

# GENOMESPACE



**(tools x connections) + recipes = integrative analysis**

Ted Liefeld  
Broad Institute  
July 15, 2014



**How many  
bioinformatics tools  
and databases are  
there?**



**~10,000 tools  
~ 5,000 databases**



**What could  
researchers  
accomplish if those  
tools could talk to  
each other?**

# Case study: Studying the regulatory control of human hematopoiesis

Cell

Resource

## Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis

Noa Novershtern,<sup>1,2,3,11</sup> Aravind Subramanian,<sup>1,11</sup> Lee N. Lawton,<sup>4</sup> Raymond H. Mak,<sup>1</sup> W. Nicholas Haining,<sup>5</sup> Marie E. McConkey,<sup>6</sup> Naomi Habib,<sup>3</sup> Nir Yosef,<sup>1</sup> Cindy Y. Chang,<sup>1,6</sup> Tal Shay,<sup>1</sup> Garrett M. Frampton,<sup>2,4</sup> Adam C.B. Drake,<sup>2,7</sup> Ilya Leskov,<sup>2,7</sup> Bjorn Nilsson,<sup>1,6</sup> Fred Pfeffer,<sup>8</sup> David Dombkowski,<sup>8</sup> John W. Evans,<sup>5</sup> Ted Liefeld,<sup>1</sup> John S. Smutko,<sup>9</sup> Jianzhu Chen,<sup>2,7</sup> Nir Friedman,<sup>3</sup> Richard A. Young,<sup>2,4</sup> Todd R. Golub,<sup>1,5,10</sup> Aviv Regev,<sup>1,2,10,12,\*</sup> and Benjamin L. Ebert<sup>1,5,6,12,\*</sup>

<sup>1</sup>Broad Institute, 7 Cambridge Center, Cambridge MA, 02142, USA

<sup>2</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge MA, 02140, USA

<sup>3</sup>School of Computer Science, Hebrew University, Jerusalem 91904, Israel

<sup>4</sup>Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

<sup>5</sup>Dana-Farber Cancer Institute, Boston, MA 02115, USA

<sup>6</sup>Brigham and Women's Hospital, Boston, MA 02115, USA

<sup>7</sup>Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139

<sup>8</sup>Massachusetts General Hospital, Boston, MA 02114, USA

<sup>9</sup>Nugen Technologies, San Carlos, CA 94070, USA

<sup>10</sup>Howard Hughes Medical Institute, Chevy Chase, MD 20815-6789, USA

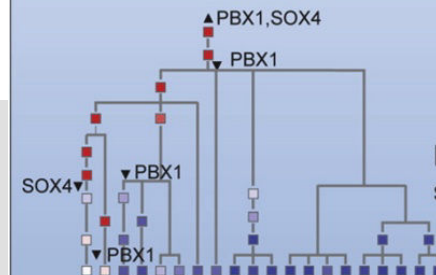
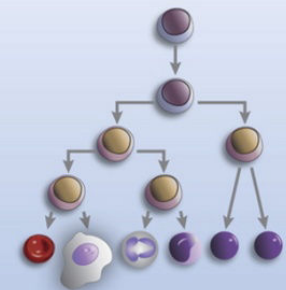
<sup>11</sup>These authors contributed equally to this work

<sup>12</sup>These authors contributed equally to this work

\*Correspondence: aregev@broad.mit.edu (A.R.), bebert@partners.org (B.L.E.)

DOI 10.1016/j.cell.2011.01.004

Gene expression profiling from 38 states of human hematopoiesis



Expression- and sequence-based models

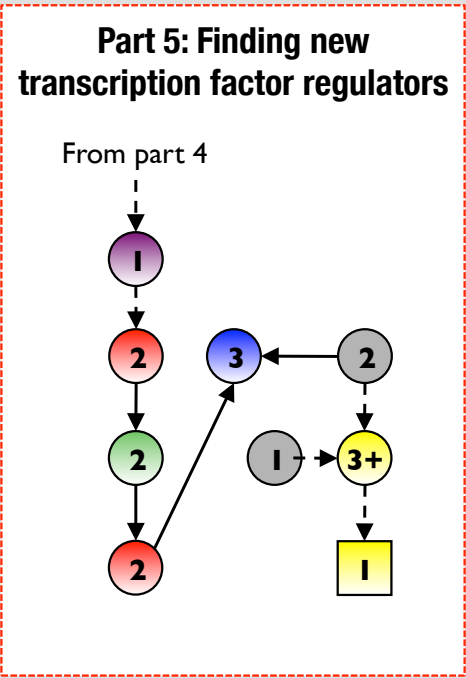
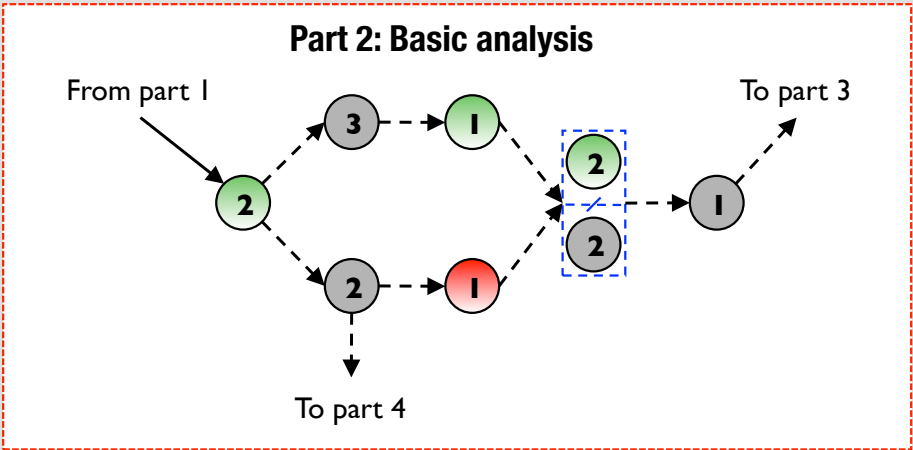
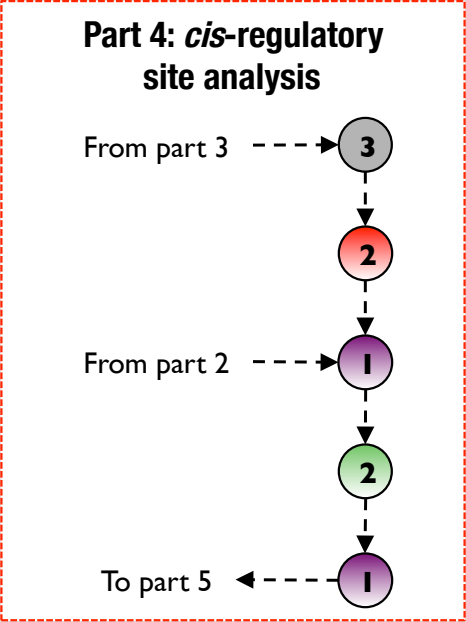
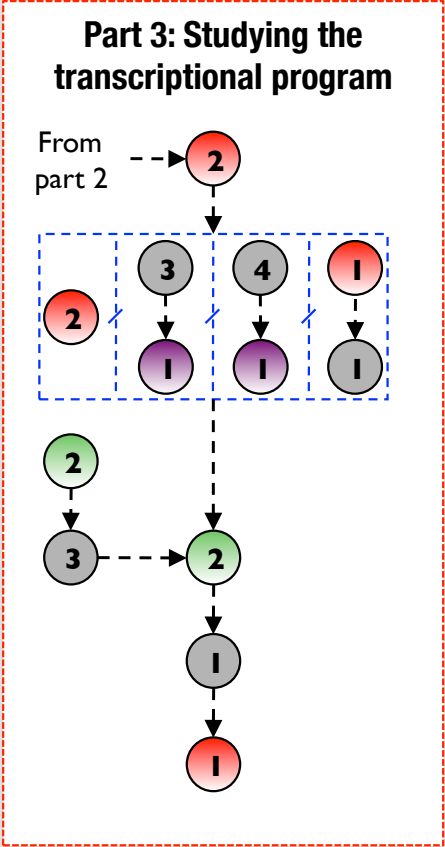
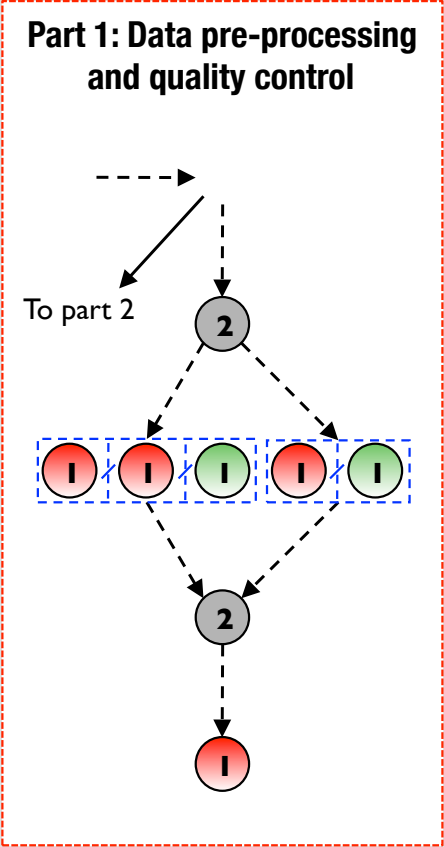
Validation with shRNA knockdown and ChIP sequencing



e.g. MEIS1, TAL1, IKAROS, PU.1, AFF1, YY1, HIF3A



# To do integrative research, you need to move data between many tools



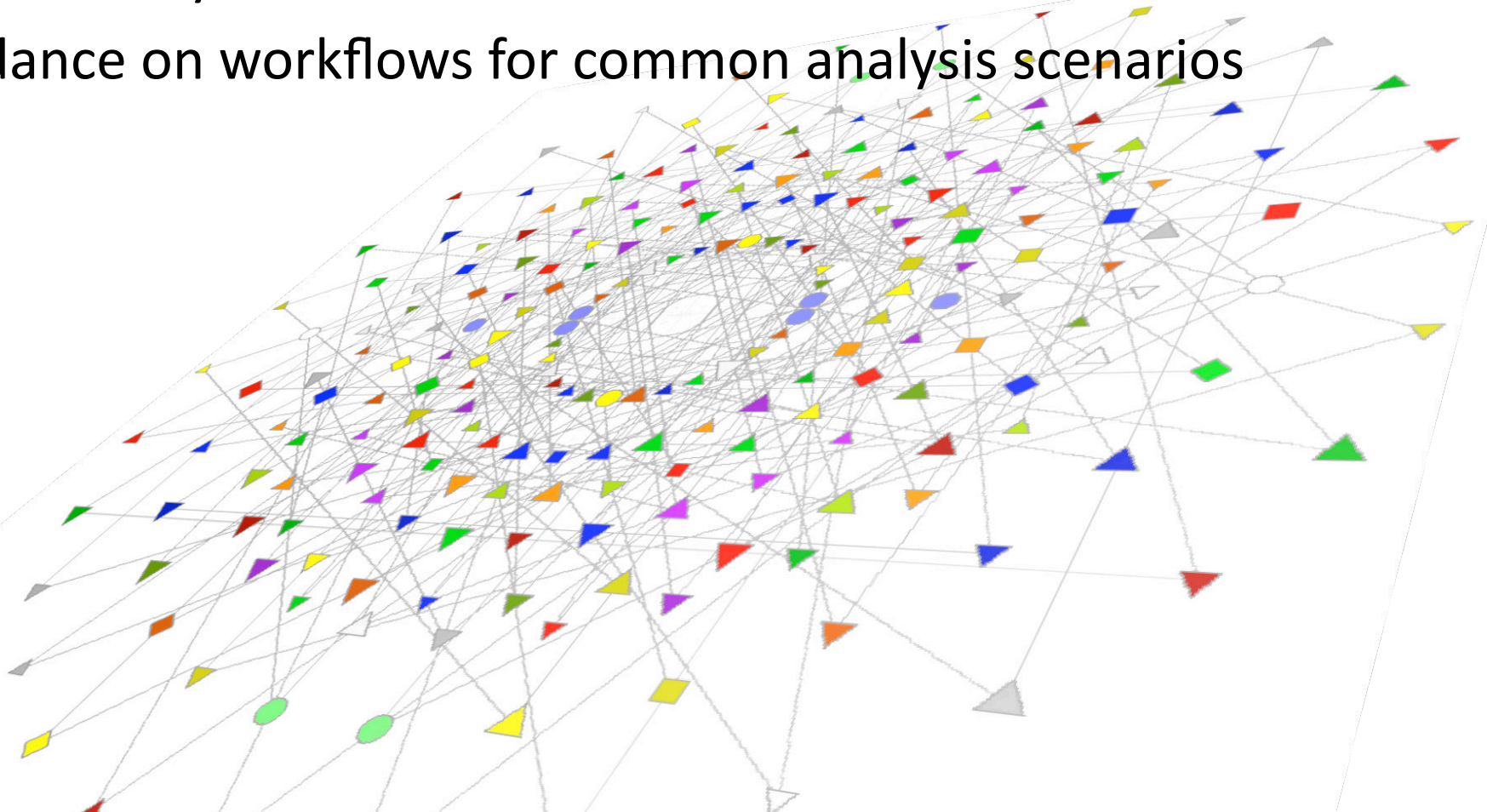
- Genomica
- GenePattern
- IGV
- Cytoscape
- Galaxy
- Manual step

---

- Analysis section
- Analysis step (# steps)
- Analysis conclusion
- Optional choices
- Currently integrated
- Not yet integrated

# We need a connection layer between 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 for new tools
- Guidance on workflows for common analysis scenarios



**WIRED**

## Twitter Founder Reveals Secret Formula for Getting Rich Online

BY RYAN TATE 09.30.13 9:30 AM

“The key to making a fortune online ...  
**is to remove extra steps from  
common activities”**

– Ev Williams

# GENOMESPACE

An online community to share diverse computational tools

The screenshot shows the GenomeSpace website homepage in a browser window. The browser title is "GenomeSpace: Home" and the address bar shows "www.genomespace.org". The page features a navigation menu with links for "What is GenomeSpace?", "Tools", "Recipes", "Documentation", "Developers", "Support", and "About". A search bar is also present. The main content area includes the "GENOMESPACE BETA" logo, the tagline "Frictionless connection of bioinformatics tools", and two prominent buttons: "Register" (blue) and "User Login" (green). Below this, there is a "STATUS" section indicating "All systems are operating normally." and a "WHAT'S NEW" section with tabs for "News Highlights" and "GenomeSpace Blog". The "GenomeSpace Blog" section contains two news items: "GenomeSpace at ISMB 2013 Berlin" and "GenomeSpace at BOSC-Broad Interoperability Hackathon". On the right side, there is a social media feed for "@genomespace" with mentions from the GenomeSpace Team and the FGED Society.

## Driving Biological Projects

lincRNAs  
Cancer stem cells  
Patient Stratification

## Seed Tools

Cytoscape  
Galaxy  
GenePattern  
Genomica  
IGV  
UCSC Browser

## Analysis Recipes

Integrative analysis  
Data transfer  
etc.

# GenomeSpace Architecture

## Server

### Amazon EC2

Identity Service (Open ID)

Analysis Task Manager (ATM)  
REST

Data Manager (DM)  
REST



HTTP

HTTP

File Conversion

File Transfers

	HTTP		
Protocols	REST		
	CDK		
Tools	Cytoscape GenePattern Genomica etc.	GenomeSpace UI IGV Galaxy etc.	



# GenomeSpace UI

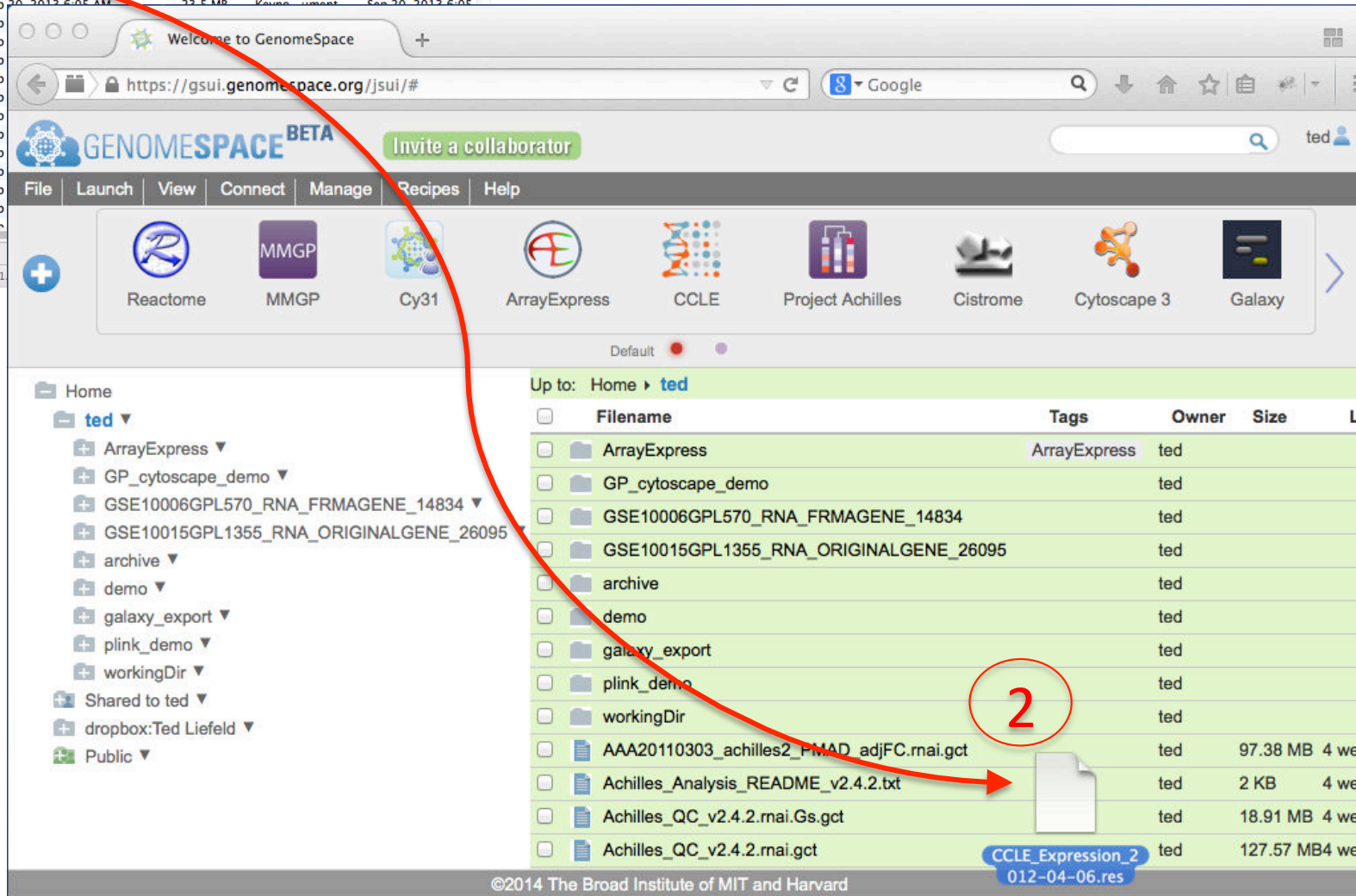
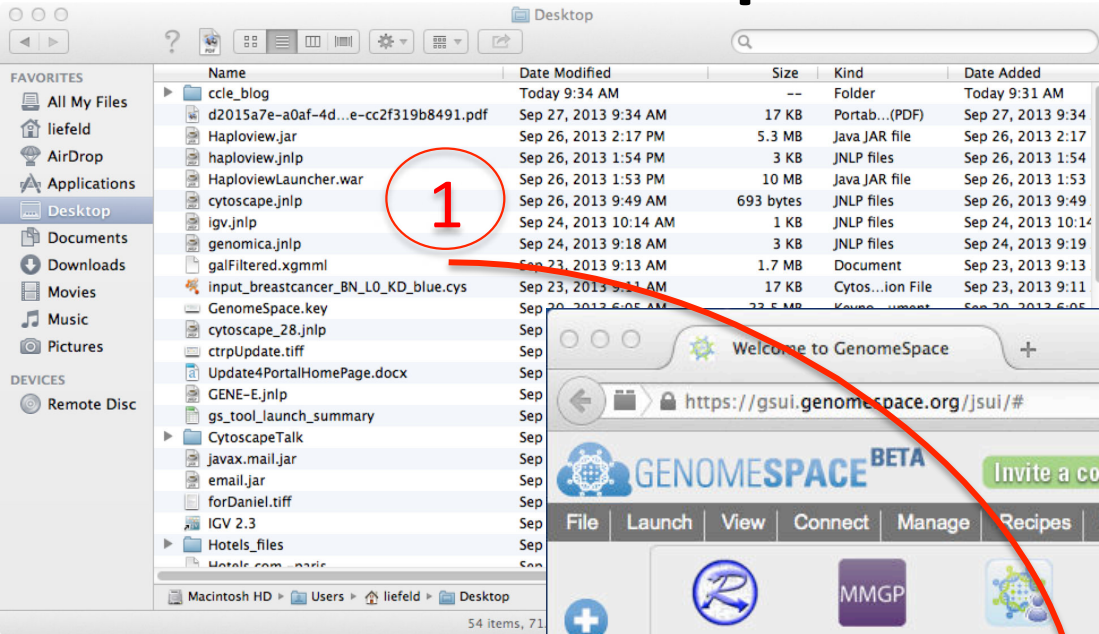
The screenshot displays the GenomeSpace user interface. At the top, a browser window shows the URL <https://gsui.genomespace.org/jsui/#>. The main header features the "GENOMESPACE BETA" logo, an "Invite a collaborator" button, and a search bar with the user name "ted". A navigation menu includes "File", "Launch", "View", "Connect", "Manage", "Recipes", and "Help". Below this is a row of application icons: Reactome, MMGP, Cy31, ArrayExpress, CCLE, Project Achilles, Cistrome, Cytoscape 3, and Galaxy. The main content area shows a file browser for the user "ted" in the "Home" directory. A sidebar on the left lists sub-directories like "ArrayExpress", "GP\_cytoscape\_demo", and "archive". The main pane displays a table of files and folders.

Filename	Tags	Owner	Size	La
ArrayExpress	ArrayExpress	ted		
GP_cytoscape_demo		ted		
GSE10006GPL570_RNA_FRMAGENE_14834		ted		
GSE10015GPL1355_RNA_ORIGINALGENE_26095		ted		
archive		ted		
demo		ted		
galaxy_export		ted		
plink_demo		ted		
workingDir		ted		
AAA20110303_achilles2_PMad_adjFC.rnai.gct		ted	97.38 MB	4 week
Achilles_Analysis_README_v2.4.2.txt		ted	2 KB	4 week
Achilles_QC_v2.4.2.rnai.Gs.gct		ted	18.91 MB	4 week
Achilles_QC_v2.4.2.rnai.gct		ted	127.57 MB	4 week

©2014 The Broad Institute of MIT and Harvard



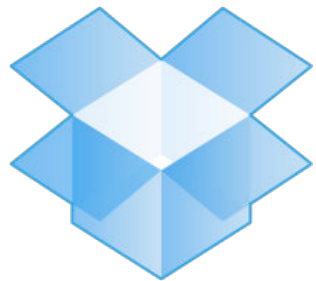
# Uploading Data



# Connection to cloud-based storage systems

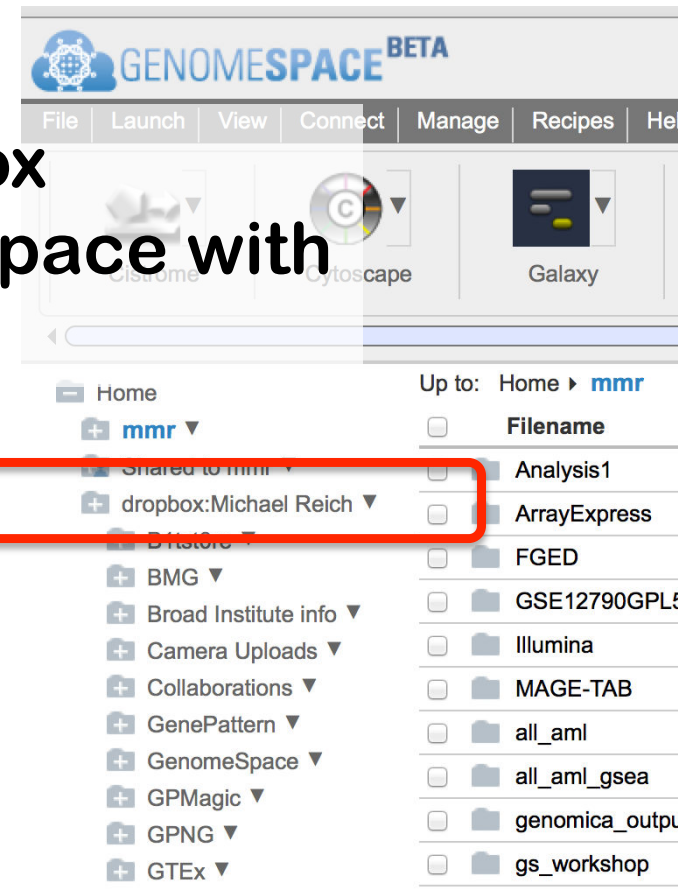


Use GenomeSpace S3 storage  
or add your own Amazon  
account



**Dropbox**

Connect your Dropbox  
account to GenomeSpace with  
one click

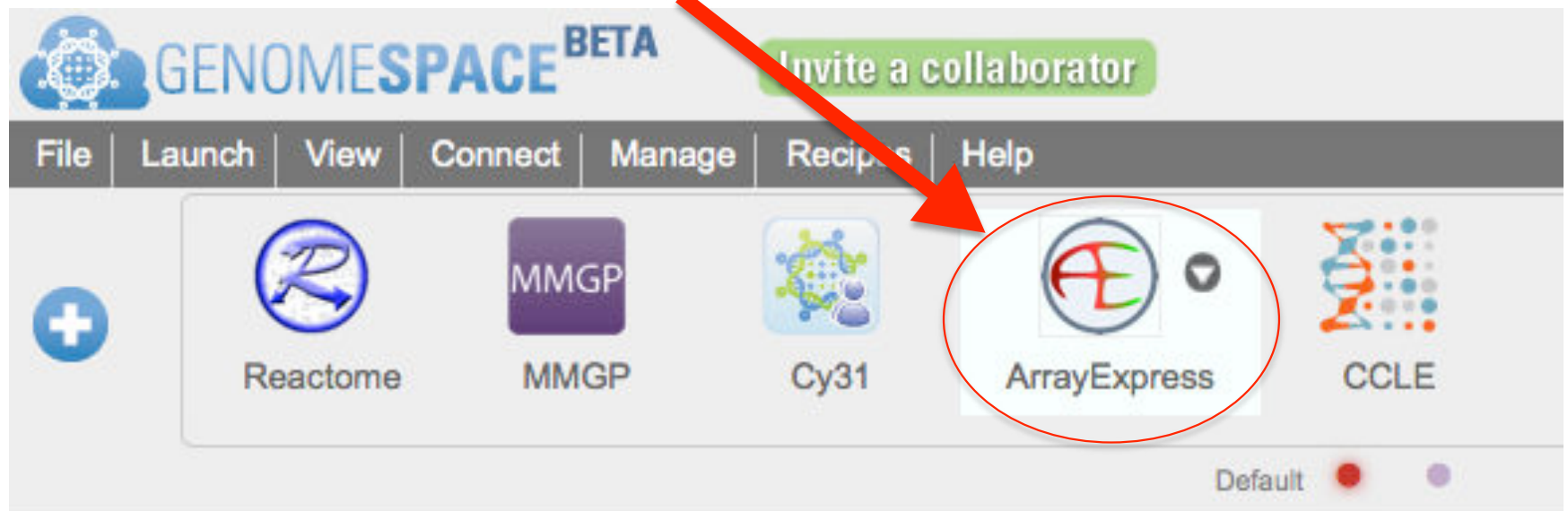


in development



# Launching tools

Click on the tool's icon



Home

ted

ArrayExpress

GP\_cytoscape\_demo

GSE10006GPL570\_RNA\_FRMAGENE\_14834

Up to: Home > ted

Filename

ArrayExpress

GP\_cytoscape\_demo

GSE10006GPL570\_RN

# File Preview

- 01speedzonecomp.pdf
- SNPs.bed.bed
- SaveFromGenomica.gct
- SaveFr
- TF.data est.filt.tx
- TF.data est.odf
- TF.gen
- all\_am
- cytosca
- fromGe
- test1.s
- testFor
- tf.data.2.tab

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

Welcome to GenomeSpace

https://gsui.genomospace.org/jsui/#

GENOMESPACE BETA Invite a collaborator

File Launch View Connect Manage Recipes Help

Normals\_Leu.gct

Preview Heatmap

Rows  
Feature Name

	HSC-2	HSC-3	HSC-4	HSC-6	HSC-5	CMP-2	CMP-3	CMP-4	CMP-1	GMP-2
1415670_at	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
1415671_at	Red	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
1415672_at	Red	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
1415673_at	Red	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue

Close

Public

- Achilles\_Analysis\_README\_v2.4.2.txt
- Achilles\_QC\_v2.4.2.mai.Gs.gct
- Achilles\_QC\_v2.4.2.mai.gct



# Extracting Rows and/or Columns

The screenshot shows the GenomeSpace web interface. The browser address bar displays <https://gsui.genomespace.org/jsui/#>. The page header includes the 'GENOMESPACE BETA' logo, an 'Invite a collaborator' button, and a search bar with the user name 'ted'. A navigation menu contains 'File', 'Launch', 'View', 'Connect', 'Manage', 'Recipes', and 'Help'. The 'File' menu is open, showing options such as 'Upload', 'Import from URL', 'Create Subdirectory', 'Preview', 'Extract rows and columns' (highlighted), 'Convert', 'Download', 'View Link URL', 'Sharing', 'Tags', 'Create Archive', 'Expand Archive', 'Rename', 'Copy/Move', and 'Delete'. Below the menu, a toolbar features icons for various tools: MMGP, Cy31, ArrayExpress, CCLE, Project Achilles, Cistrome, Cytoscape 3, and Galaxy. The main content area displays a file list with columns for file names, owners, sizes, and dates. The file 'Normals\_Leu.gct' is selected and highlighted in yellow.

File Name	Owner	Size	Date
GP_cytoscape_demo	ted		
GSE10006GPL570_RNA_FRMAGENE_14834	ted		
GSE10015GPL1355_RNA_ORIGINALGENE_26095	ted		
archive	ted		
demo	ted		
galaxy_export	ted		
plink_demo	ted		
workingDir	ted		
AAA20110303_achilles2_PMad_adjFC.mai.gct	ted	97.38 MB	4 week
Achilles_Analysis_README_v2.4.2.txt	ted	2 KB	4 week
Achilles_QC_v2.4.2.mai.Gs.gct	ted	18.91 MB	4 week
Achilles_QC_v2.4.2.mai.gct	ted	127.57 MB	4 week
E-MTAB-2205.raw.1.zip	ted	134.11 MB	4 month
<b>Normals_Leu.gct</b>	ted	<b>4.19 MB</b>	<b>4 month</b>
Paper_sampleannotation.txt	ted	13 KB	4 week

©2014 The Broad Institute of MIT and Harvard

# 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

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"/>	
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	-0.49470805
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	0.25927109
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	-0.21882417
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	1.7195824
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	0.03669621
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	0.47363602

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



- Sharing links

- With other GenomeSpace users

- To people without GenomeSpace accounts

Name	Read	Write
mocana	Read	Write & Delete
GS-Developers	Read	Not writeable

**Share with User**  
User name:   
 Read  
 Read, Write & Delete

**Share with Group**  
showing  groups  
Group name:    
 Read  
 Read, Write & Delete

**Share with Public**  
 Allow public access  
 Read  
 Read, Write & Delete

**Grant Permissions** **Close**

# GenomeSpace-enabled tools, 7/2012



Galaxy



IGV



Cytoscape 2.8



UCSC Table Browser



GenePattern



Genomica



InSilicoDB



Cistrome

# GenomeSpace-enabled tools, 7/2014



Galaxy



Cytoscape 2.8



GenePattern



ArrayExpress



ISACreator



Synapse



GiTools



Cytoscape 3



MMGP



CCLE



IGV



UCSC Table Browser



Genomica



InSilicoDB



Cistrome



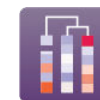
GeWorkbench



MSigDB



Reactome



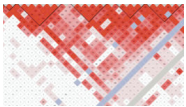
Project Achilles

# GenomeSpace-enabled tools, 7/2014

## In development



**cBio Portal** (*Memorial Sloan Kettering Cancer Center*)



**PLINK & HaploView** (*Broad Institute/MGH*)



Synapse



GiTools



Cytoscape 3



MMGP



CCLE



GeWorkbench



MSigDB



Reactome



Project Achilles

# Or add your own tools

GENOMESPACE BETA

File Launch View Connect Manage

Cytoscape ArrayExpress

Home

- ted
- Shared to ted
- dropbox: Ted Liefeld
- Public
- s3:1000genomes

### Edit HaploViewPlink

Name:

Description:

Tool Provider:

Base URL:

Help URL:

File Parameter: Parameter name:  Required:

Allow multiple files:  Multiple file Delimiter:   
example call with dummy files  
`http://vcancerportal05.broadinstitute.org:3003/HaploviewLauncher/haploview.jsp?plinkFiles=https%3A//gs.org/File1.txt,https%3A//gs.org/File2.txt`

Available formats	Formats this tool accepts
mitab	
mut	
ndb	
nnf	
odf	
owl	

[Request a new format be added to the available formats list.](#)

Icon:   haploview.tiff

Share with groups: GS-Developers ✖

Search

bench Gtools HaploV

owner	Size	Last Modified
stem		
stem		
stem		

# Automatic data transformation



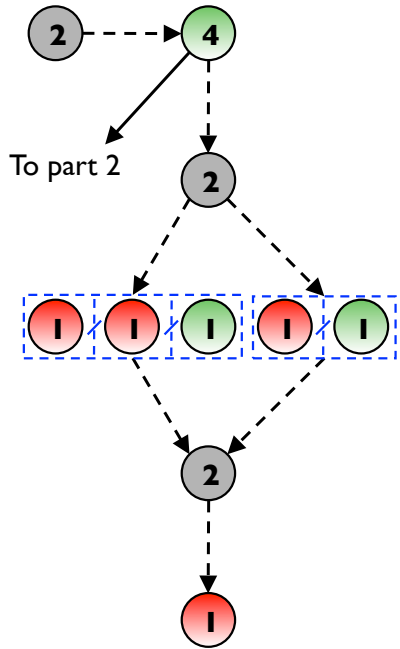
From	To	Description
gct	Cytoscape ATTR	Converts GenePattern gct file to Cytoscape attribute format
Cytoscape ATTR	Cytoscape GeneMania attr	Converts attr to GeneMania attr
res	genomicatab	Converts res to Genomica tab format
adj	xgmml	Converts adjacency file to XGMML
gmt	genomicatab	Converts from gmt to Genomica tab using Perl script
gct	gxp	Converts gct stream to gxp format
res	gxp	Converts res stream to gxp format
genomicatab	gct	Converts Genomica Tab format to gct format
odf	Cytoscape ATTR	Converts GenePattern CMS results ODF file to Cytoscape attribute format
gxp	gct	Converts gxp stream to gct format
gct	genomicatab	Converts gct to Genomica tab format
gct	geneset.tab	Converts gct to Genomica geneset.tab
gct	geWorkbench exp	Converts gct file to exp file.
res	geneset.tab	Converts res to Genomica geneset.tab

**Data transformations added upon request  
and by user contribution**

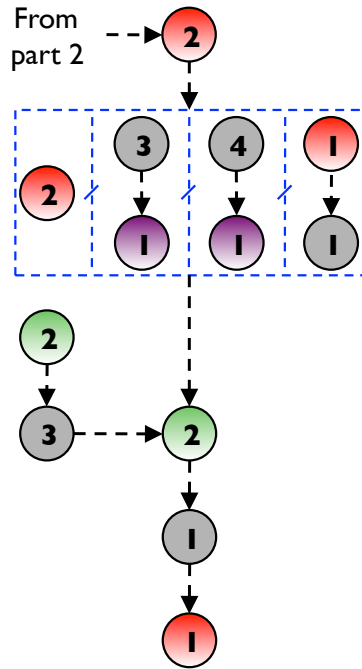


# Studying the regulatory control of human hematopoiesis Using GenomeSpace

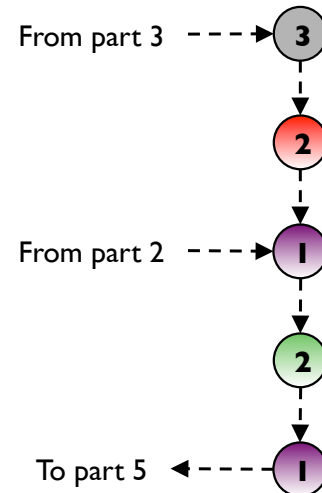
## Part 1: Data pre-processing and quality control



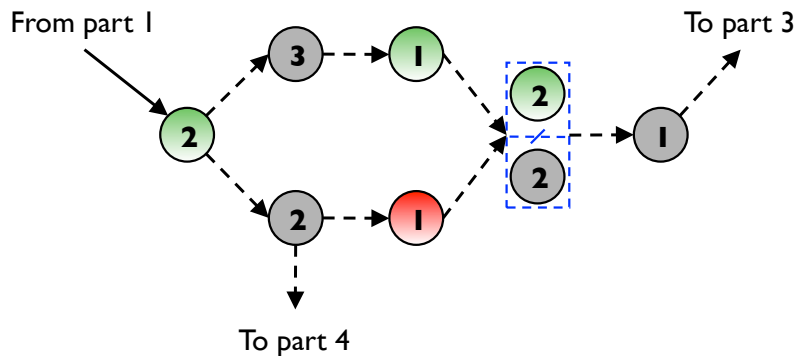
## Part 3: Studying the transcriptional program



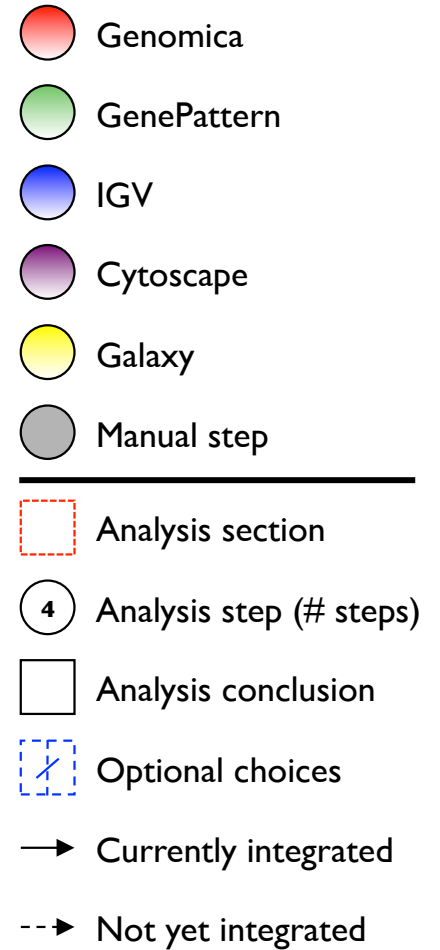
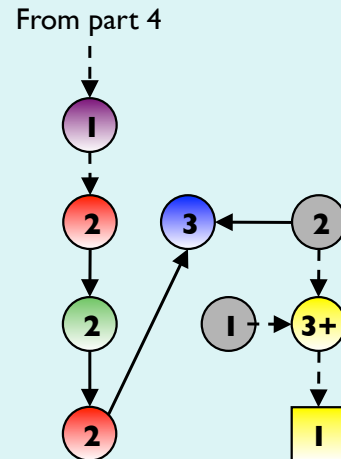
## Part 4: cis-regulatory site analysis



## Part 2: Basic analysis

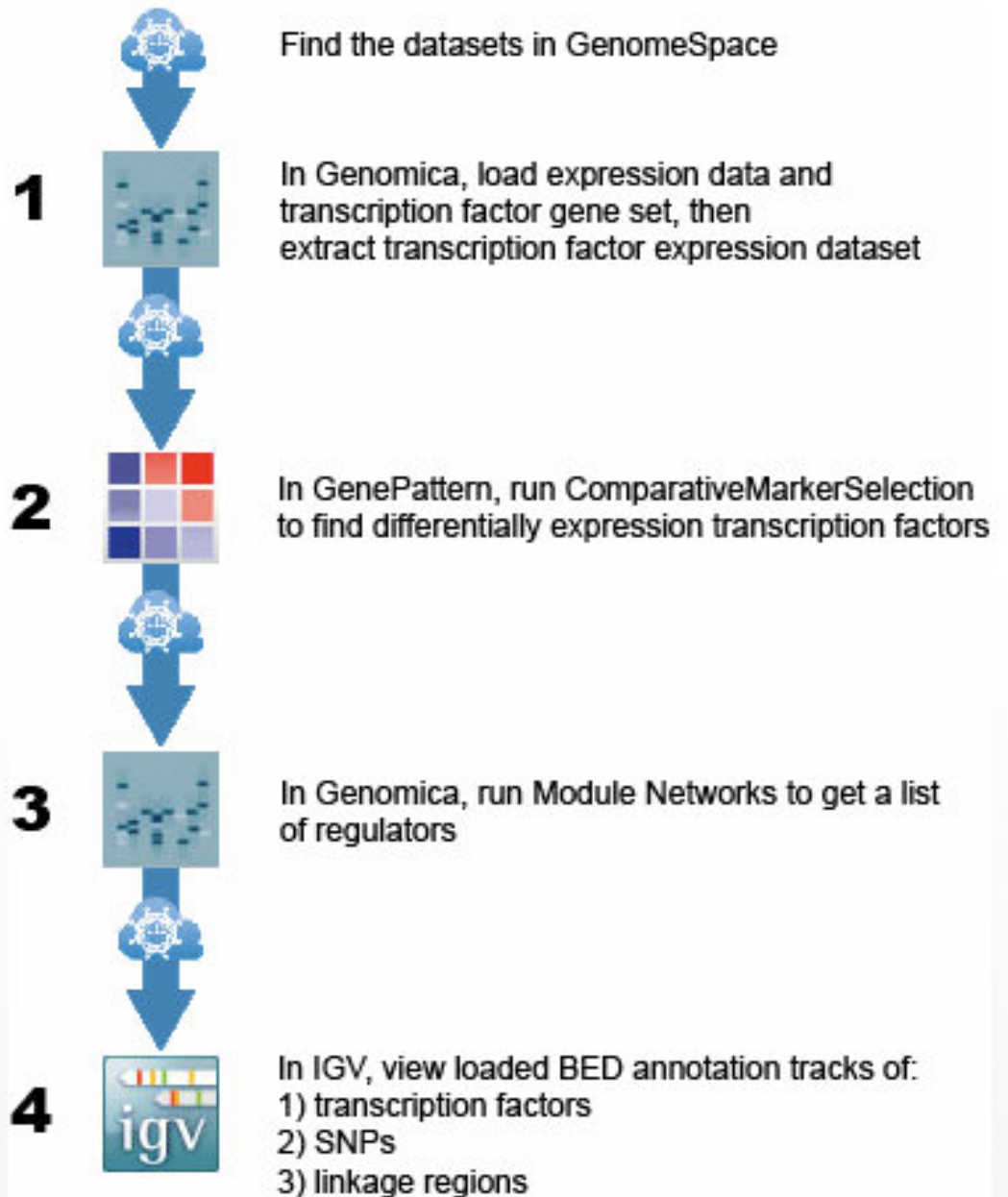
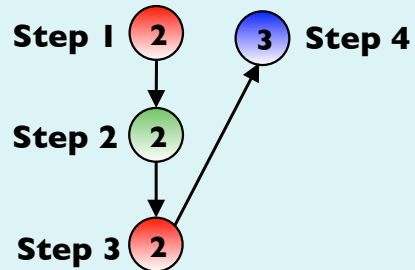


## Part 5: Finding new transcription factor regulators

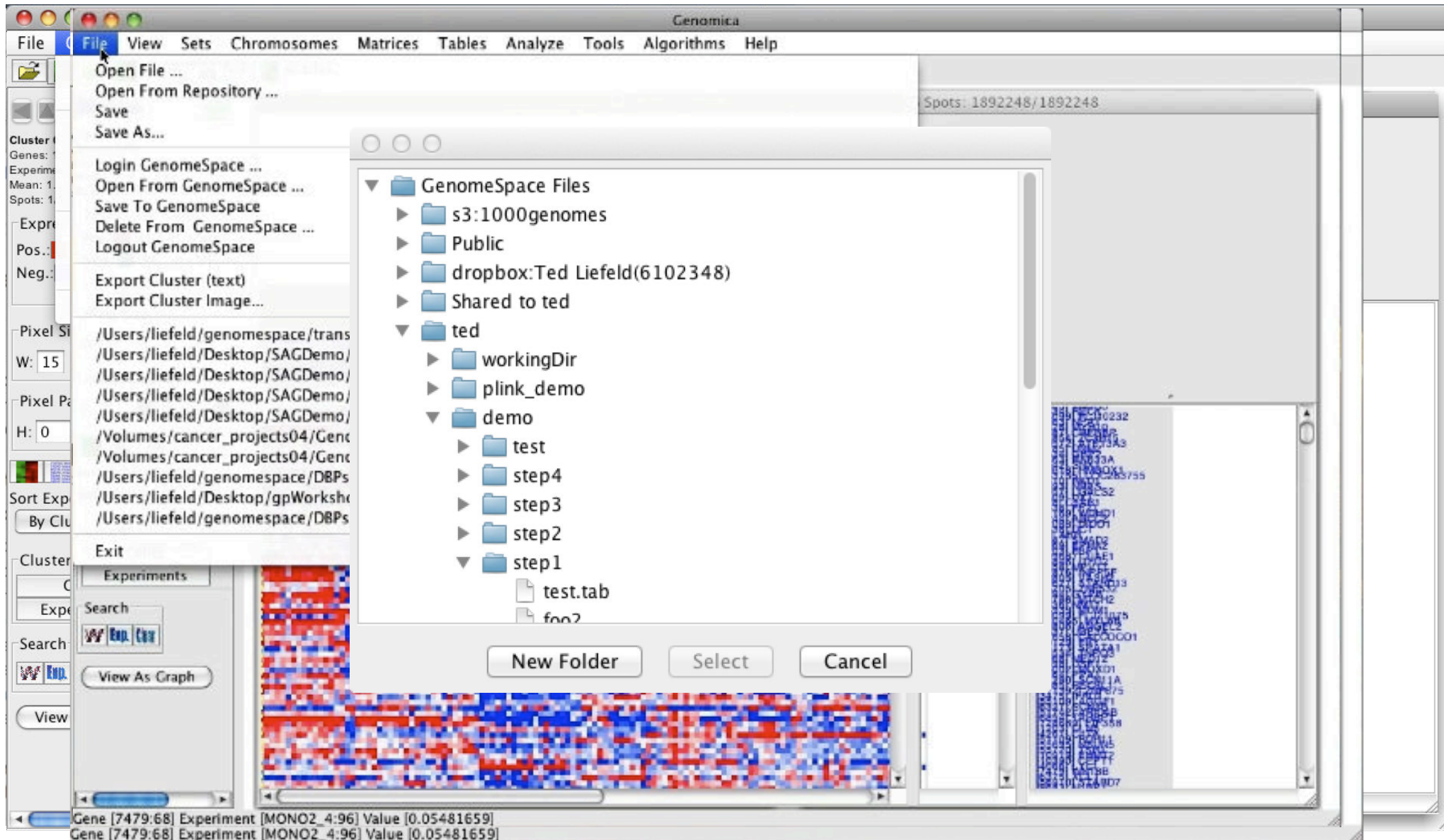


## Part 5: Finding new transcription factor regulators

From part 4



# Step 1: Create transcription factor dataset in Genomica and save to GenomeSpace



## Step 2: Send transcription factor datasets into GenePattern

Receive GenomeSpace File

genepattern.broadinstitute.org/gp/pages/genomespace/receiveFile.jsf?files=h

GenePattern [System Message](#) | [My Settings](#) | [Sign out](#) ted1


[Modules & Pipelines](#) [Suites](#) [Job Results](#) [Resources](#) [Downloads](#) [Help](#) [GenomeSpace](#)

### GenomeSpace has sent you a file

**DMP.classess.cls**  
Save a Copy:

**TF.data.tab** *Converted to gct*  
Save a Copy:

Send to Module:

[About GenePattern](#) | [Contact Us](#) ©2003-2014 Broad Institute, MIT 

# Step 2: Perform differential expression analysis in GenePattern

The screenshot shows the GenePattern web interface in a browser window. The address bar displays the URL: `genepattern.broadinstitute.org/gp/pages/index.jsf?lsid=urn:lsid:broad.mit.edu:cancer.software.genepattern`. The page title is "GenePattern" and the user is logged in as "ted1". The navigation menu includes "Modules & Pipelines", "Suites", "Job Results", "Resources", "Downloads", "Help", and "GenomeSpace".

The main content area is titled "ComparativeMarkerSelection" with version "10". Below the title, there is a description: "Identify differentially expressed genes that can discriminate between distinct classes of samples." A note indicates that fields marked with an asterisk are required. There are "Reset" and "Run" buttons.

The configuration fields are:

- input file\***: A file selection box showing "Hide Files...(Selected 1 files)" and a text input field containing the URL: `https://dm.genomespace.org/datamanager/file/Home/ted/demo/step2/TF.data.tab?dataformat=http://www.genomespace.org/datamanager/dataformat/gct`.
- cls file\***: A file selection box showing "Hide Files...(Selected 1 files)" and a text input field containing the URL: `https://dm.genomespace.org/datamanager/file/Home/ted/demo/step2/DMAP.classes.cls`.

Below the "cls file\*" field, there is a note: "The class file - .cls".

The footer of the page contains "About GenePattern | Contact Us" on the left and "©2003-2014 Broad Institute, MIT" on the right.



# Step 3: Send differentially expressed genes to Genomica

The screenshot shows the GenePattern web interface. The browser address bar displays `genepattern.broadinstitute.org/gp/pages/index.jsf?lsid=urn:lsid:broad.mit.edu:cancer.software.genepattern`. The main navigation bar includes **GenePattern** and links for [System Message](#), [My Settings](#), and [Sign out](#) (ted1). Below the navigation bar are tabs for **Modules & Pipelines**, **Suites**, **Job Results**, **Resources**, **Downloads**, **Help**, and **GenomeSpace**.

The left sidebar shows a file tree with folders **step1**, **step2**, **step3**, **step4**, and **test**. Under **step3**, the file **TF.data.genesymbol.gct** is selected. The main content area displays the file name **TF.data.genesymbol.gct** and offers actions: **Delete File** (Permanently delete this file.), **Save File** (Save a copy of this file to your local computer.), and a **Send to GenomeSpace Tool** section. The **Send to Genomica** option is highlighted with a red box. Other options include **Send to Galaxy** and **Send to Gitools**.

The right sidebar shows a **Documentation** tab and a **Run** button. Below the **Run** button, there are file links in a dashed box, including `ce.org/datamanager/file/Home/ted/demo/step2/TF.data.tab?dataformat` and `pace.org/datamanager/dataformat/gct`. A note states: "Note that if your data is log transformed, you will need to set the 'log' parameter below to 'yes'".

At the bottom of the page, there is a footer with [About GenePattern](#) | [Contact Us](#) on the left and [©2003-2014 Broad Institute, MIT](#) on the right.



# Step 3: perform module network analysis in Genomica



# Step 4: Visualize regulators with known SNPs and linkage regions

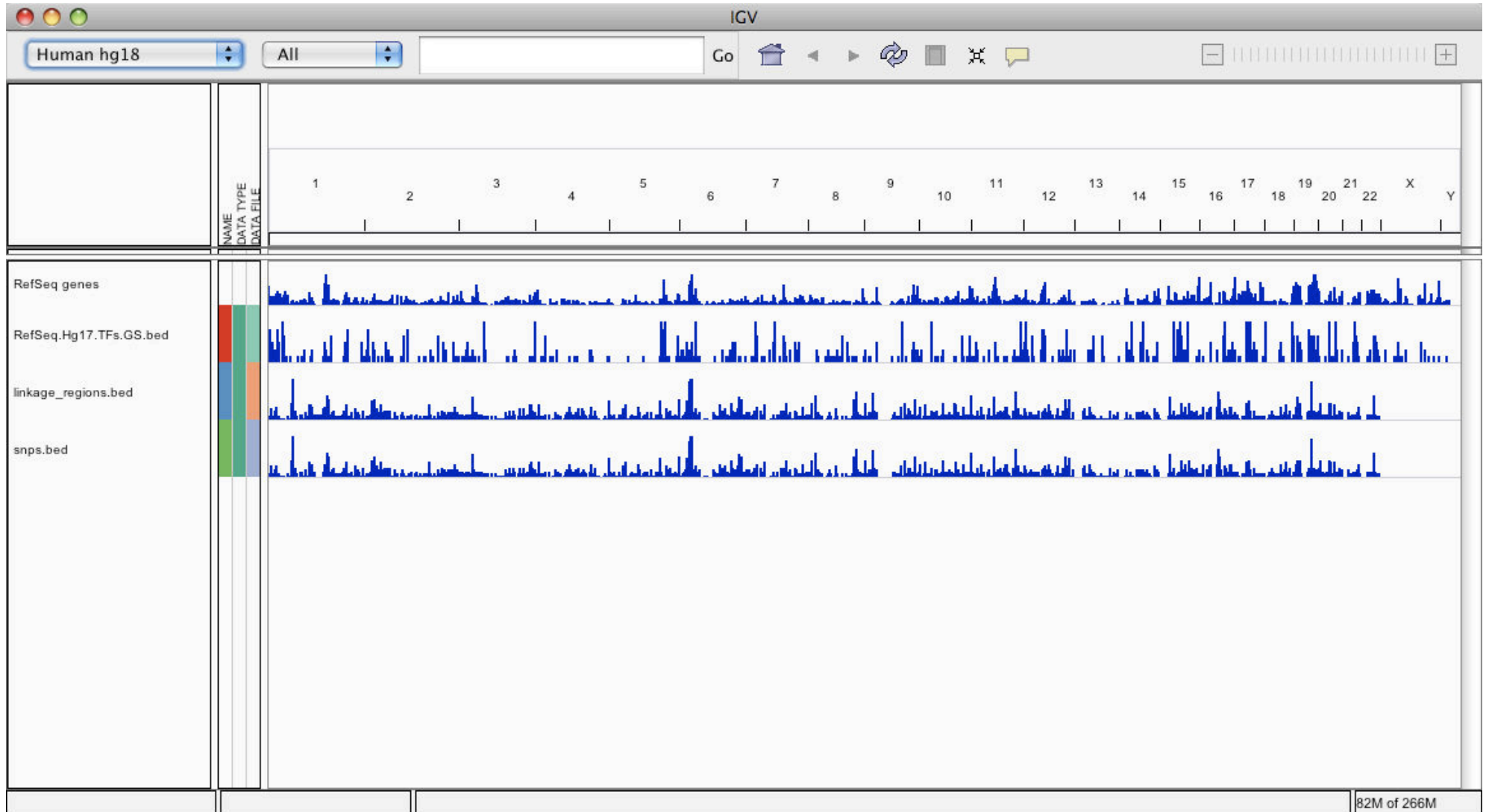
The screenshot shows the GENOMESPACE BETA web interface. The top navigation bar includes 'File', 'Launch', 'View', 'Connect', 'Manage', 'Recipes', and 'Help'. Below this is a toolbar with various tools: Pattern, Genomics, geWorkbench, Gitoos, HaploViewPlink, IGV, InSilicoDB, and ISACreator. The main content area shows a file browser with the path 'Home > ted > demo > step4'. A table of files is displayed, with three files selected: 'LinkageRegions.HematologicDisorders.bed.bed', 'Regulators.bed.bed', and 'SNPs.bed.bed'. A 'Launch IGV' dialog box is open, showing the 'Currently Selected Files' and 'Launch' and 'Close' buttons.

Filename	Tags	Owner	Size	Last Modified
<input type="checkbox"/> Galaxy.SNPsLinkageRegionsTFs.xls		ted	501 Kb	19 months ago
<input type="checkbox"/> HSCRegulatorsGeneSet.gmt		ted	63 bytes	19 months ago
<input checked="" type="checkbox"/> LinkageRegions.HematologicDisorders.bed.bed		ted	3 Kb	19 months ago
<input checked="" type="checkbox"/> Regulators.bed.bed		ted	47 Kb	19 months ago
<input checked="" type="checkbox"/> SNPs.bed.bed		ted	127 Kb	19 months ago
<input type="checkbox"/> TF.data.cytosacpe.attr		ted	1.31 Mb	19 months ago
<input type="checkbox"/> Launch IGV			0 Kb	19 months ago

**Currently Selected Files**

- .../LinkageRegions.HematologicDisorders.bed.bed
- /ted/demo/step4/Regulators.bed.bed
- /ted/demo/step4/SNPs.bed.bed

## Step 4: Visualize regulators with known SNPs and linkage regions



# How about a non-retrospective example?

[Cancer Discov.](#) 2013 Mar;3(3):308-23. doi: 10.1158/2159-8290.CD-12-0418. Epub 2013 Feb 21.

## Targeting MYCN in neuroblastoma by BET bromodomain inhibition.

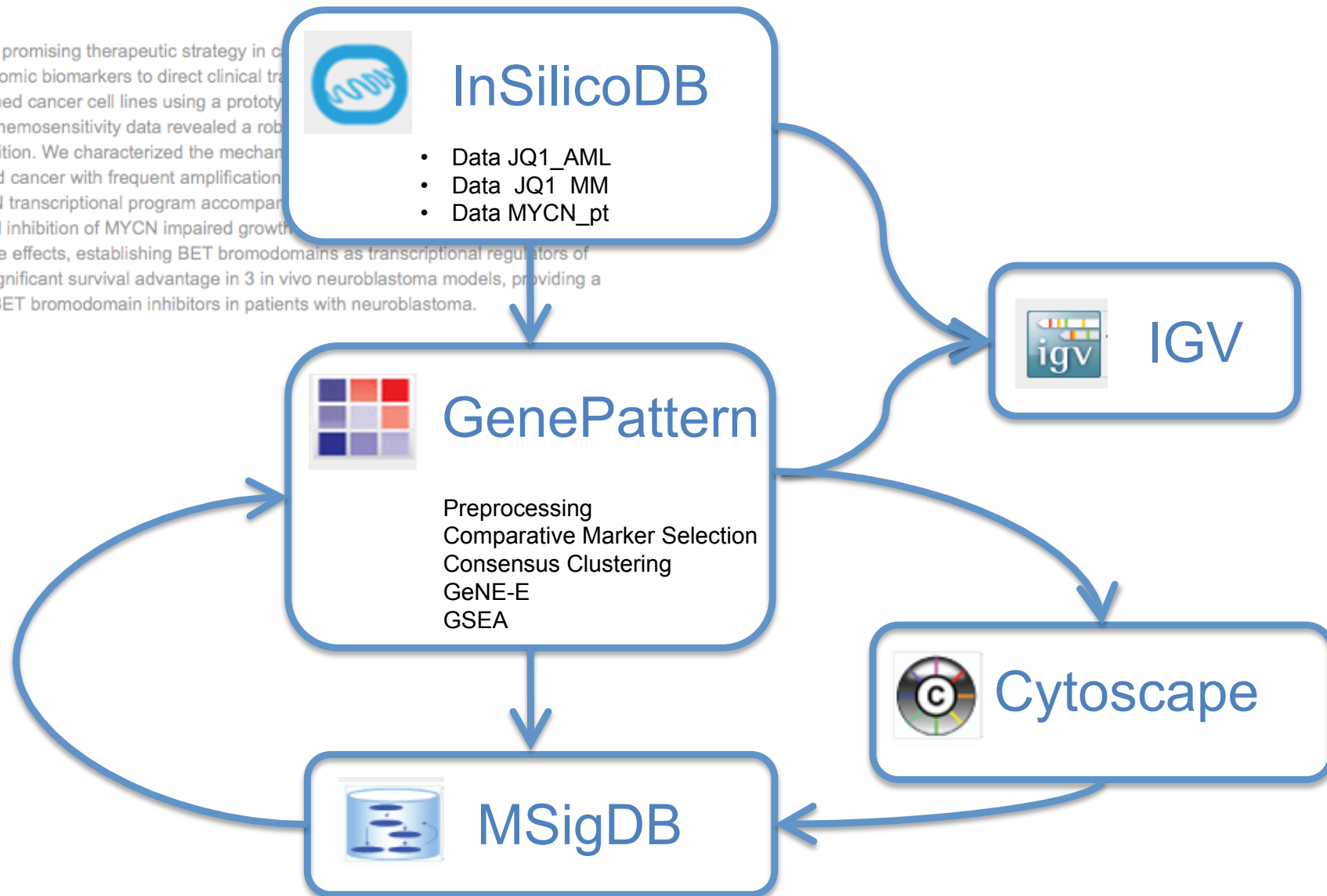
[Puissant A](#), [Frumm SM](#), [Alexe G](#), [Bassil CF](#), [Qi J](#), [Chanthery YH](#), [Nekritz EA](#), [Zeid R](#), [Gustafson WC](#), [Greninger P](#), [Garrett MJ](#), [McDermott U](#), [Benes CH](#), [Kung AL](#), [Weiss WA](#), [Bradner JE](#), [Stegmaier K](#).

Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02215, USA.

### Abstract

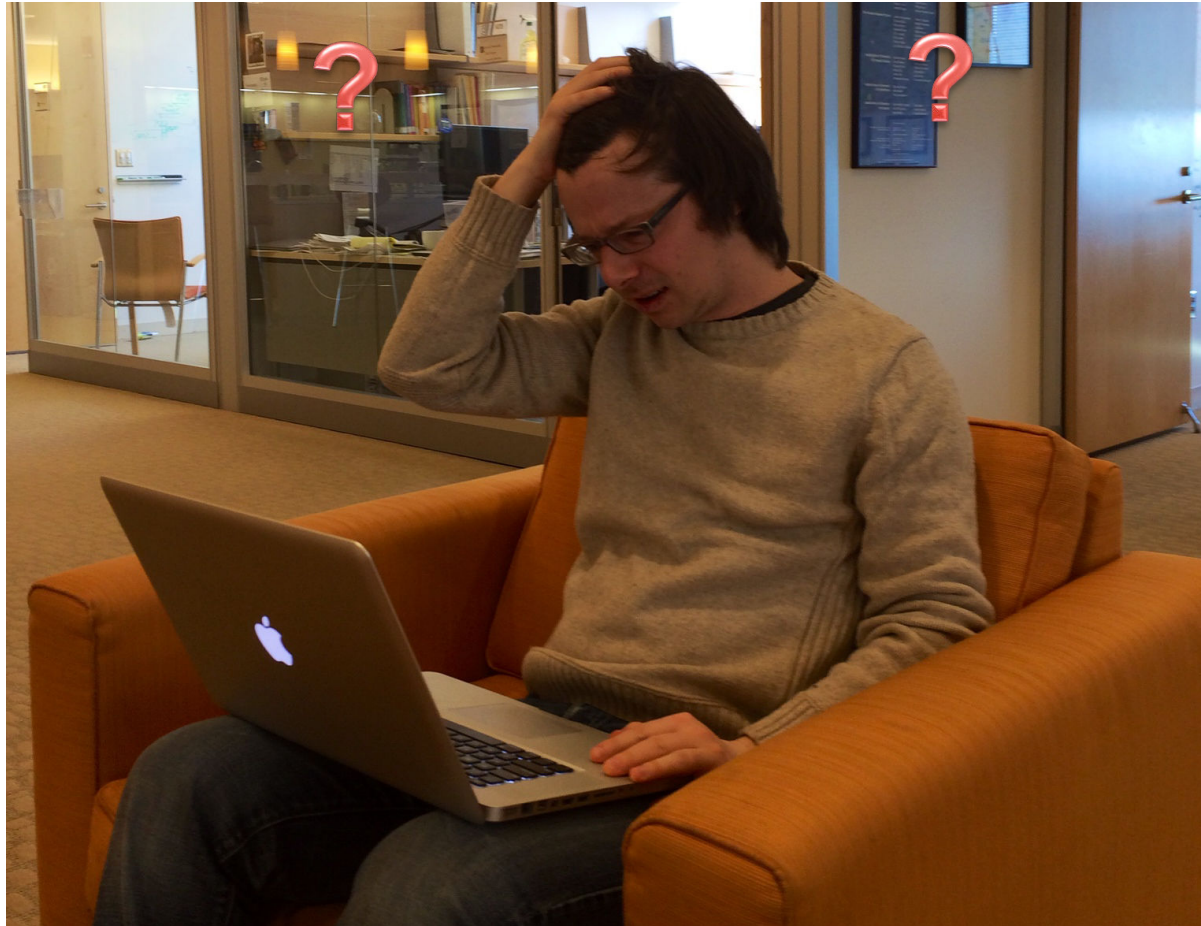
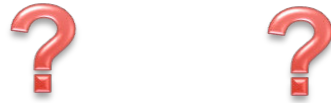
Bromodomain inhibition comprises a promising therapeutic strategy in cancer malignancies. To date, however, genomic biomarkers to direct clinical trials are lacking. We performed a cell-based screen of genetically defined cancer cell lines using a prototype integrative approach. Integration of genetic features with chemosensitivity data revealed a robust and sensitive biomarker for response to bromodomain inhibition. We characterized the mechanism of action of bromodomain inhibition in neuroblastoma, a childhood cancer with frequent amplification of MYCN. We found that bromodomain inhibition showed downregulation of the MYCN transcriptional program accompanied by increased cell growth. Functionally, bromodomain-mediated inhibition of MYCN impaired growth of neuroblastoma cells. BRD4 knockdown phenocopied these effects, establishing BET bromodomains as transcriptional regulators of MYCN. BET inhibition conferred a significant survival advantage in 3 in vivo neuroblastoma models, providing a compelling rationale for developing BET bromodomain inhibitors in patients with neuroblastoma.

### Comment in

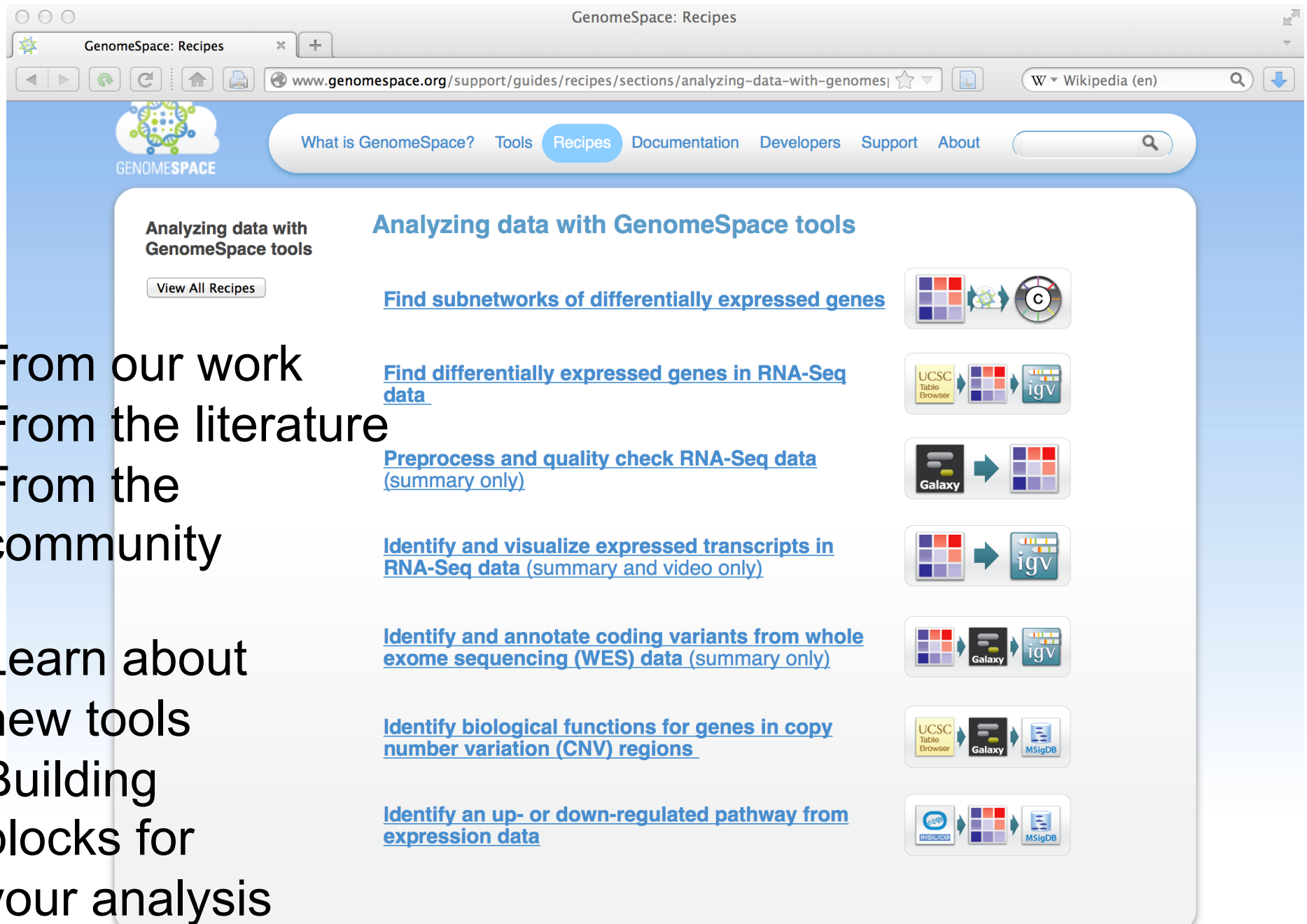




# How do I use it???



# GenomeSpace Recipe Collection



The screenshot shows the GenomeSpace website interface. The browser address bar displays the URL: [www.genomespace.org/support/guides/recipes/sections/analyzing-data-with-genomes](http://www.genomespace.org/support/guides/recipes/sections/analyzing-data-with-genomes). The website header includes the GenomeSpace logo and navigation links: "What is GenomeSpace?", "Tools", "Recipes", "Documentation", "Developers", "Support", and "About". A search bar is located on the right side of the header.

The main content area is titled "Analyzing data with GenomeSpace tools" and features a "View All Recipes" button. Below this, there is a section titled "Analyzing data with GenomeSpace tools" which lists several recipes, each with a brief description and a corresponding icon representing the tools used in the recipe.

1. From our work
2. From the literature
3. From the community

- Learn about new tools
- Building blocks for your analysis

The recipes listed are:

- [Find subnetworks of differentially expressed genes](#) (Icon: Heatmap, Network, Cytoscape)
- [Find differentially expressed genes in RNA-Seq data](#) (Icon: UCSC Table Browser, Heatmap, IGV)
- [Preprocess and quality check RNA-Seq data \(summary only\)](#) (Icon: Galaxy, Heatmap)
- [Identify and visualize expressed transcripts in RNA-Seq data \(summary and video only\)](#) (Icon: Heatmap, IGV)
- [Identify and annotate coding variants from whole exome sequencing \(WES\) data \(summary only\)](#) (Icon: Heatmap, Galaxy, IGV)
- [Identify biological functions for genes in copy number variation \(CNV\) regions](#) (Icon: UCSC Table Browser, Galaxy, MSigDB)
- [Identify an up- or down-regulated pathway from expression data](#) (Icon: INSIGO, Heatmap, MSigDB)

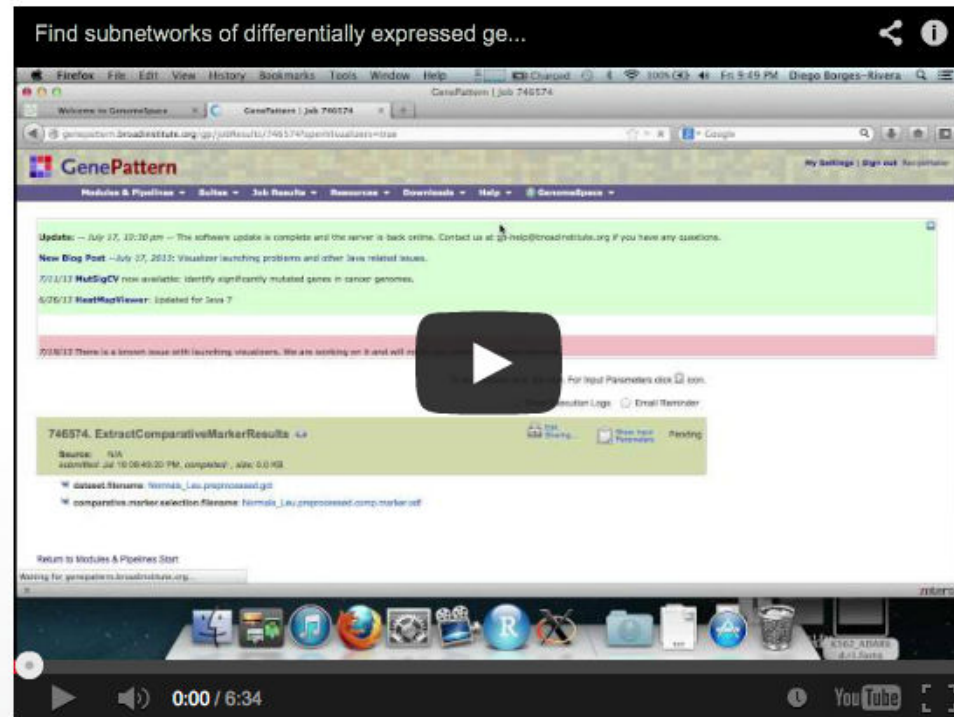
# Example Recipe: find subnetworks of differentially expressed genes.

## Find subnetworks of differentially expressed genes

1. Go to GenePattern from GenomeSpace
2. Filter out rows having less than 3x expression variation
3. Find differentially expressed genes
4. Select the top 50 genes by rank
5. Get a gene list
6. Get the network of interacting proteins
7. Navigate and explore the protein-protein network
8. Find subnetworks

[View All Recipes](#)

## Find subnetworks of differentially expressed genes



## Summary

This recipe provides one method to finding subnetworks of interacting proteins that are differentially expressed between normal cells and tumor cells.

### Input:

- A gene expression dataset, containing data from two different conditions or phenotypes, such as tumor and normal
- A class file identifying the classes of samples in the gene expression dataset

### Steps:

1. Find the list of the 50 top differentially expressed genes in a gene expression dataset using **PreprocessDataset**, **ComparativeMarkerSelection**, and **ExtractComparativeMarkerResults** (in **GenePattern**)
2. Find the protein-protein interaction subnetworks associated with these genes using **GeneMANIA** and **MCODE** (in **Cytoscape**)

Workflow →

← Video

Summary →

Input →

Steps and Tools →



# Example Recipe: find subnetworks of differentially expressed genes.

Use case →

## Recipe Details

We describe the detailed steps below in the context of an example, where we are using gene expression data for normal progenitor and leukemic samples from a study of transformation from committed progenitor cells to leukemia stem cells initiated by MLL-AF9 to find differentially expressed subnetworks. The example data is from mouse (*Mus musculus*) cell lines. Expression input data is in GenePattern's [GCT format](#), and the two classes of cells are identified in GenePattern's [CLS format](#). The example data files are available in the GenomeSpace Public folder (*Public>SharedData>Demos>Analysis Recipe 1*):

- Normal\_Leu.gct (Gene expression data for normal *Mus musculus* progenitor and leukemic samples from a study of transformation from committed progenitor cells to leukemia stem cells initiated by MLL-AF9)
- Normal\_Leu.cls (Class assignments -- tumor or normal -- for all the samples in the GCT file)



## GenePattern: Find differentially expressed genes

Detailed steps →

### 1. Go to GenePattern from GenomeSpace

- a. Log into GenomeSpace.
- b. Click the GenePattern button in the GenomeSpace tool bar.
- c. In the GenePattern window of the browser, click *Sign In Using GenomeSpace*.

**Note:** If you have not yet associated your GenomeSpace login with your GenePattern login, you will be asked to do so, or you can automatically generate a new GenePattern account that will be associated with your GenomeSpace login.

A screenshot of the GenePattern web interface. At the top, there is a header with the GenePattern logo and name. Below the header, there is a login section with the text "Sign in to GenePattern" and a link "Click to Register". There are two input fields: "Username:" and "Password:". Below the password field is a "Sign in" button and a link "Forgot your password?". To the right of the login section is the GenomeSpace logo and a link "Sign in Using GenomeSpace".

### 2. Filter out rows having less than 3x expression variation

The **PreprocessDataset** module trims out many of the rows of data for genes with expression variation less than the 3x (up- or down-regulated) threshold, and also normalizes the data.

- a. In the *Modules & Pipelines* search box, start typing `PreprocessDataset`, and click the name when it appears in the options dropdown under the box.



# Where we're at...

- Over 5000 registered users
- 19 public tools enabled + 15 private tools
  - ~400 tool launches/week, ~1500 up/downloads, ~3TB user data
- 19 recipes posted
  - RNA-Seq, Microarray, CNV, Network data, etc.
- Publications
  - GenomeSpace manuscript in draft
  - 'Cytoscape: the network visualization tool for GenomeSpace workflows' July 2014, in F1000Research
- Exploring moving deployment from AWS to TACC and Bionimbus
- Developing new, dynamic, recipe system and new recipes



# Explore GenomeSpace

sign up at [www.genomespace.org](http://www.genomespace.org)

- **Researchers**

- Analyze data
- Request format converters
- Contribute analysis recipes

- **Cores**

- Contribute and share recipes for your customers
- Add your local tools

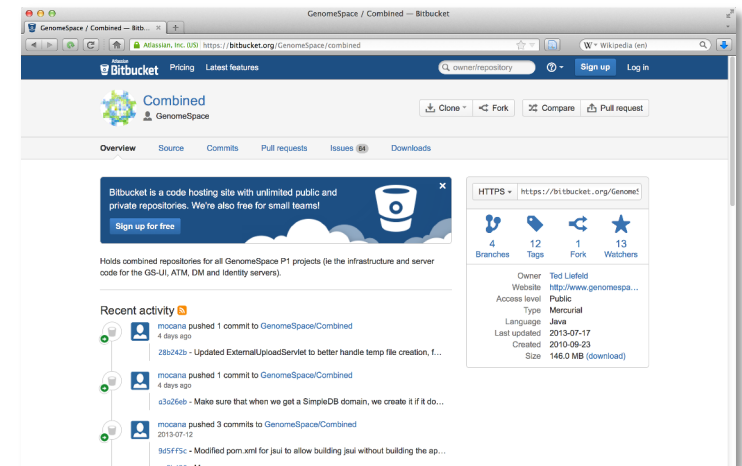
- **Developers**

- Add your tools
- Contribute format converters
- Build new infrastructure
- Hook up your app to ours



- **Data providers**

- Link your data



*Fork us on Bitbucket*



# Acknowledgements

## GenomeSpace Collaborators

**Cytoscape:** Trey Ideker Lab, UCSD

**Galaxy:** Anton Nekrutenko Lab, Penn State University

**Genomica:** Eran Segal Lab, Weizmann Institute

**UCSC Browser Team**

**GenePattern Team**

**IGV Team**

## **Driving Biological Projects**

**Howard Chang Lab** – Stanford University

**Aviv Regev Lab** – Broad Institute

## Broad Institute

Sara Garamszegi

Marco Ocana

Michael Reich

Jim Robinson

Thorin Tabor

Helga Thorvaldsdottir

Felix Wu

**Jill Mesirov, PI**




## Funding



National Human  
Genome Research  
Institute



[www.genomespace.org](http://www.genomespace.org)

Long read assembly	bioRxiv citation	More than 100 words on a slide, or a poster with no diagrams	"Big Data" or "Big Science"	Any controversy regarding the reproducibility of an analysis
 Poor/detracting use of animation in a talk	Any reference to a Lior Pachter blog post	Inappropriate left-handed DNA	<b>STAP cells</b>	Arguments over which RNA-seq algorithm to use (DESeq, EdgeR, ...)
Unlabelled axis	Talk starts early or late	<b>#ISMB Bingo 2014</b>	Oxford Nanopore praised & PacBio insulted, or vice-versa	Tweet read out
Ridiculogram/confusogram (i.e. unintelligible network diagram)	 Bingo card appears in a talk	 Any poor use of colour	<b>ROC curve</b>	<b>Absolutely anything related to XKCD</b>
That Moore's Law plot	Apophenia anecdote (i.e. someone finds a 'result' in random data)	<b>CRISPR/Cas9</b>	Speaker (who is not affiliated with PLoS) wearing a free PLoS t-shirt	"I won't go into the mathematics"

Huge thanks to everybody live-tweeting at #ISMB this year :) @jonathancairns



FIN

