

# Introduction to the 6<sup>th</sup> Genboree Epigenome Informatics Workshop

Bioinformatics Research Laboratory



March 4-5, 2013  
Houston, Texas

# **Workshop Objective:**

## **Catalyze Conversion of Epigenomic Profiling Data into Biological Insights through Integrative Analysis**

- Introduction to Workshop, Epigenome Informatics, Genboree
- Methods (Assays, Data Processing)
- Standards (Metadata, Interoperability)
- Data Resources (Human Epigenome Atlas, ENCODE)
- Tools (Epigenomic Toolset, Genboree Workbench, WashU Browser)
- Use Cases
- Collaborative Opportunities / Networking / Exchange of Experience

## **Monday, Mar 4<sup>th</sup>, 2013**

8:00 – 8:45 am      Badge pickup & continental breakfast (8<sup>th</sup> floor, Hilton )

### **Session 1**

8:45 – 9:00 am      Introduction to the Workshop – Matt Roth

9:00 – 9:45 am      *Introduction to Epigenome Analysis & Genboree – Aleks Milosavljevic*

### **Session 2**

9:45 – 10:45 am      *Use case preparation – Setting up projects, databases, groups, accessing files, user privileges, navigating Genboree, toolsets, submitting jobs, etc. (BRL staff)*

**10:45 – 11:00 am Break**

11:00 – 1:00 pm      Hands-on Case Studies: Epigenomic variation between tissues, individuals, and in Cancer (BRL staff)

**1:00 – 2:00 pm      Lunch**

## Monday, Mar 4<sup>th</sup>, 2013

### Session 3

2:00 – 3:00 pm	Analysis of Infinium Methylation Arrays (BRL staff)
3:00 – 3:45 pm	Visualizing Human Epigenomic Data via WashU Genome Browser (Xin Zhou, Wash U)
4:00 – 5:30 pm	Track A: Analysis of individual data (for those who uploaded data), or continuation of use cases (attendees & BRL staff)
4:00 – 5:30	Track B: Discussion of the “programmable web”, data and tool integration, and REST APIs (Aleks Milosavljevic)
6:00 pm	<b>Depart for dinner together (place TBD) <u>or</u> dinner on your own</b>

## Tuesday, Mar 5<sup>th</sup>, 2013

8:30 – 9:00 am	Continental breakfast (outside meeting room)
----------------	--

### Session 4

9:00 – 9:15 am	Review of Day 1 and preview of Day 2 – Matt Roth
----------------	--

## Tuesday, Mar 5<sup>th</sup>, 2013

9:15 – 10:00 am     *Quantitative profiling of histone modifications, peak calling and segmentation of epigenomic signals, Chip-Seq, RNA-Seq (BRL staff)*

10:00 – 12:00 pm     Hands-on Case Study: Chip-Seq & RNA-Seq analysis  
(attendees & BRL staff)

**12:00 – 1:00 pm     Boxed lunch (outside auditorium)**

### Session 5

1:00 – 3:00 pm     Track A: Analysis of individual data (for those who uploaded data), or continuation of use cases (*attendees & BRL staff*)

1:00 – 3:00 pm     Track B: Discussion of the “programmable web”, data and tool integration, and REST APIs (Aleks Milosavljevic)

3:00 – 4:30 pm     Wrap-up data analysis: semi-structured time for completing individualized data analysis, case studies, informal discussions (attendees & BRL staff)

**4:30 – 5:00 pm     Open discussion and wrap-up, adjourn**

# BRL Hosted Genboree Epigenome Informatics Workshops (on-site - Houston, TX)

**March, 2012:** Genboree end-user focused workshop, introduction to epigenome data analysis  
-people wanted more hands-on with tools

**May, 2012:** Launch of RNA-Seq & Chip-Seq tools

**October, 2012:** Attendees could upload own data. Introduced Spark (epigenomic data visualization, *Nielsen et al.* Genome Research) & “programmable web” session

**March, 2013:** Attendees could upload own data. Expanded Spark demo, WashU demo

## **Workshop Participants:**

**Brief introductions to facilitate networking;**

**Please describe your goals in attending workshop**

# Welcome to The Genboree Workbench!

The screenshot displays the Genboree Workbench web interface. At the top is a navigation bar with links: Home, Workbench, Browser, Profile, Groups, Projects, Databases, Tools, Log Out, and Help. Below this is a banner featuring the 'GENBOREE' logo and the BCM logo. A secondary navigation bar contains tabs for System/Network, Data, QC and Pre-processing, Genome, Transcriptome, Cistrome, and Epigenome. The main content area begins with a 'Welcome to the Genboree Workbench!' message and a '[Getting Started]' link. On the left, the 'Data Selector' panel shows a 'Refresh' button, a 'Data Filter' dropdown, and a tree view of data sources under 'www.genboree.org', including Atlas Tools Access, EDACC, Epigenome Informatics Workshop (May 2012), Epigenome ToolSet Demo Input Data, Epigenomics Roadmap Repository, GenboreeUser\_group, GMT\_Tutorial, Group1, JonathanMill\_Lab, paithank\_group, Public, ROI Repository, and Targeted Atlases. On the right, the 'Details' panel has an 'Attribute' section. Below it, the 'Input Data' section includes up, down, and delete icons. At the bottom right, the 'Output Targets' section also includes up, down, and delete icons. Five callout boxes provide context: one points to the 'Data Selector' tree, another to the 'Details' panel, a third to the 'Input Data' section, a fourth to the 'Output Targets' section, and a fifth to the 'Data Selector' filter area.

Home Workbench Browser Profile Groups Projects Databases Tools Log Out Help

GENBOREE BCM

System/Network Data QC and Pre-processing Genome Transcriptome Cistrome Epigenome

Welcome to the Genboree Workbench! [Getting Started]

**Data Selector**

Refresh Data Filter: Select a filter...

- www.genboree.org
  - Atlas Tools Access
  - EDACC
  - Epigenome Informatics Workshop (May 2012)
  - Epigenome ToolSet Demo Input Data
  - Epigenomics Roadmap Repository
  - GenboreeUser\_group
  - GMT\_Tutorial
  - Group1
  - JonathanMill\_Lab
  - paithank\_group
  - Public
  - ROI Repository
  - Targeted Atlases

**Details**

Attribute

**Input Data**

↑ ↓ ✕

**Output Targets**

↑ ↓ ✕

Specific information on files/samples selected in the "Data Selector"

Tells the tool to use this data/file

Tells the tool where to deposit results

Various Data Types (tracks, files, ROIs, etc)



# Create & Manage Groups

The screenshot displays the GENBOREE Workbench interface. At the top, a navigation bar includes links for Home, Workbench, Browser, Profile, Groups, Projects, Databases, Tools, Log Out, and Help. Below this, the GENBOREE logo is prominently displayed on the left, and the BCM (Baylor College of Medicine) logo is on the right. A secondary navigation bar features tabs for System/Network, Data, QC and Pre-processing, Genome, Transcriptome, Cistrome, Epigenome, Metagenome, and Visualization. The main content area is titled "Welcome to the Genboree Workbench! [Getting Started]". On the left, a sidebar menu is open, showing options like User Profile, Groups, Hosts, Jobs, and Request Feature. The "Groups" option is selected, revealing a sub-menu with actions: Create Group, Edit Group Info, Delete Group, Add Existing User To Group, Add New User To Group, Update Roles, Copy Users, and Message to Group. Below the sub-menu, a list of existing groups is shown, including Atlas Tools Acc, EDACC, Epigenome Inf, Epigenome To, Epigenomics F, GenboreeUser\_group, GMT\_Tutorial, Group1, JonathanMill\_Lab, paithank\_group, Public, ROI Repository, Targeted Atlases, and vamin\_group. On the right side of the main content area, there are three panels: "Details" with an "Attribute" and "Value" table, "Input Data" with up/down arrows and a delete icon, and "Output Targets" with similar controls.

# Create & Manage Databases & Projects

The screenshot displays the GENBOREE web application interface. At the top, a navigation bar includes links for Home, Workbench, Browser, Profile, Groups, Projects, Databases, Tools, Log Out, and Help. Below this, a banner features the GENBOREE logo and the BCM Baylor College of Medicine logo. A secondary navigation bar contains tabs for System/Network, Data, QC and Pre-processing, Genome, Transcriptome, Metagenome, and Visualization. A dropdown menu for 'Databases' is open, showing options: Create Database, Rename Database, Delete Database, Edit Database Info, Unlock/Lock Database, and Publish/Retract Database. Another dropdown menu for 'Projects' is also open, showing a 'Create Project' option. A 'Data Selector' panel on the left lists various data sources, including Atlas Tool, EDACC, Epigenome Informatics Workshop (May 2012), Epigenome ToolSet Demo Input Data, Epigenomics Roadmap Repository, GenboreeUser\_group, GMT\_Tutorial, Group1, JonathanMill\_Lab, paithank\_group, Public, ROI Repository, Targeted Atlases, and vamin\_group. The main content area shows a table with columns for 'Route' and 'Value', and an 'Output Targets' section with up, down, and delete icons.

Home Workbench Browser Profile Groups Projects Databases Tools Log Out Help

GENBOREE

BCM  
Baylor College of Medicine

System/Network Data QC and Pre-processing Genome Transcriptome Metagenome Visualization

Welcome to

Data Selector

Refresh

www.genboree.org

Atlas Tool

EDACC

Epigenome Informatics Workshop (May 2012)

Epigenome ToolSet Demo Input Data

Epigenomics Roadmap Repository

GenboreeUser\_group

GMT\_Tutorial

Group1

JonathanMill\_Lab

paithank\_group

Public

ROI Repository

Targeted Atlases

vamin\_group

Databases

Entity Lists

Entrypoints

Files

Projects

Samples & Sample Sets

Tracks

Create Database

Rename Database

Delete Database

Edit Database Info

Unlock/Lock Database

Publish/Retract Database

Create Project

Route Value

Output Targets

# Welcome to the Genboree Workbench!

The screenshot shows the Genboree Workbench interface. At the top, there is a navigation bar with tabs: System/Network, Data, QC and Pre-processing, Genome, Transcriptome, Cistrome, **Epigenome**, Metagenome, and Visualization. The **Epigenome** tab is selected, and its dropdown menu is open, showing options: QC, Search for Similar Signals by Correlation, Analyze Signals, Compute Similarity Matrix (heatmap), Slice Epigenomic Data, and Analyze Signals in the Context of Epigenome Atlas. On the left, a 'Welcome to the Genboree Workbench!' message is displayed with a list of instructions. A red box highlights the 'Data Selector' tree on the left and the 'Input Data' area in the instructions.

**Welcome to the Genboree Workbench!**

- The **Data Selector** tree on the left shows the data entities to which you have access.
- Drag items to be used as tool **inputs** over to the **Input Data** area.
- Drag items to be used as **output destinations** for tool results over to the **Output Target** area.
- Tools which can be run on your selections will be highlighted in **green**.
- Unsure about what kinds of items a particular tool needs in the **Input Data** and **Output Target** areas?
  - Just click the tool button when it is *not highlighted* to see help information.

The screenshot shows the Genboree Workbench interface with the **Genome** tab selected. The dropdown menu for **Genome** is open, showing options: Analyze Structural Variants, SNPs, Collect Insert Sizes, Breakpoint Detection by Breakout, Intersect SVs, Report Multiple SVs, Epigenomic Enrichment, and Epigenomic Variation. On the left, the 'Data Selector' tree is visible, and a red arrow points to the 'Call SNPs by Atlas-SNP2' tool. The 'Welcome to the Genboree Workbench!' message is also present.

**Welcome to the Genboree Workbench!**

- The **Data Selector** tree on the left shows the data entities to which you have access.
- Drag items to be used as tool **inputs** over to the **Input Data** area.
- Drag items to be used as **output destinations** for tool results over to the **Output Target** area.
- Tools which can be run on your selections will be highlighted in **green**.
- Unsure about what kinds of items a particular tool needs in the **Input Data** and **Output Target** areas?
  - Just click the tool button when it is *not highlighted* to see help information.

The screenshot shows the Genboree Workbench interface with the **Transcriptome** tab selected. The dropdown menu for **Transcriptome** is open, showing options: Analyze RNA-Seq Data, Analyze Small RNA-Seq Data, Map Reads and Splice Junctions by TopHat, Assemble and Measure Transcripts by Cufflinks, and Detect Transcription Changes by Cuffdiff. On the left, the 'Data Selector' tree is visible, and a red arrow points to the 'Filter Reads' tool. The 'Welcome to the Genboree Workbench!' message is also present.

**Welcome to the Genboree Workbench!**

- The **Data Selector** tree on the left shows the data entities to which you have access.
- Drag items to be used as tool **inputs** over to the **Input Data** area.
- Drag items to be used as **output destinations** for tool results over to the **Output Targets** area.
- Tools which can be run on your selections will be highlighted in **green**.
- Unsure about what kinds of items a particular tool needs in the **Input Data** and **Output Targets** areas?
  - Just click the tool button when it is *not highlighted* to see help information.

The screenshot shows the Genboree Workbench interface with the **Cistrome** tab selected. The dropdown menu for **Cistrome** is open, showing options: Call ChIP-Seq Peaks by, and MACS. The 'Welcome to the Genboree Workbench!' message is also present.

**Welcome to the Genboree Workbench!**

# Epigenome Atlas Release 8 over 2000 experiments

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



Home



Releases




Informatics

Publications

Forums

Contributors

## Human Epigenome Atlas Release 4 (hg19)

- [Data Access Policy](#)
- Data embargo period: from 04/14/2011 - 01/14/2012 or earlier as specified [here](#)
- Select cells by clicking and dragging, then use the "View Selections in" pulldown in the top left corner (below) to view selections in the Atlas Gene Browser or the UCSC Genome Browser
- To see data authors, other metadata, and to download data, click a sample name in the first column or an assay type in the header row
- Human Epigenome Atlas releases are intended to be cumulative: e.g. Release 3 includes all Release 2 data and additional submissions
- NOTE: Some pages may not be accessible over low bandwidth internet connections. This page has been tested with the following browsers:   

## Human Epigenome Atlas Release 4 (hg19)

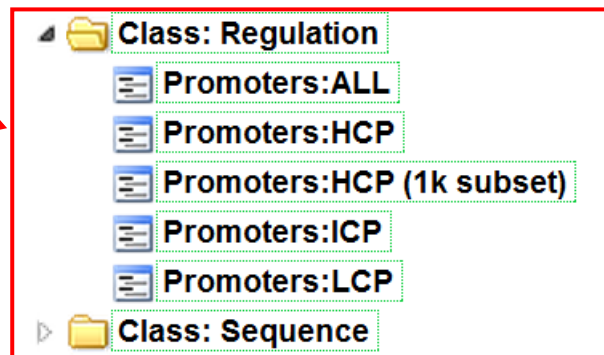
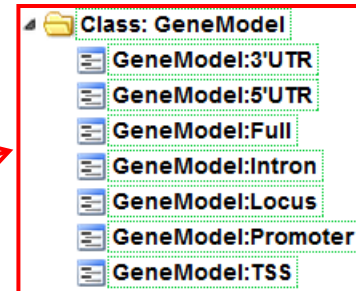
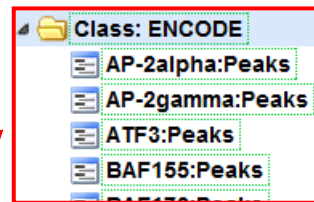
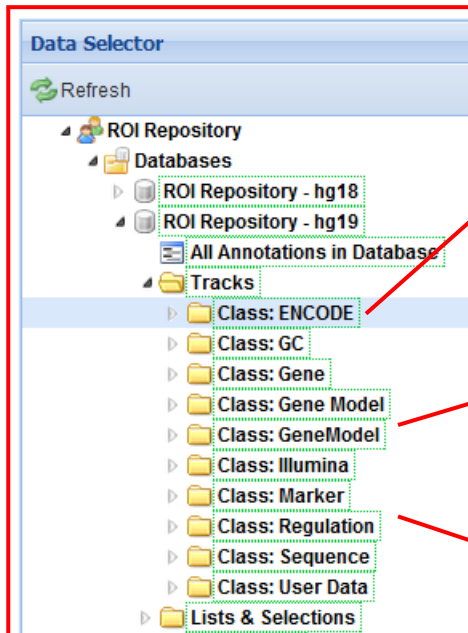
View Selections in  Clear Selections

Sample	Assay															
	Bisulfite-Seq	MeDIP-Seq	MRE-Seq	RRBS	DNAse Hypersensitivity	Digital Genomic Footprinting	mRNA-Seq	smRNA-Seq	Expression Array	ChIP-Seq Input	Histone H3K27me3	Histone H3K36me3	Histone H3K4me1	Histone H3K4me3	Histone H3K9ac	Histone H3K9me3
Brain Substantia Nigra										1	1	1	1	1	1	1
Breast Luminal Epithelial Cells	4	5					2				1	1	1			1
Breast Myoepithelial Cells	3	3					2			2	2	2	2	2	2	2
Breast Stem Cells	4	4					1									
Breast vHMEC	1	1			2		1	1		2	1	1	2	1		1
CD14 Primary Cells					2											
CD15 Primary Cells				1							1			1		
CD19 Primary Cells				1	3					1	2	2		2	2	

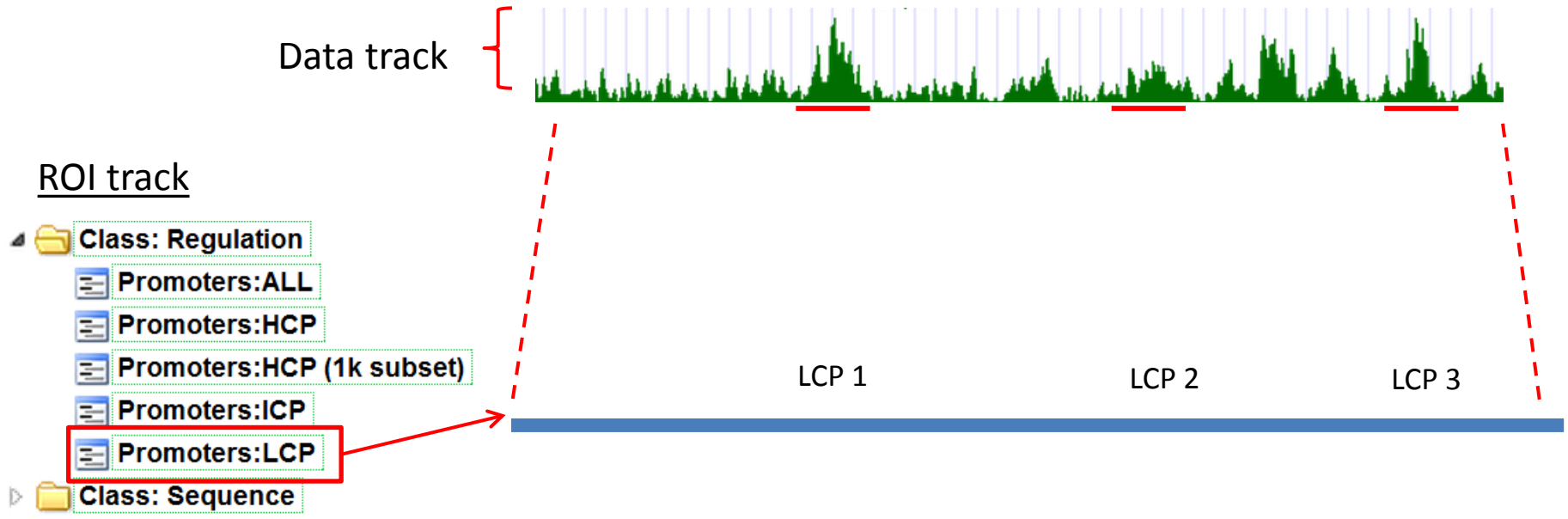
# Biology Across Key Genetic Elements (promoters, exons, UTR, etc): Many ROI (i.e. annotation) Tracks Available



## ROI tracks available

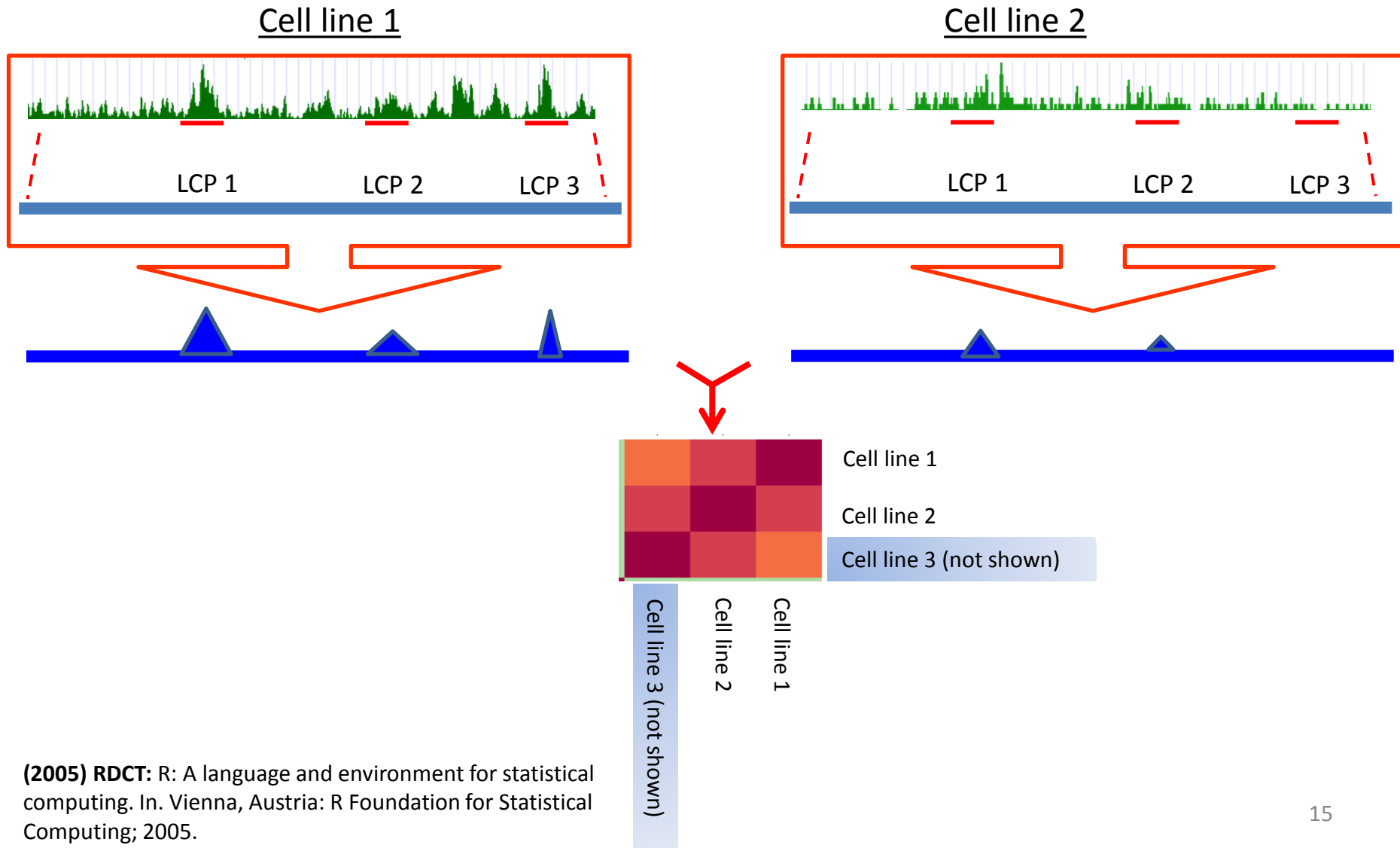


# Data Tracks from Epigenomic Experiments Projected On To ROI (i.e. annotation) Tracks



Promoters: LCP = Low-CpG promoters (as defined in Weber et al., *Nature Genetics* (2007))

# Compute Pearson Correlation Coefficient Between Experiments: Similarity Matrix is Output as Heatmap





# Viewing selections


Human Epigenome Atlas [Release 4](#) (hg19)



- [Data Access Policy](#)
- Data embargo period: from 04/14/2011 - 01/14/2012 or earlier as specified [here](#)
- Select cells by **clicking and dragging**, then use the "View Selections In" pulldown in the top left corner (below)
- To see data authors, other metadata, and to download data, click a sample name in the first column or an assay type
- Human Epigenome Atlas releases are intended to be cumulative: e.g., Release 3 includes all Release 2 data and
- **NOTE:** Some pages may not be accessible over low bandwidth Internet connections. This page has been tested

Human Epigenome Atlas Release 4 (hg19)

View Selections In  Clear Selections

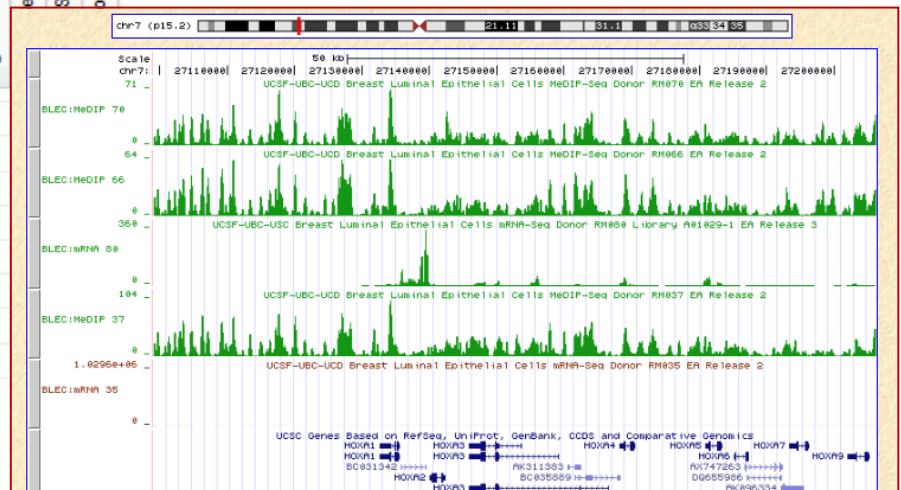
Atlas Gene Browser  Assay

Genome Browser  Choose a Mirror

-  Local UCSC browser mirror (Fast)
-  UCSC genome browser (Slow)

Sample  
Filter:  (e.g. "cell line")

Sample	Bisulfite-Seq	MeDIP-Seq	MRE-Seq	RRBS	DNAse-seq	Hypersensitivity	Digital Genomic Footprinting
Brain Substantia Nigra							
Breast Luminal Epithelial Cells		4	5				
Breast Myoepithelial Cells		3	3				
Breast Stem Cells		4	4				
Breast vHMEC		1	1		2		
CD14 Primary Cells					2		
CD15 Primary Cells				1			
CD19 Primary Cells				1	3		
CD20 Primary Cells					1		







Home Workbench Browser Profile Groups Projects Databases Tools Log Out Help

**GENBOREE**

System/Network Data QC and Pre-processing Genome Transcriptome Cistrome **Epigenome** Metagenome Visualization

**Welcome to the Genboree Workbench!**

- The **Data Selector** tree on the left shows the data entities to which you have access.
- Drag items to be used as tool **inputs** over to the **Input Data** area.
- Drag items to be used as **output destinations** for tool results over to the **Output Target** area.
- Tools which can be run on your selections will be highlighted in **green**.
- Unsure about what kinds of items a particular tool needs in the **Input Data** and **Output Target** areas? Just click the tool button when it is *not highlighted* to see help information.

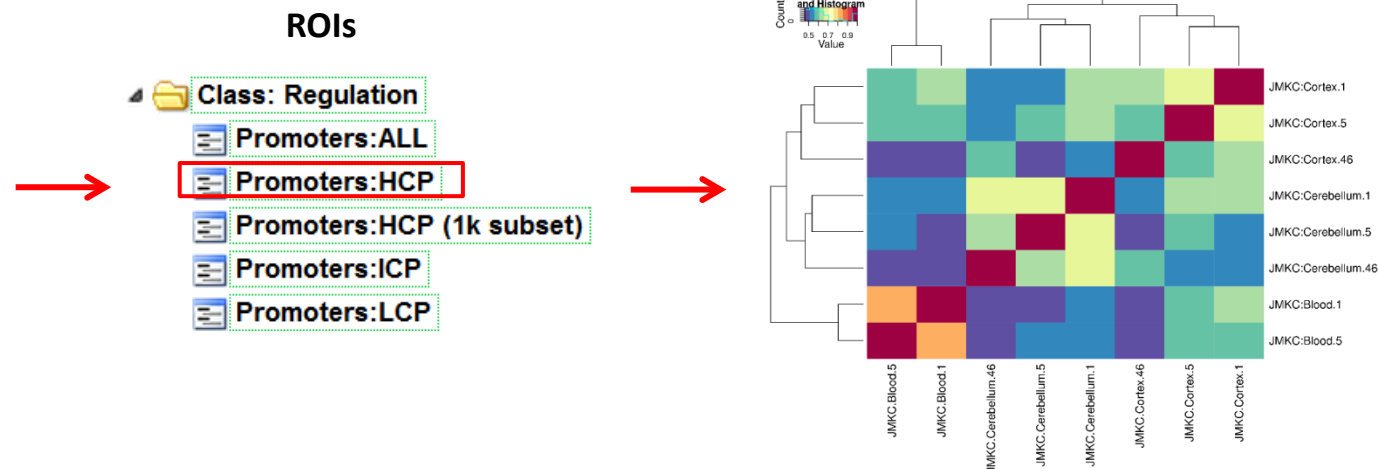
QC  
Search for Similar Signals by Correlation  
**Analyze Signals**  
**Compute Similarity Matrix (heatmap)**  
Slice Epigenomic Data  
Analyze Signals in the Context of Epigenome Atlas

## Use Case 5: Methylation of some features discriminate tissue type better than others

Similar to Use Case 1 & 2 but uses different ROIs to illustrate how different features produce different similarity matrices (heatmaps).

### Epigenomic Tracks

- Blood
- Cerebellum
- Cortex



## Use Case 9: Coordinated Changes of Epigenomic Marks Across Tissue Types

Welcome to the Genboree Workbench!

- The **Data Selector** tree on the left shows the data entities to which you have access.
- Drag items to be used as tool *inputs* over to the **Input Data** area.
- Drag items to be used as *output destinations* for tool results over to the **Output Target** area.
- Tools which can be run on your selections will be highlighted in **green**.
- Unsure about what kinds of items a particular tool needs in the **Input Data** and **Output Target** areas? Just click the tool button when it is *not highlighted* to see help information.

Epigenome

- QC
- Search for Similar Signals by Correlation
- Analyze Signals
- Compute Similarity Matrix (heatmap)
- Slice Epigenomic Data**
- Analyze Signals in the Context of Epigenome Atlas

Download Epigenomic Data Slice

Download Epigenomic Data Slice by Gene Element

### Epigenomic Tracks:

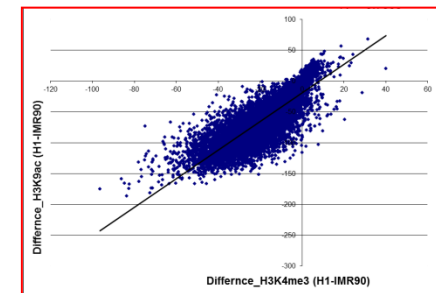
- H1 cell line
- IMR90 cell line

Collates score tracks into one data matrix, export to Excel

Bisulfite data				H1.H3K9ac						IMR90.H3K9ac						H1.H3K4me3					
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R			
1	Index	H1.BS.Con IMR90.BS Combined			H1.H3K9ac H1.H3K9ac H1.H3K9ac H1.H3K9ac H1.H3K9ac 68						IMR90.H3K90 H3K9ac 46			H1.H3K4m H1.H3K4m H1.H3K4m H1.H3K4m H1.H3K4m H1.H3K4m H1.							
2	HSAP04065	0	0.044444			0.314286 0.314286 0.314286 0.314286 0.314286						0.314286 0.314286			0.628571 0.628571 0.795714 0.314286 0.314286						
3	HSAP04065	0.047353	0.034789			4.51304 3.29043 2.35826 2.34435 2.73913						17.9513 16.5722			8.52 28.9374 23.92 11.84 2.3113						
4	HSAP04065	0.208431	0.215174			5.79688 8.58438 5.85313 3.75312 8.56875						15.5813 22.7719			17.1375 39.5844 22.9281 14.525 5.03125 22.9						
5	HSAP04065	0.209214	0.212334			1.07769 2.62314 1.87107 0.581818 1.20496						8.35537 9.14876			5.90248 14.443 27.2893 3.07769 0.363636						

Column headers = experiments  
Rows = ROIs

### Scatter plots



## Use Case 12: Assess breast cancer cell type of origin

Home Workbench Browser Profile Groups Projects Databases Tools Log Out Help

GENBOREE

System/Network Data QC and Pre-processing Genome Transcriptome Cistrome Epigenome Metagenome Visualization

Welcome to the Genboree Workbench!

- The **Data Selector** tree on the left shows the data entities to which you have access.
- Drag items to be used as tool *inputs* over to the **Input** area.
- Drag items to be used as *output destinations* for tool outputs to the **Output** area.
- Tools which can be run on your selections will be highlighted with a green background.
- Unsure about what kinds of items a particular tool needs? Just click the tool button when it is *not highlighted* to see help information.

Find Differences By Regression  
Cluster by Spark  
Compare by LIMMA

QC  
Search for Similar Signals by Correlation  
Analyze Signals  
Compute Similarity Matrix (heatmap)  
User Supplied Data Matrix  
Tracks  
Track with Sample Metadata

Data Selector

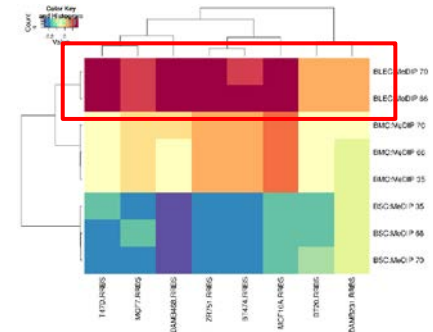
LIMMA: Smyth, G. K. Statistical Applications in Genetics and Molecular Biology (2005)

### “Your” Epigenomic Tracks (RRBS):

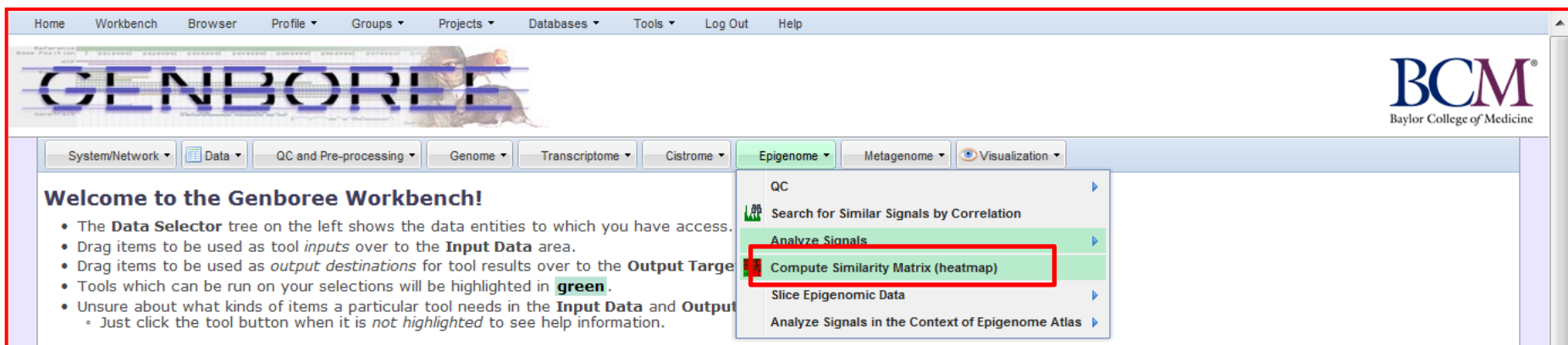
- Breast Luminal Epithelium
- Breast Myoepithelial
- Breast Stem Cell

### Public Epigenomic Tracks (MeDIP):

- Breast Luminal Epithelium
- Breast Myoepithelial
- Breast Stem Cell



## Use Case 13: Analysis of epigenomic variation in breast tumors (450K)



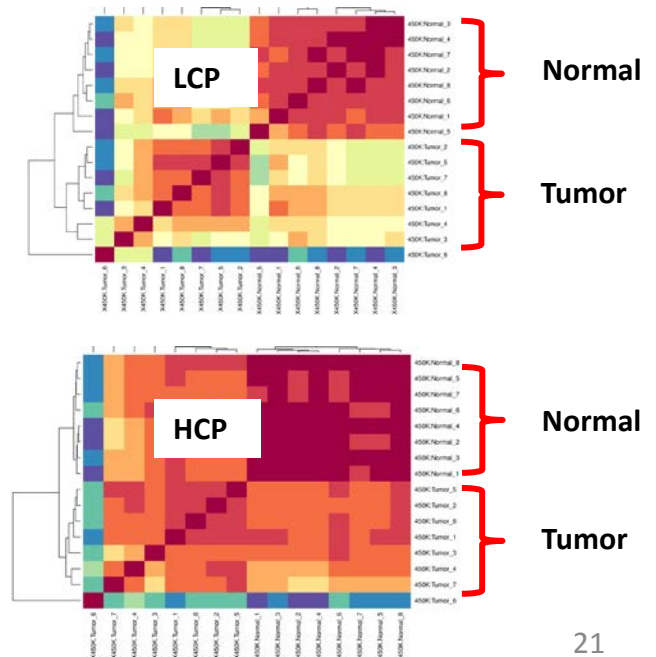
### Use Case 13a: Cluster all 16 breast tissue samples

16 450 K Samples (Dedeurwaerder, S. et al. (2011))

-8 normal breast samples

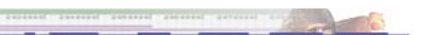

-8 cancerous breast samples

ROIs (HCP vs LCP)



### Use Case 13: Analysis of epigenomic variation in breast tumors

[Home](#)
[Workbench](#)
[Browser](#)
[Profile](#)
[Groups](#)
[Projects](#)
[Databases](#)
[Tools](#)
[Log Out](#)
[Help](#)

System/Network

Data

QC and Pre-processing

Genome

Transcriptome

Cistrome

Epigenome


Metagenome

Visualization


### Welcome to the Genboree Workbench!

- The **Data Selector** tree on the left shows the data entities to which you have access.
- Drag items to be used as tool *inputs* over to the **Input Data** area.
- Drag items to be used as *output destinations* for tool results over to the **Output Targets** area.
- Tools which can be run on your selections will be highlighted in **green**.
- Unsure about what kinds of items a particular tool needs in the **Input Data** and **Output Targets** areas?
  - Just click the tool button when it is *not highlighted* to see help information.

QC

 Search for Similar Signals by Correlation

Analyze Signals

 Compute Similarity Matrix (heatmap)

Slice Epigenomic Data

Analyze Signals in the Context of Epigenome Atlas

Use Case 13b: Compare 450K profiles (8 tumor, 8 normal) *against reference epigenomes* from the Epigenome Atlas

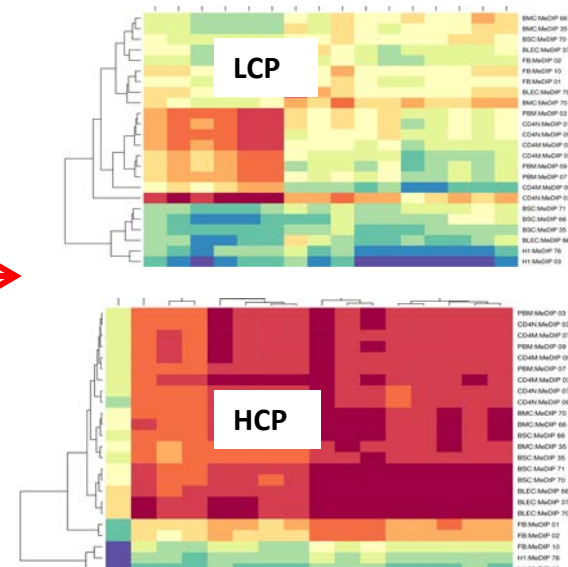
## 16 450 K Samples (Dedeurwaerder, S.et al. (2011))

-8 normal breast samples

-8 cancerous breast samples

## ROIs

- Class: Regulation
  - Promoters:ALL
  - Promoters:HCP
  - Promoters:HCP (1k subset)
  - Promoters:ICP
  - Promoters:LCP





## Use Case 13: Analysis of epigenomic variation in breast tumors

Use Case 13c: Since most breast tumor samples appear to contain excess of blood & immune cells, comparison of normal and tumor tissue may reveal differentially methylated genes (and corresponding pathways). Identify differentially methylated probes, genes, and pathways using LIMMA & online resources

16 450 K Samples (Dedeurwaerder)

-8 normal breast samples

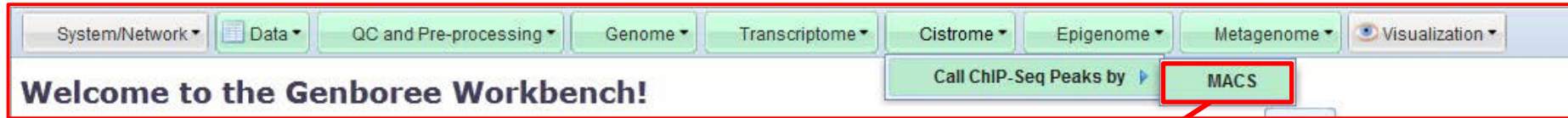
-8 cancerous breast samples

ROIs

Gene List

	B
1	PAQR7
2	VGLL4
3	TESK1
4	FGFR1
5	UCKL1
6	MGC1612
7	LCAT
8	SLC12A4
9	MTSS1L
10	FGAM4

## Use Case 14: Chip-Seq Data Analysis



Zhang et al, *Genome Biology* (2008)

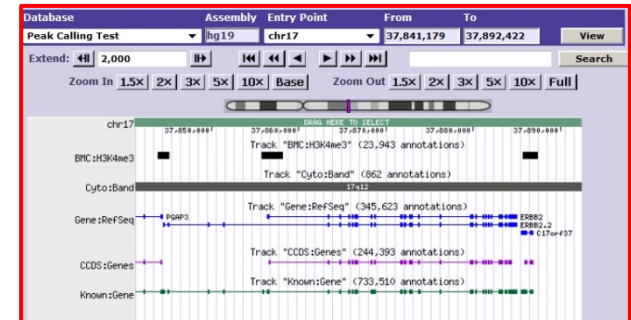
MACS results (file in Genboree)

BED  
files

	chr	start	end	length	summit	tags	-10*LOG10 fold_enrichment
20	chr1	9861	10677	817	360	292	2034.63
21	chr1	713307	715489	2183	1297	83	250.89
22	chr1	724760	727160	2401	2202	92	223.86
23	chr1	761468	763266	1799	865	133	463.34
24	chr1	833096	834002	907	577	20	96.47
25	chr1	839750	840203	454	262	15	77.67
26	chr1	859636	861418	1783	741	37	83.16
27	chr1						
28	chr1						7.95
29	chr1						2.67
30	chr1						4.84
31	chr1	911145	912090	952	598	25	125.98

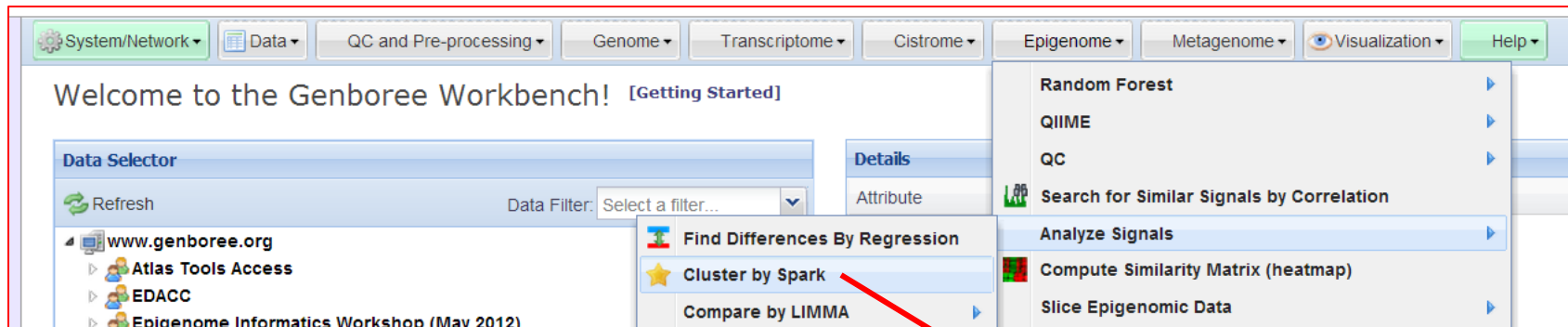
H3K4me3, H3K4me1

Visualize in Genboree

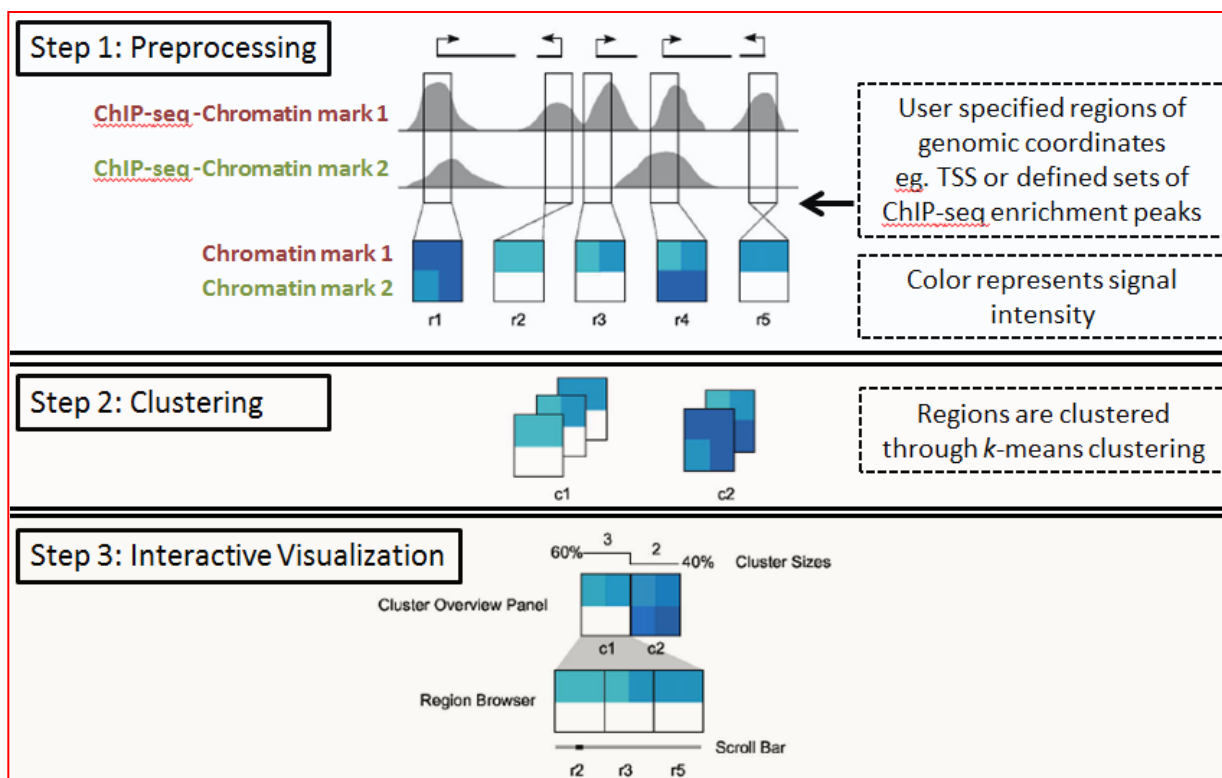




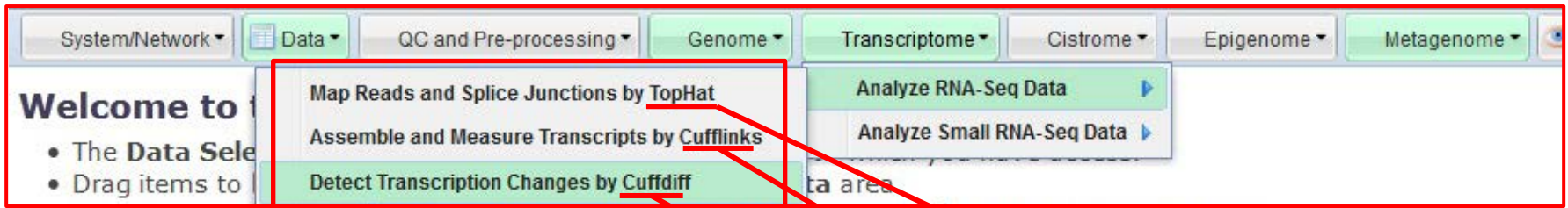
## Use Case 15: Spark Analysis



Nielsen et al, *Genome Biology* (2012)



## Use Case 16: RNA-Seq Cuffdiff Data Analysis (Use Case 17: Cufflinks & TopHat)



Trapnell et al, *Bioinformatics* (2009)

Trapnell et al, *Nature Biotech* (2010)

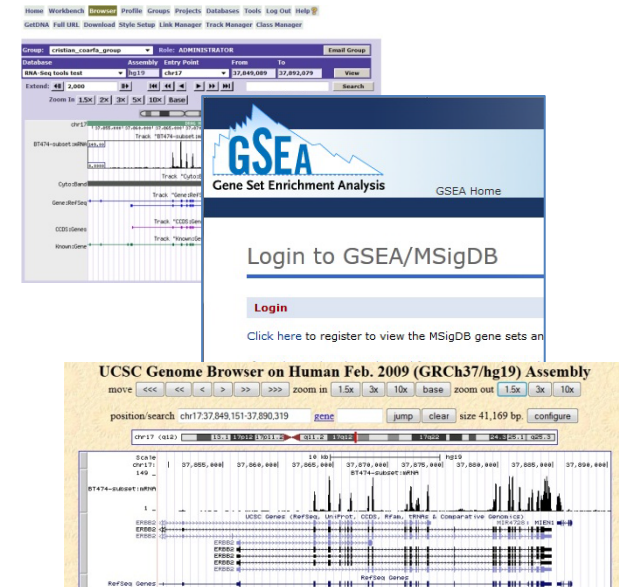
Trapnell et al, *Nature Biotech* (2010)

## Visualization/pathway analysis

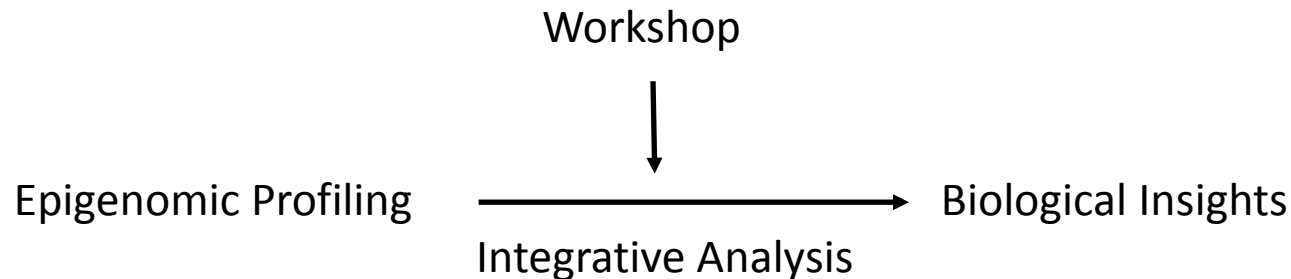
Gene expression diffs (file in Genboree)

	A	B	C	D	E	F	G
1	test_id	gene_Nar	gene_id	gene	locus	sample	sample
5	NM_0000	ACADS	NM_0000	-	chr12:121	Luminal	BasalA
10	NM_0000	ADA	NM_0000	-	chr20:432	Luminal	BasalA
32	NM_0000	AR	NM_0000	-	chrX:6676	Luminal	BasalA
41	NM_0000	ATP7B	NM_0000	-	chr13:5250	Luminal	BasalA
51	NM_0000	C3	NM_0000	-	chr19:667	Luminal	BasalA
88	NM_0001	CYBA	NM_0001	-	chr16:8870	Luminal	BasalA
91	NM_0001	CYP1B1	NM_0001	-	chr2:3829	Luminal	BasalA
195	NM_0002	ITGA6	NM_0002	-	chr2:1732	Luminal	BasalA
199	NM_0002	JAG1	NM_0002	-	chr20:106	Luminal	BasalA
254	NM_0002	NPC1	NM_0002	-	chr18:210	Luminal	BasalA
290	NM_0003	CTSA	NM_0003	-	chr20:445	Luminal	BasalA
327	NM_0003	SOX9	NM_0003	-	chr17:701	Luminal	BasalA

FASTQ,  
BAM  
files



# Workshop Evaluation (link)



- Introduction to Workshop, Epigenome Informatics, Genboree
- Methods (Assays, Data Processing)
- Standards (Metadata, Interoperability)
- Data Resources (Human Epigenome Atlas)
- Tools (Epigenomic Toolset, Genboree Workbench)
- Use Cases / Case Studies
- Collaborative Opportunities / Networking / Exchange of Experience

# Acknowledgments

- BRL
  - Aleksandar Milosavljevic
  - Cristian Coarfa, Alan R Harris, Viren Amin
  - Elke Norwig-Eastaugh
- BRL core
  - Matt Roth, Kevin Riehle, Tim Charnecki
- Genboree programming team
  - Andrew Jackson, Sameer Paithankar, Sriram Raghuram
- Former BRL members
  - Chia-Chin Wu, Arpit Tandon
- WashU
  - Xin Zhou
  - Ting Wang
- Peak Calling
  - Anshul Kundaje
  - Bob Thurman (UW)
  - Noam Shores (BI)
  - Martin Hirst (BCGSC)
  - Lee Daniels (NIH)
  - Wei Li (BCM)
- RNA-Seq
  - Adrian Lee, Joe Gray
  - Laising Yen, Kalpana Kannan
  - Sean McGuire, Christine Pichot