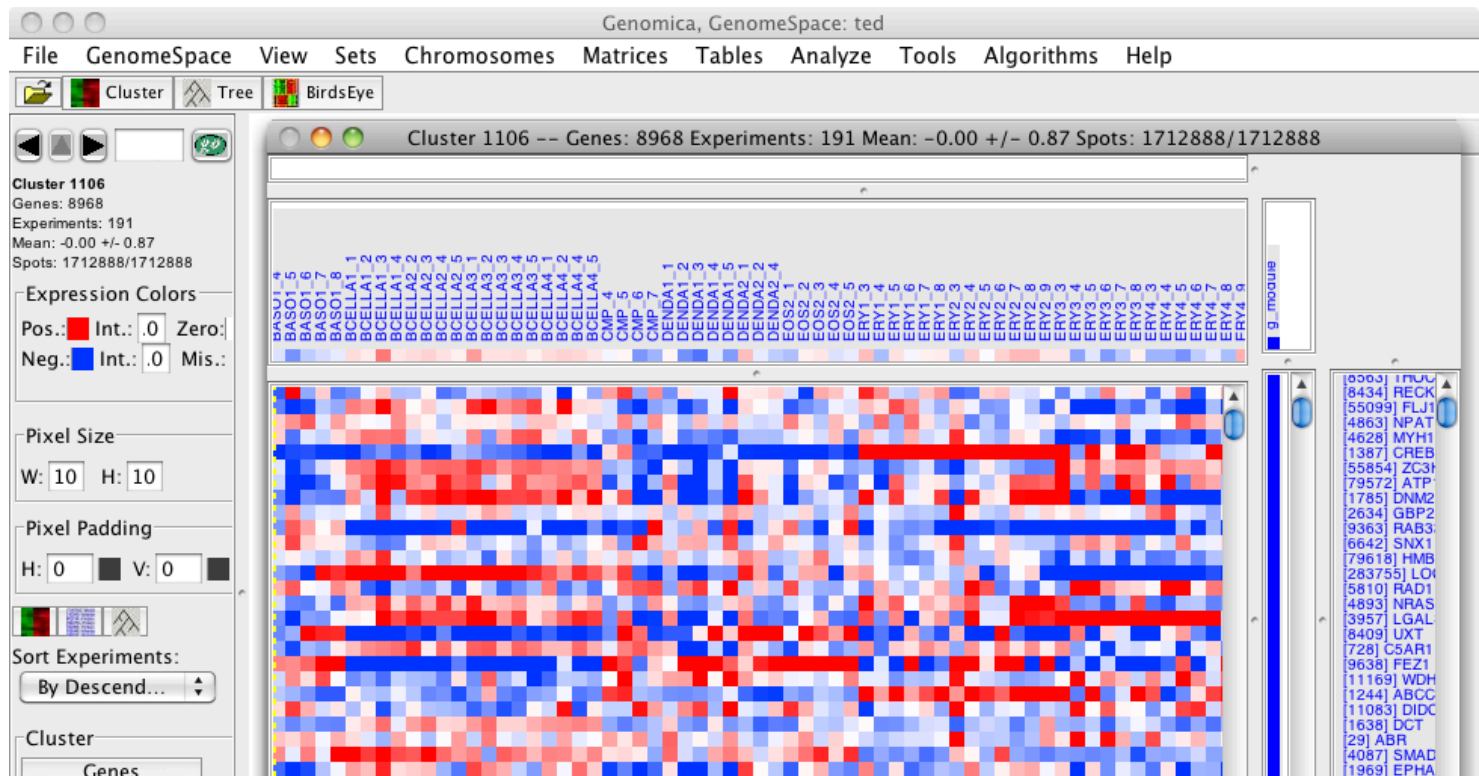
Galaxy is an open-source, scalable framework for tool integration that allows users to analyze multiple alignments, compare genomic annotations, and profile metagenomic samples, among many possible analyses; workflows allow the linking together of analyses.

The screenshot shows the Galaxy web interface in a browser window. The address bar displays `https://main.g2.bx.psu.edu`. The navigation menu includes **Analyze Data**, **Workflow**, **Shared Data**, **Visualization**, **Cloud**, **Help**, and **User**. A yellow banner at the top indicates a server shutdown: "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 main content area features a central graphic with the text "Galaxy is hiring" and "OSLO" (University of Oslo). The left sidebar lists various tool categories such as "Get Data", "Send Data", "ENCODE Tools", and "Text Manipulation". The right sidebar shows a "History" panel with a list of recent jobs, including "imported: Exons vs Repeats" (416.4 MB) and several "GenomeSpace importer" jobs. Below the main content, there is a "Live Quickies" section with four featured jobs: "Imported fastQ", "454 Mapping: Single End", "Uploading Data using FTP", and "Managing account histories".



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 interface. The browser address bar displays `genepattern.broadinstitute.org/gp/pages/index.jsf`. The page features a navigation menu with options: Modules & Pipelines, Suites, Job Results, Resources, Downloads, Help, and GenomeSpace. A search bar is located under 'Modules & Pipelines'. The main content area includes a notification for '3/8/13 GSEALeadingEdgeViewer updated for compatibility with Java 7.', a 'Welcome to GenePattern' message, and a section titled 'Analyzing genomic data in GenePattern' with the question 'what do you want to do?'. This section lists two options: '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.' Below this, there are two protocol cards: '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.) and 'Differential Expression Analysis' (Find genes that are significantly differentially expressed between classes of). On the right side, there is a 'Recent Jobs' panel with a list of jobs including 'ConvertLineEndings', 'PreprocessDataset', and 'ExtractColumnName'.



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 website homepage. At the top, there is a browser window with the address bar showing 'www.ebi.ac.uk/arrayexpress/'. Below the browser window, there is a dark grey banner with a cookie notice: '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'. Below the banner, there is a navigation bar with the EMBL-EBI logo and the text 'ArrayExpress'. The navigation bar includes links for 'Services', 'Research', 'Training', 'Industry', and 'About us'. Below the navigation bar, there is a search bar with the text 'Examples: E-MEXP-31, cancer, p53, Geuvadis' and a 'Search' button. Below the search bar, there is a secondary navigation bar with links for 'Home', 'Experiments', 'Arrays', 'Submit', 'Help', 'About ArrayExpress', 'Feedback', and 'Login'. The main content area features the heading 'ArrayExpress - functional genomics data' and a paragraph describing the database. To the right, there is a 'Data Content' section with a bar chart icon and the text 'Updated today at 06:00'. Below this, there is a list of statistics: '36287 experiments', '1050421 assays', and '13.52 TB of archived data'.

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.

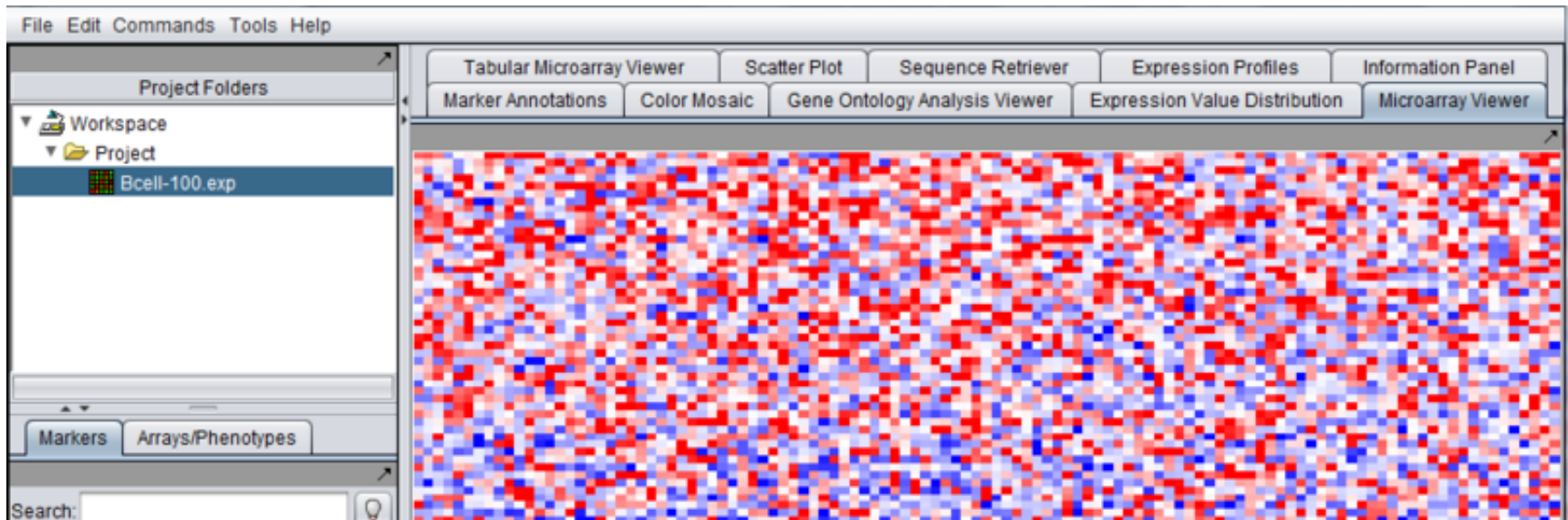
Data Content
Updated today at 06:00

- 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.

A screenshot of a web browser displaying the Galaxy/Cistrome interface. The browser's address bar shows 'cistrome.org/ap/root'. The page features a dark navigation bar with 'Galaxy / Cistrome' and several menu items: 'Analyze Data', 'Workflow', 'Shared Data', 'Lab', 'Visualization', 'Help', and 'User'. A status indicator on the right of the navigation bar says 'Using 0 bytes'. On the left, there is a 'Tools' sidebar with a search box and a list of tool categories: 'CISTROME TOOLBOX' (with sub-items: 'Import Data', 'Data Preprocessing', 'Gene Expression', 'Integrative Analysis', 'Liftover/Others') and 'GALAXY TOOLBOX' (with sub-item: 'Get Data'). The main content area is dominated by a large blue error message box with an information icon. The message reads: 'Galaxy/Cistrome! is experiencing problems'. Below this, it states: '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!'. An 'UPDATE' section follows: '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**'. The final paragraph says: '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.' On the right side of the interface, there is a 'History' panel showing 'Unnamed history' with '0 bytes' and a message: 'Your history is empty. Click 'Get Data' on the left pane to start'.



Gitools

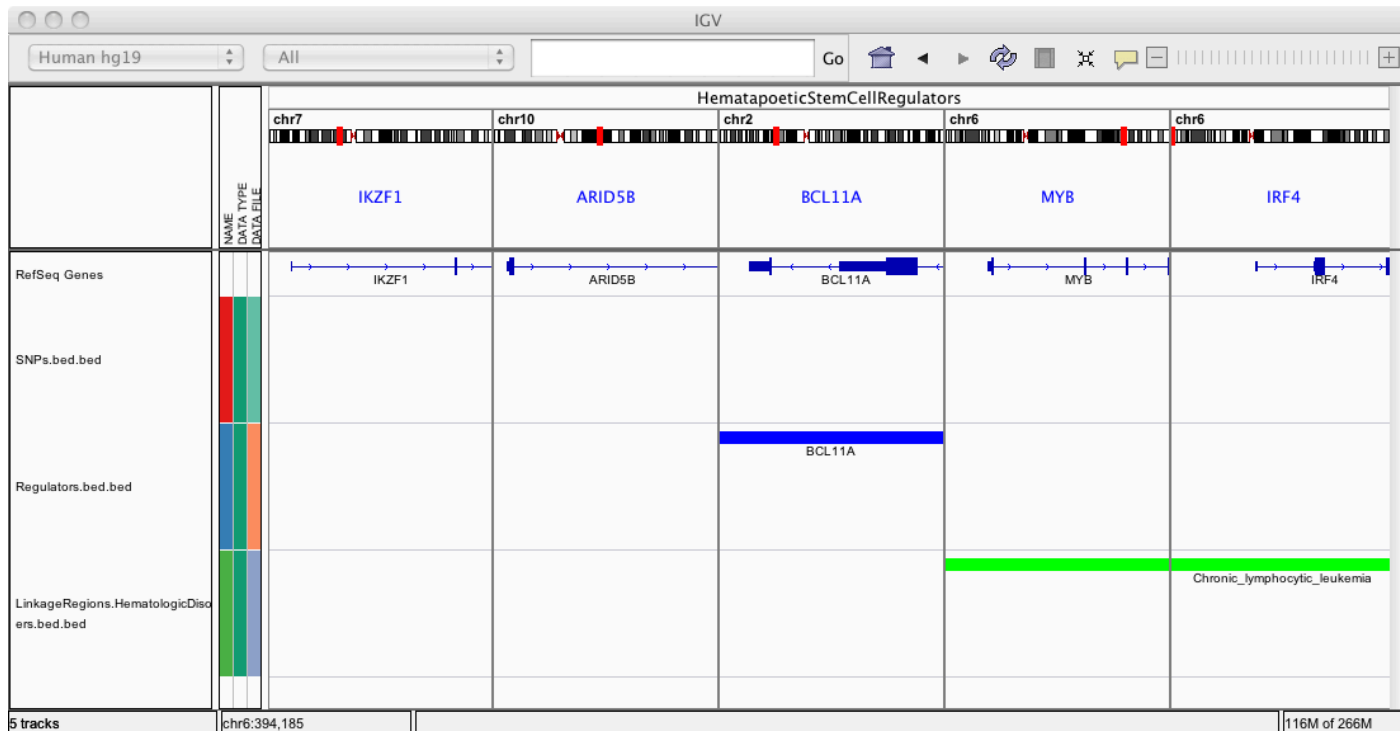
- Gitools is a framework for analysis and visualization of genomic data using interactive heatmaps.





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 displays the InSilicoDB web interface. At the top, there is a navigation bar with the InSilicoDB logo, a search bar containing 'gse14990', and a 'Samples basket' for 'liefeld@broadinstitute.org'. Below the navigation bar, the main content area is divided into a left sidebar and a main panel. The sidebar contains filter sections: 'DataSets source' with 'My safe' (2/2) and 'Public' (6/6); 'Curation' with 'Manually curated' (checked) and 'To curate' (unchecked); 'Platforms' with 'MicroArray' (13/13) and 'High Throughput Sequencing' (13/13); and 'Measurement type' with 'MicroArray' (1/1) and 'High Throughput Sequencing' (3/3). The main panel shows a table with one dataset entry: 'GSE14990 MYC regulation of a "poor prognosis" metastatic cancer cell state'. The table has columns for 'Dataset', 'Sharing', 'Relevance', and '#Samples'. The dataset is public, has 100% relevance, and 15 samples. Below the table, there are buttons for 'Edit/Show clinical annotation', 'Download', and 'Export', along with social media icons for Twitter and Email.

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



UCSC Table Browser

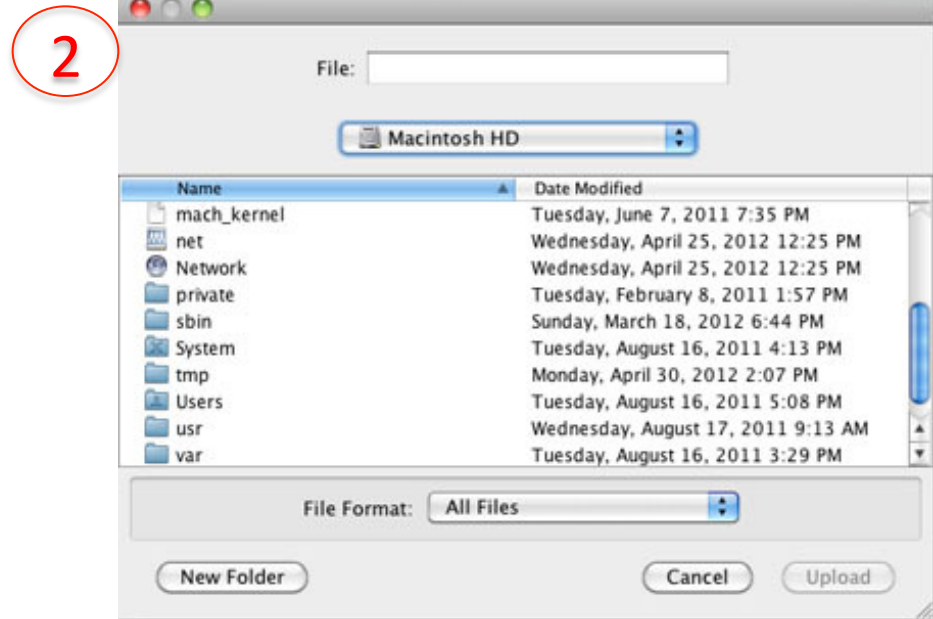
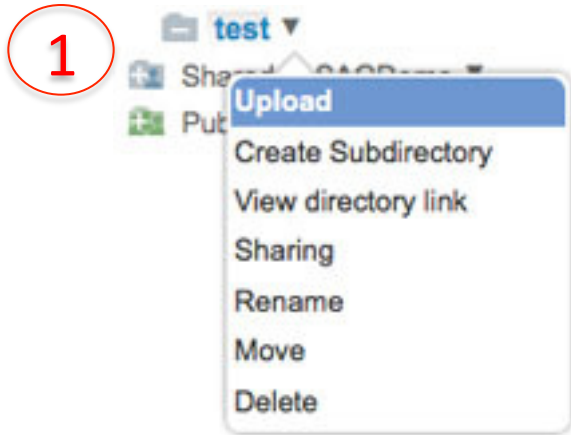
The Table Browser allows you to retrieve data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. After you select the options for your output file, you can opt to send your output file to your GenomeSpace cloud storage.

The screenshot shows the UCSC Table Browser interface. At the top is a navigation bar with links: Home, Genomes, Genome Browser, Blat, Tables, Gene Sorter, PCR, Session, FAQ, Help. Below this is the 'Table Browser' section. It contains a paragraph of introductory text and a series of controls for configuring a query. The controls include dropdown menus for 'clade' (Mammal), 'genome' (Human), and 'assembly' (Feb. 2009 (GRCh37/hg19)). There are also buttons for 'group' (Genes and Gene Prediction Tracks), 'track' (UCSC Genes), and 'table' (knownGene). A 'region' section has radio buttons for 'genome' and 'position', with a text input for 'chr21:33,031,597-33,041,570'. Below this are buttons for 'identifiers (names/accessions)', 'filter', 'intersection', and 'correlation'. The 'output format' is set to 'all fields from selected table'. The 'Send output to' section has checkboxes for 'Galaxy', 'GREAT', and 'GenomeSpace', with a red arrow pointing to the 'GenomeSpace' checkbox. The 'output file' field is empty, and the 'file type returned' is set to 'plain text'. At the bottom, there are buttons for 'get output' and 'summary/statistics', and a note: 'To reset all user cart settings (including custom tracks), [click here](#).'

Basic GenomeSpace recipes

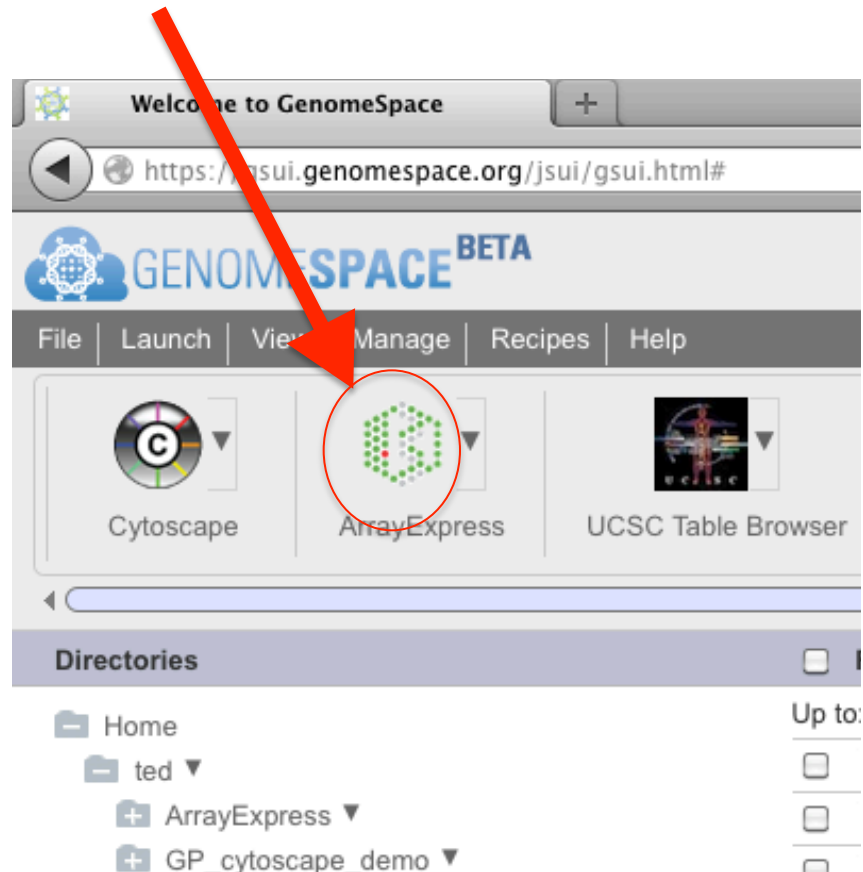
- Uploading data
- Launching tools
- Transitioning across tools

Uploading Data



Launching tools

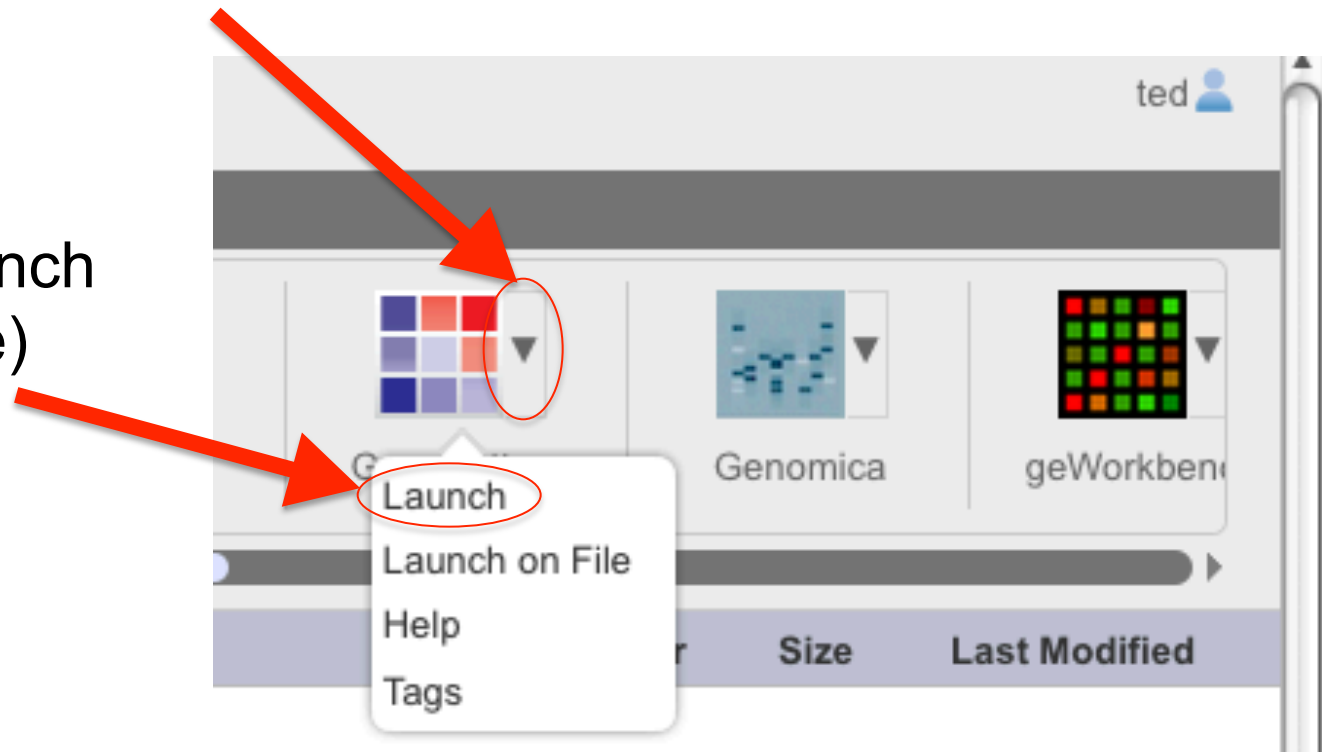
Click on the tool's icon



Launching tools

Open the tool's context menu

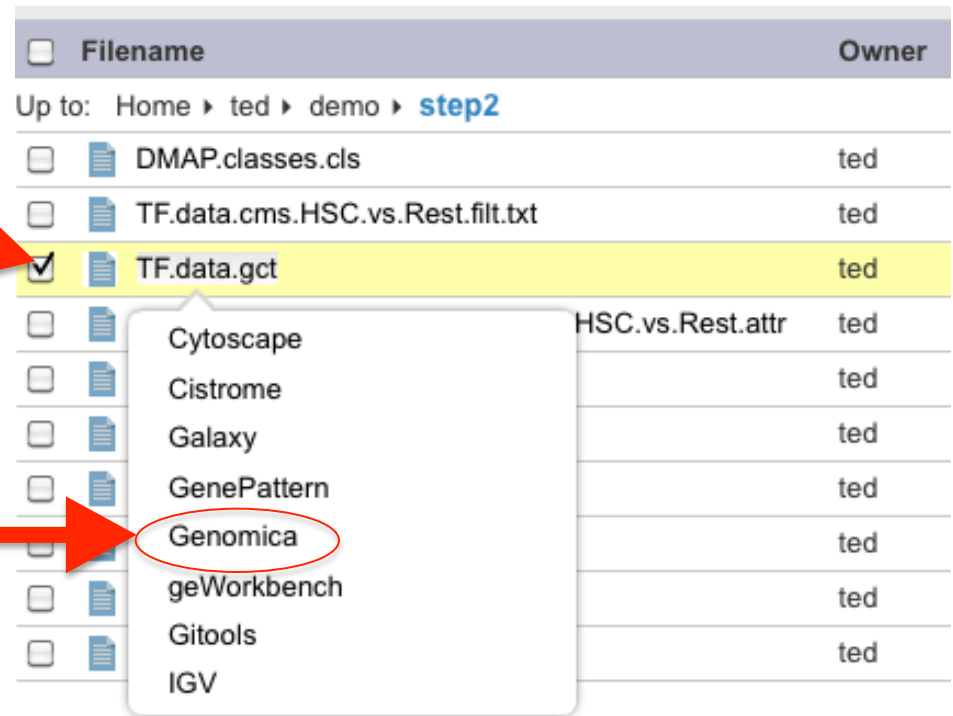
Then click on Launch
(or Launch on File)



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



The screenshot shows a file manager interface with a table of files. The file 'TF.data.gct' is selected, and a context menu is open over it. The menu items are: Cytoscape, Cistrome, Galaxy, GenePattern, Genomica (circled in red), geWorkbench, Gitools, and IGV. A red arrow points from the text 'Click the checkbox for one (or more) files' to the checkbox of 'TF.data.gct'. Another red arrow points from the text 'Then click on one of the files to get the Launch menu and pick your tool' to the 'Genomica' menu item.

<input type="checkbox"/>	Filename	Owner
Up to: Home ▶ ted ▶ demo ▶ step2		
<input type="checkbox"/>	DMAP.classes.cls	ted
<input type="checkbox"/>	TF.data.cms.HSC.vs.Rest.filt.txt	ted
<input checked="" type="checkbox"/>	TF.data.gct	ted
<input type="checkbox"/>	HSC.vs.Rest.attr	ted
<input type="checkbox"/>		ted
<input type="checkbox"/>		ted
<input type="checkbox"/>		ted
<input type="checkbox"/>		ted
<input type="checkbox"/>		ted
<input type="checkbox"/>		ted
<input type="checkbox"/>		ted

- Cytoscape
- Cistrome
- Galaxy
- GenePattern
- Genomica
- geWorkbench
- Gitools
- IGV

Launching tools

The screenshot shows the Galaxy web interface. At the top, there is a toolbar with icons for various tools: Browser, Cistrome, Galaxy, GenePattern, Genomica, and geWorkbench. Below the toolbar is a file browser showing a directory structure. A red arrow points from the GenePattern icon to a file named 'TF.data.gct'. A second red arrow points from the 'TF.data.gct' file to a 'Launch GenePattern' dialog box. The dialog box has a 'Drop files here' area with a green plus sign and a 'Launch' button. A third red arrow points from the 'Launch' button to the text 'Then click the Launch button'.

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			
<input type="checkbox"/> DMAP.classes.cls			
<input type="checkbox"/> TF.data.cms.HSC.ChIP.Rest.filt.txt			
<input type="checkbox"/> TF.data.gct			
<input type="checkbox"/> TF.data.genesymbol.comp.marker.attr			
<input type="checkbox"/> TF.data.genesymbol.gct			
<input type="checkbox"/> TF.data.genesymbol.geneset.tab			
<input type="checkbox"/> TF.data.tab	ted	1.31 Mb	12 months ago
<input type="checkbox"/> all_aml_test.res	ted	1.83 Mb	12 months ago
<input type="checkbox"/> all_aml_test.slice.gos.res	ted	433 bytes	5 months ago
<input type="checkbox"/> all_aml_train.comp.marker.attr	ted	2.02 Mb	12 months ago

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

©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 processed 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

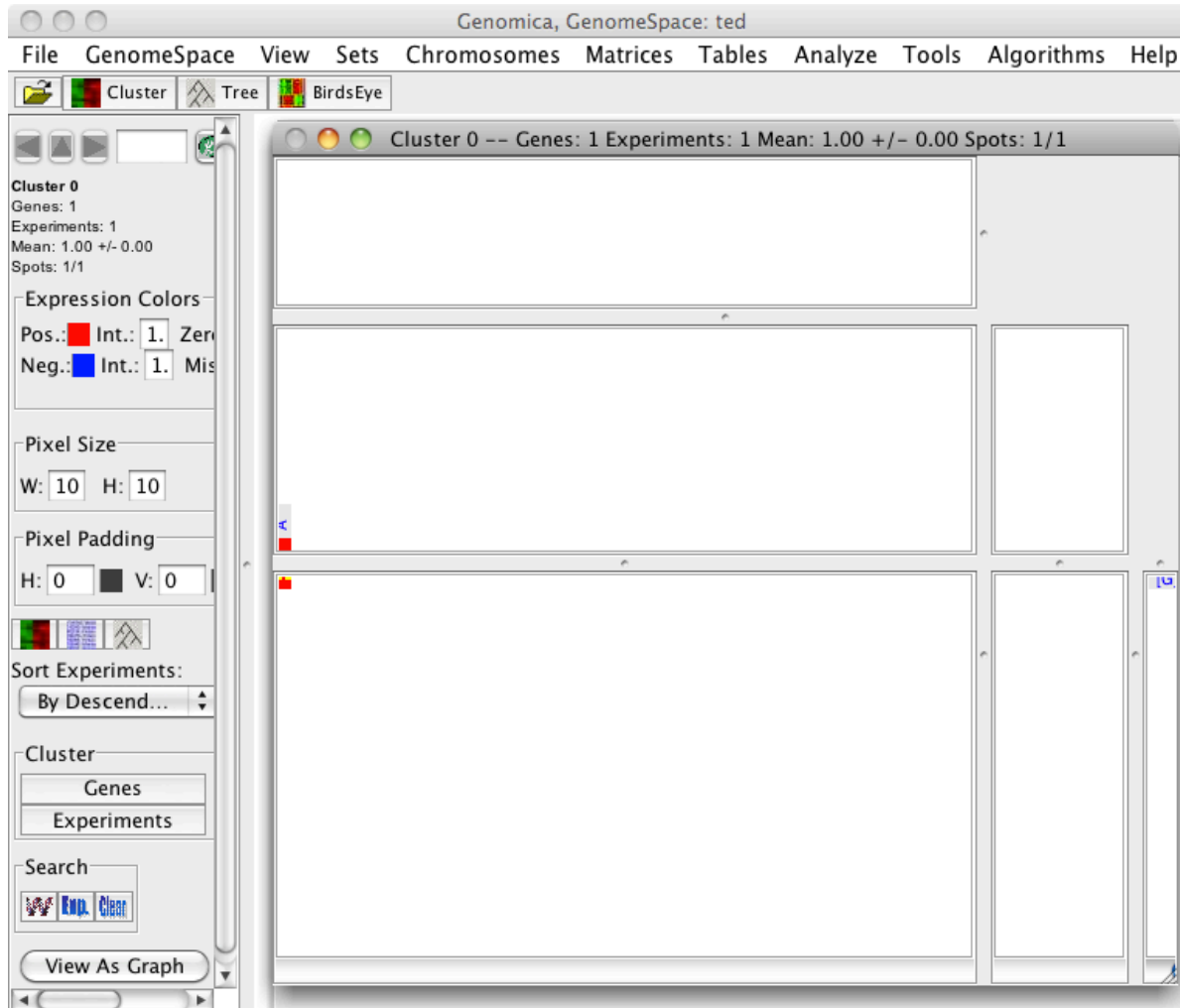
Home ▶ Public ▶ SharedData ▶ Demos ▶ Scenario ▶ step3 ▶ **80_module.gxp**

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

The image shows a screenshot of the Genomica software interface. The main window displays a heatmap with a color scale from blue to red. The top menu bar includes 'File', 'GenomeSpace', 'View', 'Sets', 'Chromosomes', 'Matrices', 'Tables', 'Analyze', 'Tools', 'Algorithms', and 'Help'. The 'GenomeSpace' menu is open, showing options such as 'Login GenomeSpace ...', 'Logout GenomeSpace', 'Open From GenomeSpace ...', 'Save To GenomeSpace', 'Export Network To GenomeSpace...', 'Launch GenomeSpace Tool...', 'Upload Local File To GenomeSpace ...', 'Download GenomeSpace File To Local File...', 'Delete From GenomeSpace ...', and 'Change GS Environment...'. The 'Save To GenomeSpace' option is highlighted. A file dialog box is open in the foreground, showing a tree view of 'GenomeSpace Files' with a subfolder 'ted'. The 'Save as:' field contains 'SavedFromGenomica.gxp'. The dialog box has 'New Folder', 'Save As', and 'Cancel' buttons.

Genomica, GenomeSpace: ted

File GenomeSpace View Sets Chromosomes Matrices Tables Analyze Tools Algorithms Help

Login GenomeSpace ...
Logout GenomeSpace
Open From GenomeSpace ...
Save To GenomeSpace
Export Network To GenomeSpace...
Launch GenomeSpace Tool...
Upload Local File To GenomeSpace ...
Download GenomeSpace File To Local File...
Delete From GenomeSpace ...
Change GS Environment...

Pixel Size
W: 10 H: 10
Pixel Padding
H: 0 V: 0
Sort Experiments:
By Descend...
Cluster
Genes
Experiments
Search
View As Graph

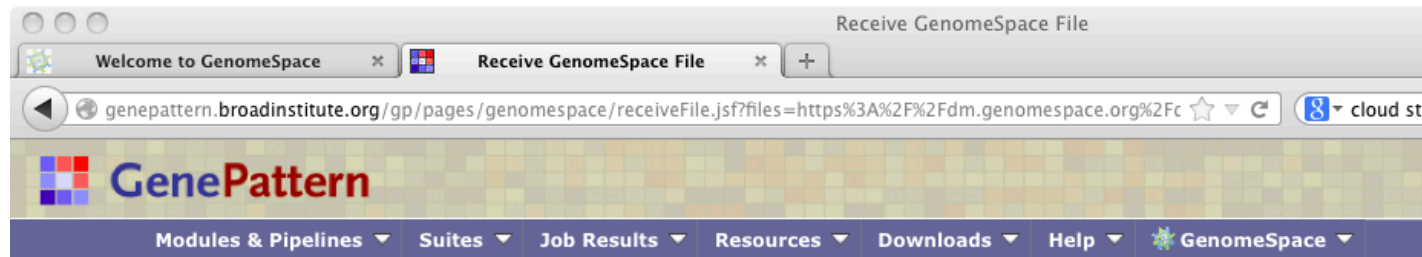
Gene [8563:0] Experiment [EOS2_5:38] Value [-0.058409]

GenomeSpace Files
ted
ArrayExpress
GP_cytoscape_demo
GSE1000GPL96_RNA_FRMAGENE_19414
GSE1000GPL96_RNA_FRMAGENE_8127
GSE10017GPL1261_RNA_FRMAGENE_17784
GSE14990GPL570_RNA_FRMAGENE_16835
NP_minidbp
demo
galaxy_export
working_dir
01speedzonecomp.pdf
CCLE_Expression_Entrez_2012-04-06.res
c5.all.v3.0.symbols.gmt
galFiltered_xomml

Save as: SavedFromGenomica.gxp
New Folder Save As Cancel

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:

Send to Module:

Process the data

- Run PreprocessDataset with default parameters

The screenshot displays the GenePattern web interface for the PreprocessDataset module. The browser address bar shows the URL: `genepattern.broadinstitute.org/gp/pages/index.jsf?sid=urn:lsid:broad.mit.edu:cancer.software.genepattern.module.an`. The page title is "GenePattern - PreprocessDataset".

The main content area is titled "PreprocessDataset" (version 4) and includes a "Run" button and a "Reset" button. The configuration parameters are as follows:

- input filename***: `https://dm.genomespace.org/datamana` (with a "Batch Parameter" checkbox)
- threshold and filter**: `yes` (with a dropdown arrow)
- 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`

The left sidebar shows a "Modules & Pipelines" search bar and a list of categories including "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 right sidebar shows a "Recent Jobs" list with the following entries:

- 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 ...

GenePattern | Job 682330

Welcome to GenomeSpace x GenePattern | Job 682330 x +

genepattern.broadinstitute.org/gp/jobResults/682330?openVisualizers=true

cloud storage dropbox

GenePattern System Message | My Settings | Sign out ted1

Modules & Pipelines Suites Job Results Resources Downloads Help GenomeSpace

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

Edit Sharing... Show Input Parameters ERR

input.filename: <https://dm.genomespace.org/genomespace.org/datamanager/dataformat/gct>

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

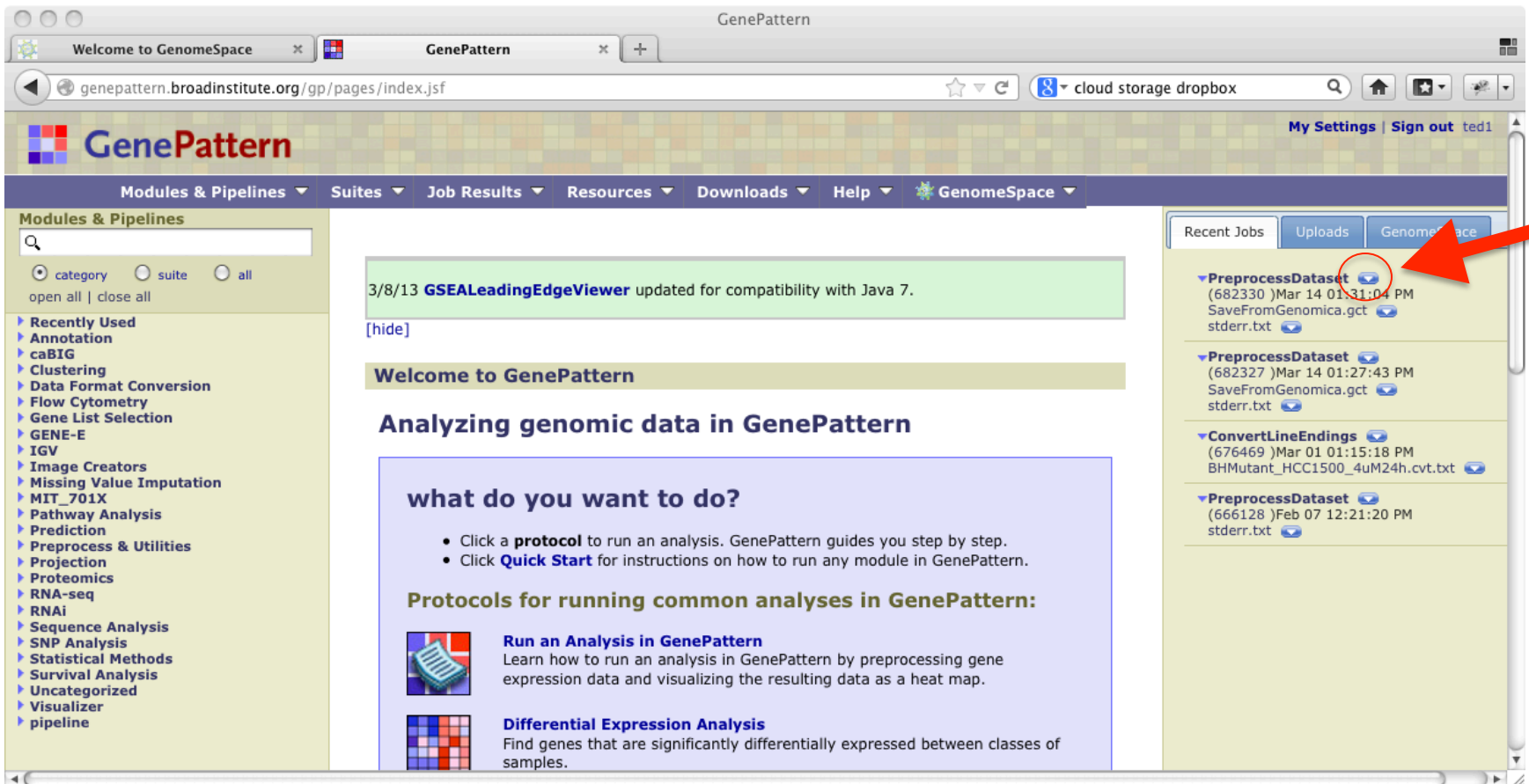
stderr.txt (1.0 KB) (Last modified: Thu Mar 14 13:30:52 EDT 2013)

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.



The screenshot shows the GenePattern web interface in a browser window. The browser tabs include "Welcome to GenomeSpace" and "GenePattern". The address bar shows "genepattern.broadinstitute.org/gp/pages/index.jsf". The page header includes "GenePattern" and "My Settings | Sign out ted1". The navigation bar contains "Modules & Pipelines", "Suites", "Job Results", "Resources", "Downloads", "Help", and "GenomeSpace".

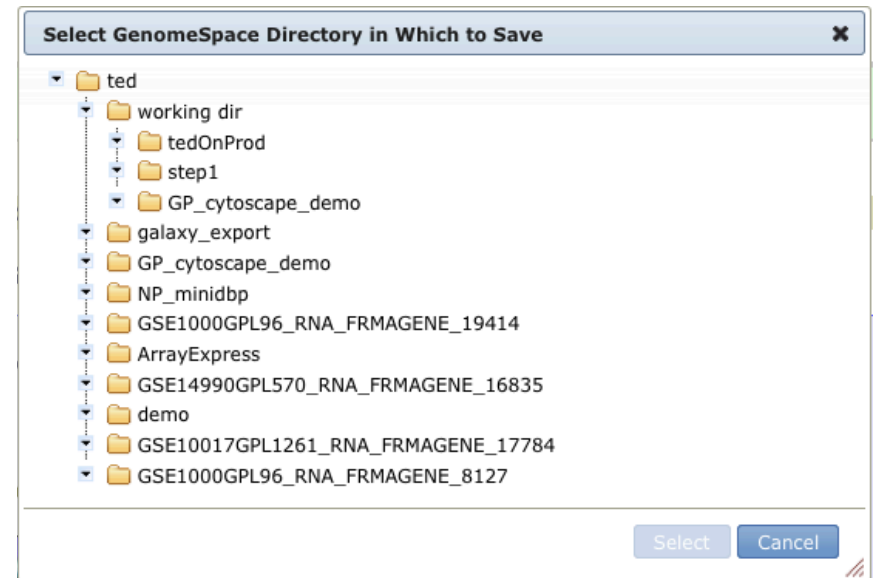
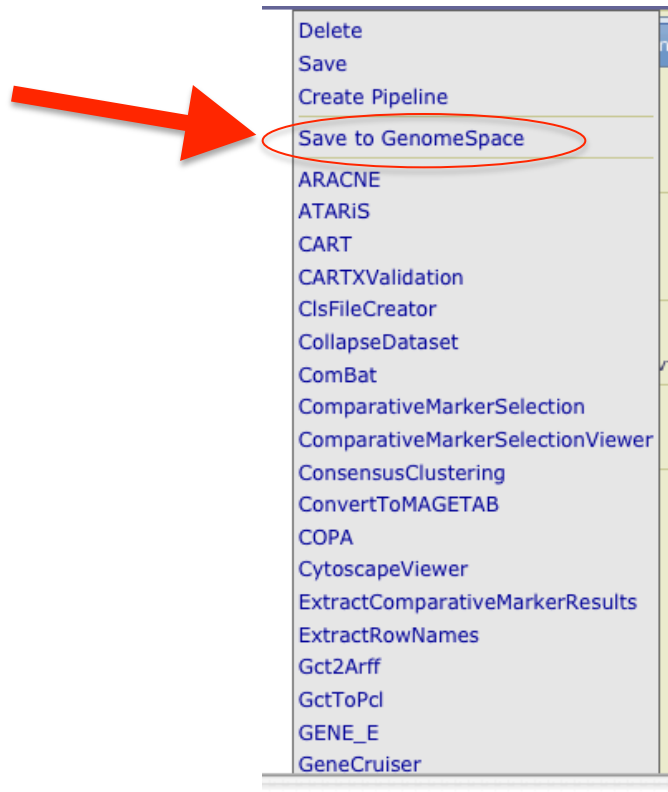
The left sidebar, titled "Modules & Pipelines", has a search box and radio buttons for "category", "suite", and "all". Below are several expandable categories, including "Recently Used", "Annotation", "caBIG", "Clustering", "Data Format Conversion", "Flow Cytometry", "Gene List Selection", "GENE-E", "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 features a green notification box: "3/8/13 GSEALeadingEdgeViewer updated for compatibility with Java 7." Below this is a "Welcome to GenePattern" banner and a section titled "Analyzing genomic data in GenePattern" with the heading "what do you want to do?". This section includes a list of instructions: "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." Below are two protocol cards: "Run an Analysis in GenePattern" (with a document icon) and "Differential Expression Analysis" (with a heatmap icon).

On the right side, there is a "Recent Jobs" panel with tabs for "Recent Jobs", "Uploads", and "GenomeSpace". A red arrow points to a context menu that is open over the first job entry, "PreprocessDataset (682330) Mar 14 01:31:04 PM". The context menu options are "SaveFromGenomica.gct" and "stderr.txt".

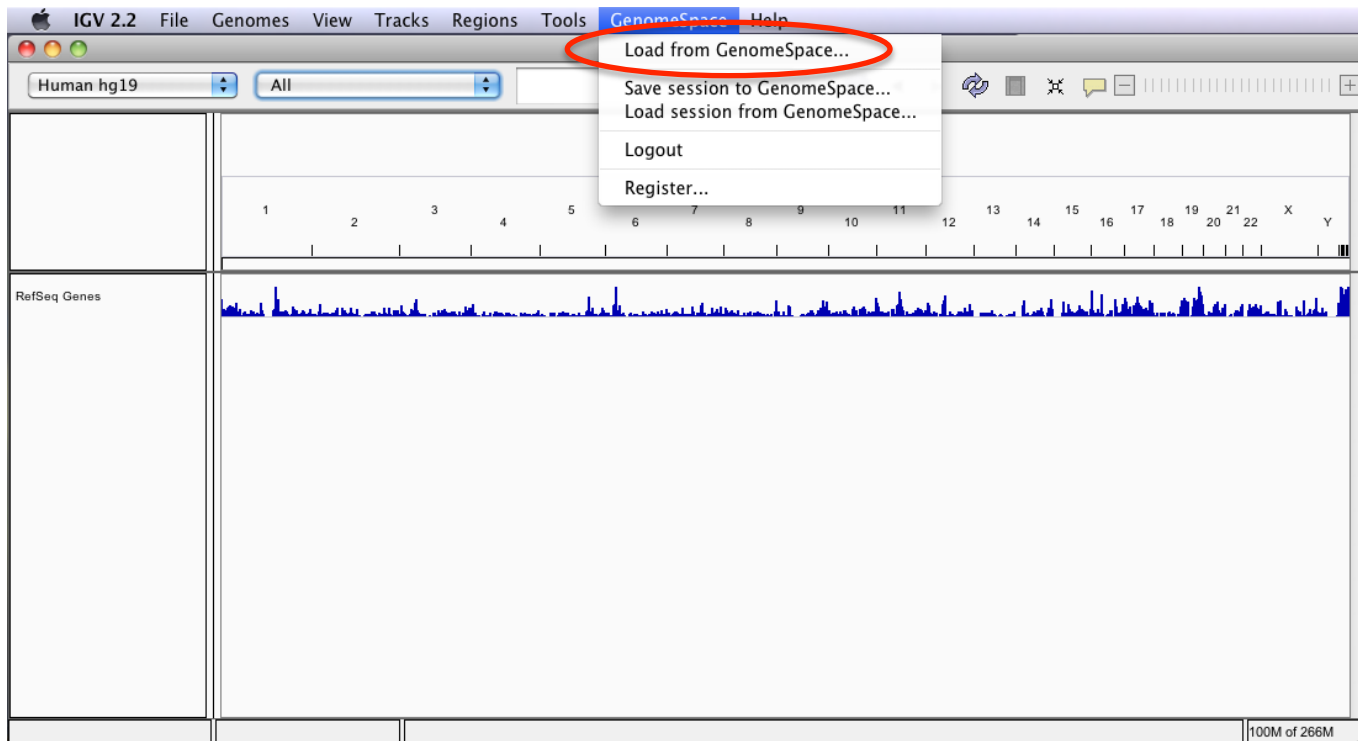
Saving to GenomeSpace

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

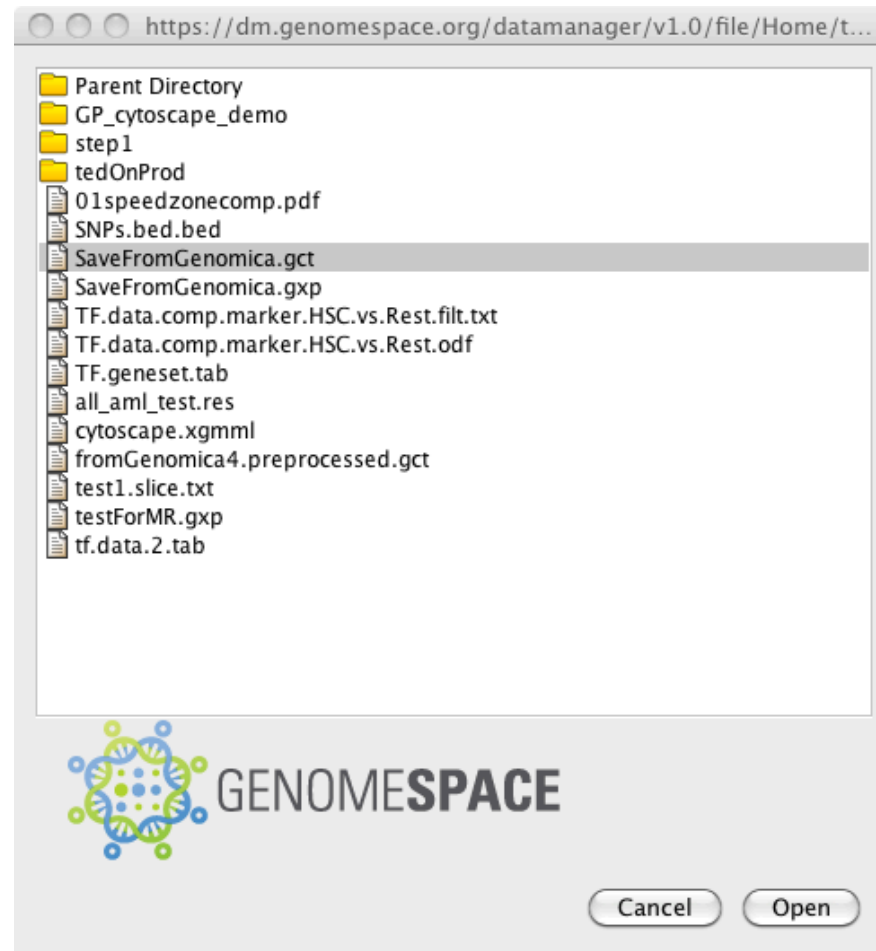


Send to IGV

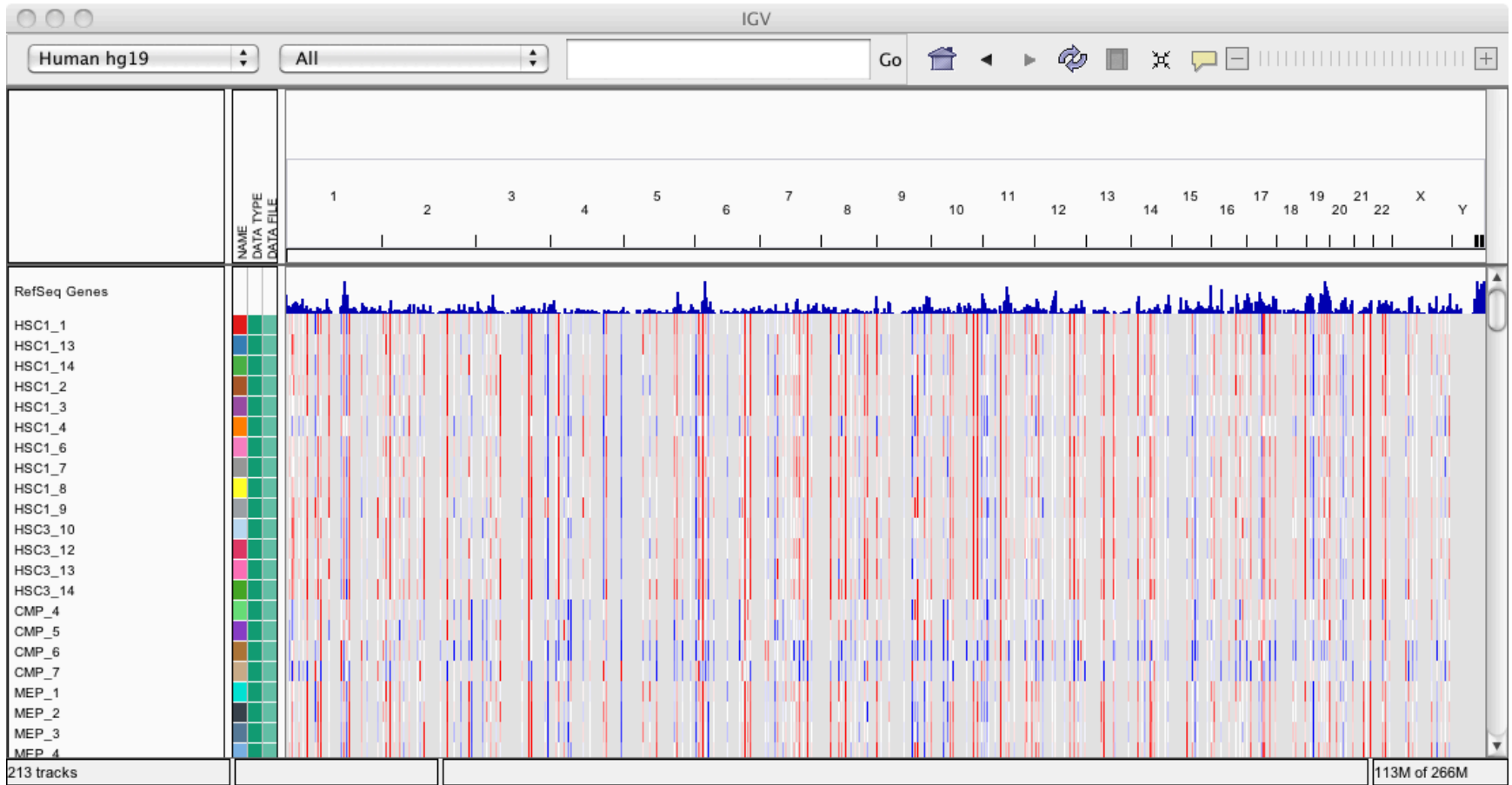
- In the GenomeSpace interface, launch IGV
 - Open the ‘GenomeSpace’ menu and ‘Load from GenomeSpace’



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

The screenshot displays the GenomeSpace web interface. On the left, a file list includes '01speedzonecomp.pdf', 'SNPs.bed.bed', 'SaveFromGenomica.gct', 'SaveF...', 'TF.data', 'TF.data', 'TF.gen', 'all_aml', 'cytosca', 'fromGe', 'test1.s', 'testFor...', and 'tf.data.2.tab'. A context menu is open over 'SaveFromGenomica.gct', with 'Preview' selected. The main window shows the 'Welcome to GenomeSpace' page with a navigation bar and tool icons. A 'File Preview' window is open, displaying the first 5000 bytes of 'SaveFromGenomica.gct'. The preview content is as follows:

```
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
```

Below the preview window, a file list shows:

File Name	Owner	Size	Time
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

Extracting Rows and/or Columns

The screenshot shows the GenomeSpace web interface. The browser address bar displays <https://gsui.genomespace.org/jsui/gsui.html#>. The page header includes the "GENOMESPACE BETA" logo and a user profile icon for "ted". A navigation menu contains "File", "Launch", "View", "Manage", "Recipes", and "Help".

The "File" menu is open, showing options: Upload, Create Subdirectory, Preview, **Extract rows and columns**, Convert, Download, View Link URL, Sharing, Rename, Move, and Delete. The "Extract rows and columns" option is highlighted in blue.

Below the menu is a toolbar with icons for various tools: ArrayExpress, UCSC Table Browser, Cistrome, Galaxy, GenePattern, Genomica, geWorkbench, Gtools, and IGV.

The main content area displays a file list with columns for "Filename", "Owner", "Size", and "Last Modified". The current directory is "ted > workingDir". The file "SaveFromGenomica.gct" is selected and highlighted in yellow.

Filename	Owner	Size	Last Modified
Up to: Home > ted > workingDir			
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.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

©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/jsui/gsui.html#

cloud storage dropbox

GENOMESPACE 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

Toggle all columns

	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1	#1.2													
2	523	211												
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	-0.21882
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

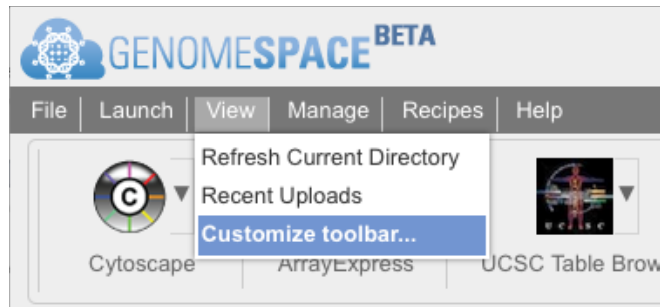
all_aml_test.res ted 1.83 Mb 6 minutes ago

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

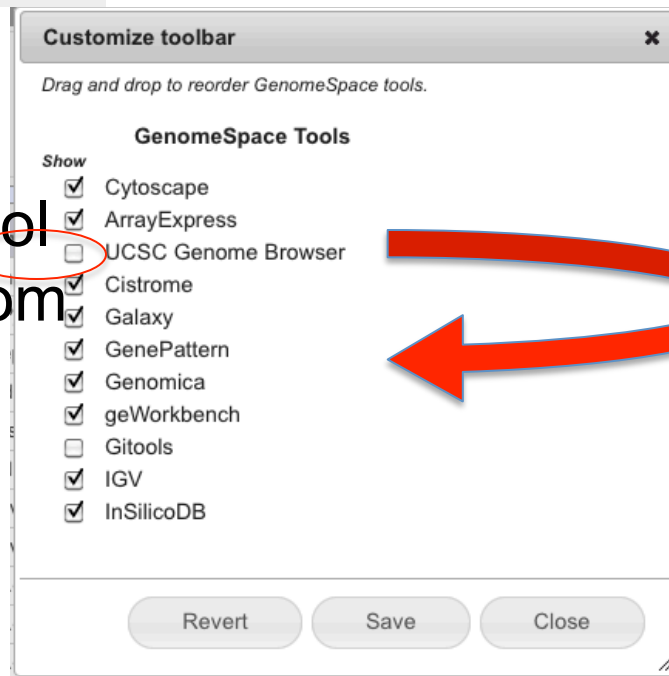
Sharing with others

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

Organizing tools



Uncheck the tool
To remove it from
The toolbar



Drag and drop
tools in the list
reorder them

Other GenomeSpace Tools



ArrayExpress



geWorkbench



Galaxy



Gitoools



Cistrome



IGV



Cytoscape



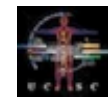
InSilicoDB



GenePattern



Genomica



UCSC Table Browser



ISAcreeator



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 website homepage. At the top, there is a browser window with the URL www.ebi.ac.uk/arrayexpress/. Below the browser window, there is a cookie notice: "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".

The main navigation bar includes "Services", "Research", "Training", "Industry", and "About us". The ArrayExpress logo is prominently displayed. Below the logo, there is a search bar with a "Search" button and an "Advanced" link. Examples of search terms are provided: "E-MEXP-31, cancer, p53, Geuvadis".

The secondary navigation bar includes "Home", "Experiments", "Arrays", "Submit", "Help", "About ArrayExpress", "Feedback", and "Login".

The main content area features the heading "ArrayExpress - functional genomics data". Below this, a paragraph states: "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."

To the right, there is a "Data Content" section with a bar chart icon, indicating "Updated today at 06:00". It lists the following statistics:

- 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

The screenshot shows a web browser window displaying the Cistrome Analysis Pipeline (AP) Module page. The browser's address bar shows the URL `cistrome.org/ap/`. The page header includes navigation tabs for `Analyze Data`, `Workflow`, `Shared Data`, `Lab`, `Visualization`, `Help`, and `User`. The main content area is titled **Cistrome Analysis Pipeline (AP) Module** and contains the following text:

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.](#)

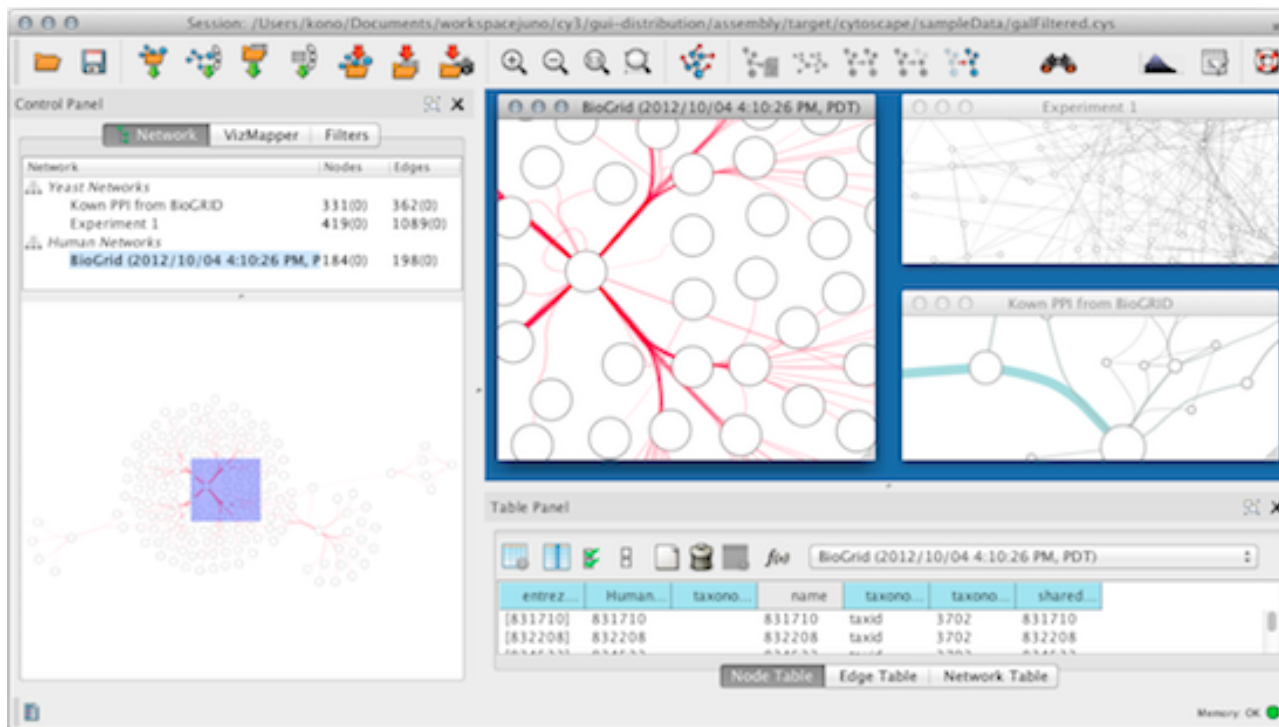
Recent News

Mar 13, 2012
Genome Space supported

The left sidebar contains a search bar and two toolboxes: **CISTROME TOOLBOX** with links for [Import Data](#), [Data Preprocessing](#), [Gene Expression](#), [Integrative Analysis](#), and [Liftover/Others](#); and **GALAXY TOOLBOX** with links for [Get Data](#), [Text Manipulation](#), [Filter and Sort](#), [Join, Subtract and Group](#), and [Convert Formats](#). The right sidebar shows a **History** panel with a single entry: **1: GenomeSpace import on TF.data.gct**, which is 1.3 MB in size.

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.

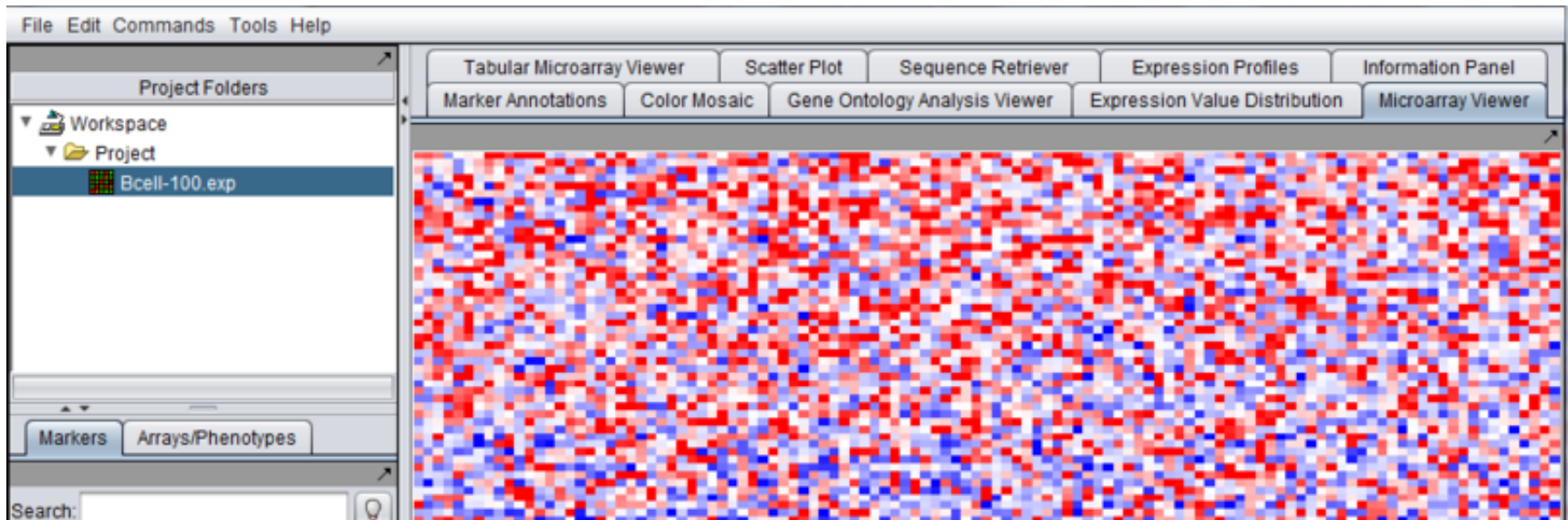
The screenshot shows the Galaxy web interface. At the top, there is a navigation bar with tabs for "Analyze Data", "Workflow", "Shared Data", "Visualization", "Cloud", "Help", and "User". A yellow banner below the navigation bar contains a warning: "Galaxy will be down on Thursday, March 14, to relocate hardware to a new server room. All jobs running at the time of the shutdown will be terminated." The main content area features a central graphic with the text "Galaxy is hiring" and "OSLO" and "UIO". Below the graphic is a "Live Quickies" section with four cards: "Completed fastQ population", "454 Mapping: Single End", "Uploading Data using FTP", and "Managing account histories". On the left, a "Tools" sidebar lists various tool categories such as "Get Data", "Send Data", "ENCODE Tools", "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". On the right, a "History" sidebar shows a list of recent jobs, including "imported: Exons vs Repeats" (416.4 MB) and several "GenomeSpace importer" jobs.



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

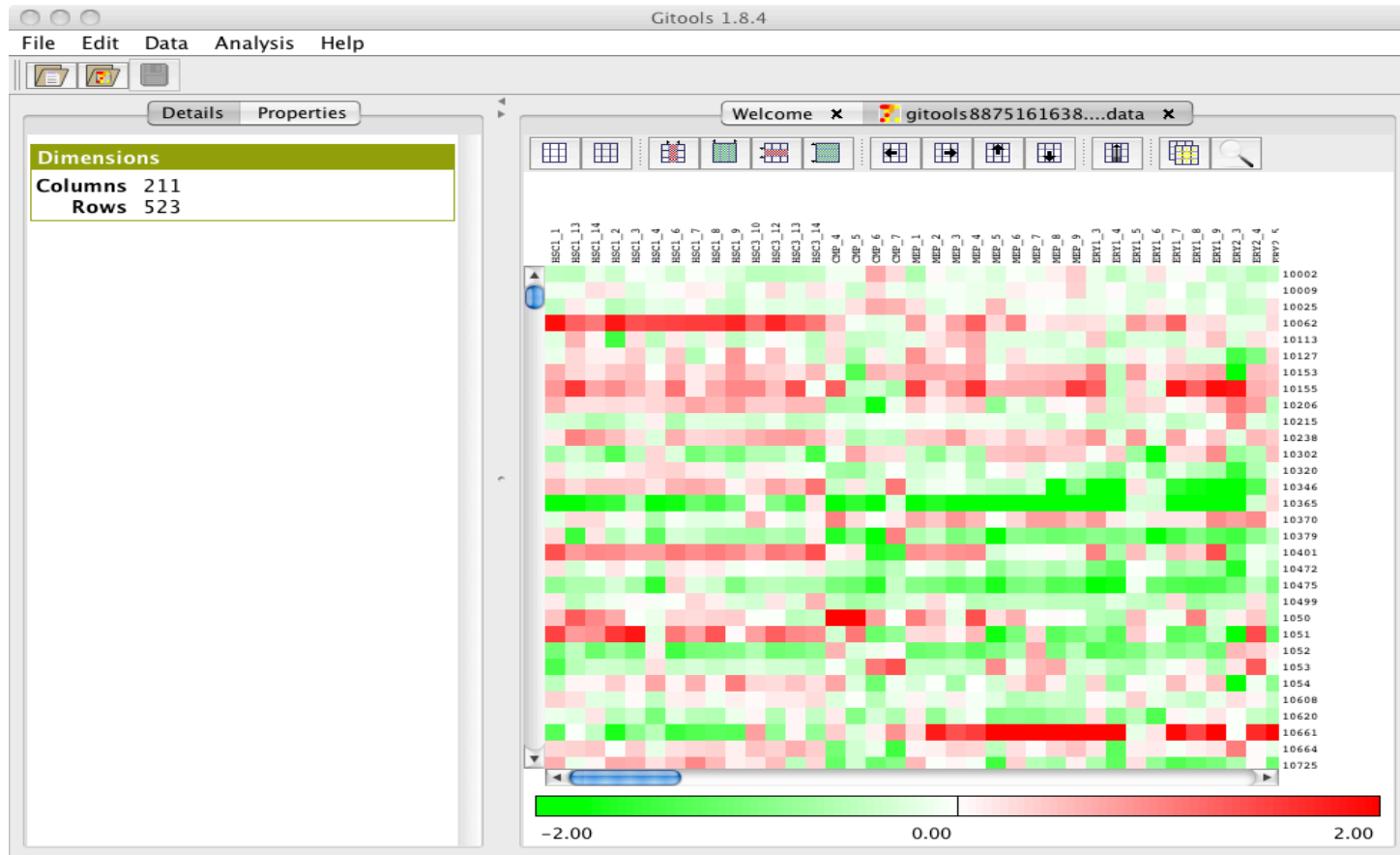




Gitools

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)





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

The screenshot displays the InSilicoDB web interface. At the top, there is a navigation bar with the 'INSILICO' logo, a search bar containing 'gse14990', and a 'Samples basket' for 'liefeld@broadinstitute.org'. Below the navigation bar, a 'Filters' sidebar on the left allows users to refine their search. The main content area shows a table of datasets, with the first entry being 'GSE14990 MYC regulation of a "poor prognosis" metastatic cancer cell state'. This entry is highlighted and includes options for 'Edit/Show clinical annotation', 'Download', and 'Export'. The table also shows the dataset's status as 'Public', its 'Relevance' as '100%', and the number of '#Samples' as '15'.

Dataset	Sharing	Relevance	#Samples	M
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

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: genome: assembly:

group: track:

table:

region: genome position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: Send output to [Galaxy](#) [GREAT](#) [GenomeSpace](#)

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

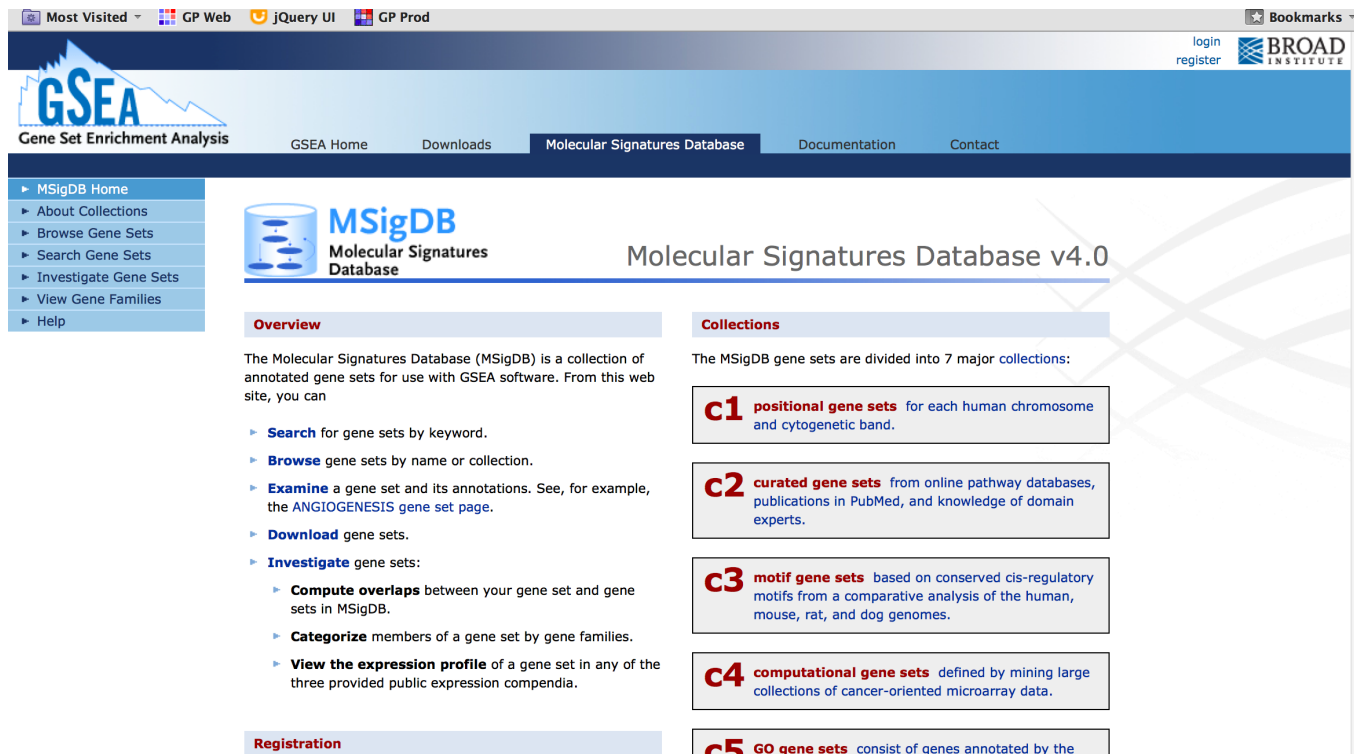
file type returned: plain text gzip compressed

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



The screenshot displays the MSigDB website interface. At the top, there is a navigation bar with links for "GSEA Home", "Downloads", "Molecular Signatures Database" (which is highlighted), "Documentation", and "Contact". A sidebar on the left contains a menu with items like "MSigDB Home", "About Collections", "Browse Gene Sets", "Search Gene Sets", "Investigate Gene Sets", "View Gene Families", and "Help". The main content area features the MSigDB logo and the title "Molecular Signatures Database v4.0". Below this, there are two main sections: "Overview" and "Collections".

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