

Use Case 18: Mapping ontogenetic pathways of cellular differentiation  
using Human Epigenome Atlas data and the epigenome toolset  
within the Genboree Workbench

---

American Society of Human Genetics  
Boston, MA 2013

Presented by the  
Bioinformatics Research Laboratory

Baylor  
College of  
Medicine

## Summary of Use Case 18

**Background:** The epigenome plays a key role in establishing and maintaining cellular phenotype during cellular differentiation. The wealth of data from large-scale sequencing projects provides a resource for biological discovery and analysis. The Human Epigenome Atlas, developed as part of the NIH Epigenome Roadmap Project, contains Chip-Seq data from over 100 different cell-types and tissues. This data repository provides a rich resource for ongoing comparative analysis on data generated outside of the NIH Epigenome Roadmap project. We demonstrate here the use of Atlas data to further our understanding of cellular differentiation by examining cell lineage relationships based on clustering over functional genetic elements, such as promoters, enhancers, and lincRNAs.

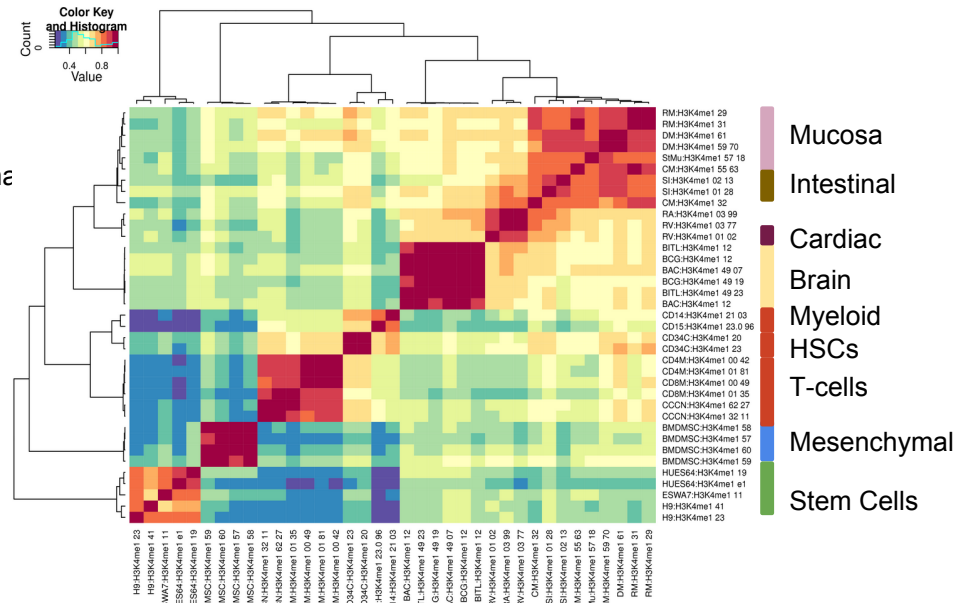
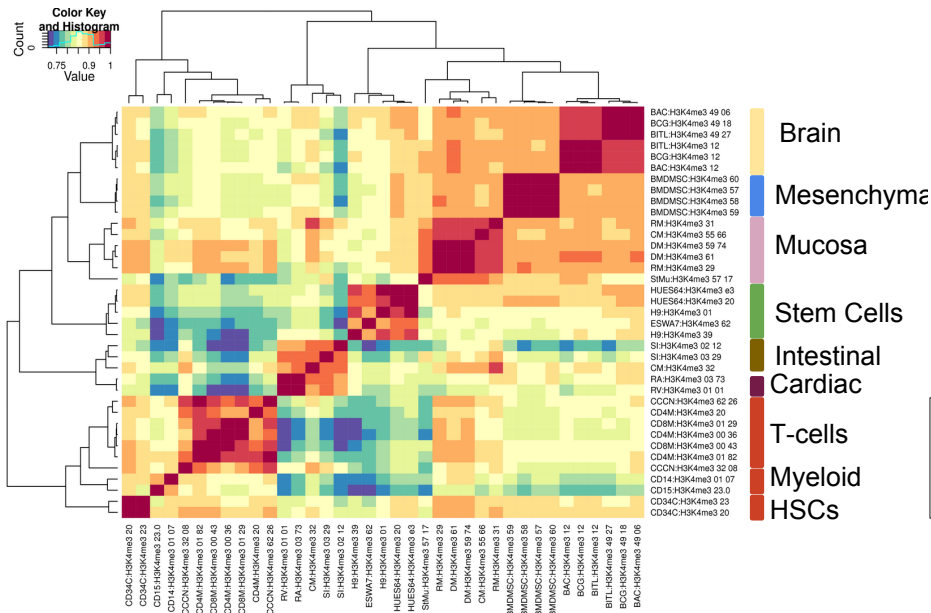
**Results:** Analysis validates current knowledge regarding H3K4me3 signals at protein coding gene promoters by clustering cell types of similar lineage<sup>1</sup>. Cluster analysis also reveals H3K4me1 signal at lincRNA promoters able to discriminate cell types/lineages, suggesting an important role for lincRNA in maintaining cellular identity.

1. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).

# Summary of Results

H3K4me3 signal over protein coding gene promoters on the NIH Roadmap Epigenome data

H3K4me1 signal over lincRNA gene promoters on the NIH Roadmap Epigenome data



Cluster analysis of the epigenomes profiled in the NIH Roadmap Epigenome project suggests that lincRNAs play an important role in maintaining cellular identity.

# Use Case Overview

**New Genboree Users** - Slides 5-13 provide steps for new Genboree users on how to create a database, a project page, and view track grid of data generated in the NIH Roadmap Epigenome Project.

**Existing Genboree Users** - If you have attended past Genboree Workshops or are familiar with the Genboree Workbench then you may briefly review these slides and start on slide 14 for the actual use case

- Methodology
- Steps for reproducing the results

# The Genboree Workbench: Web-based Data Management & Analysis

The screenshot shows the Genboree Workbench interface. At the top, there is a navigation bar with tabs for System/Network, Data, QC and Pre-processing, Genome, Transcriptome, Cistrome, Epigenome, Metagenome, Visualization, and Help. The main content area is divided into several panels:

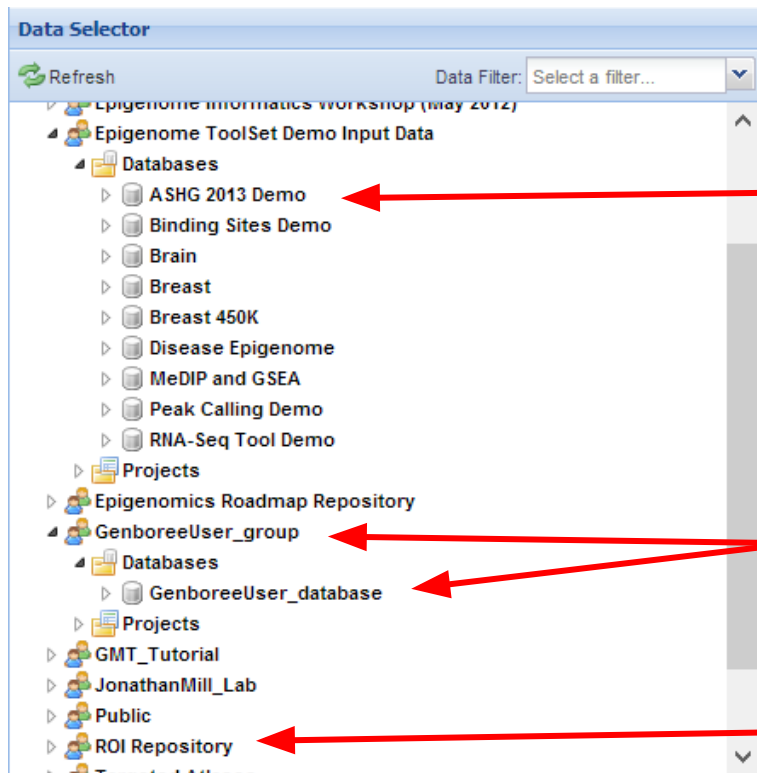
- Data Selector:** A tree view showing data sources from genboree.org, genboree.bcgsc.ca, genboree.cbrc.jp, and www.brain-research-lab.org. A callout box explains: **Data Selector:** Various Data Types (tracks, files, and ROIs (region of interests), etc).
- Details:** A table showing information for the selected data. A callout box explains: **Details:** Specific information on files/samples selected in the "Data Selector".
- Input Data:** A section for selecting input data. A callout box explains: **Input Data:** Tells the tool to use this input data/file.
- Output Targets:** A section for selecting output targets. A callout box explains: **Output Targets:** Tells the tool where to deposit results.

Red arrows point from the callout boxes to the corresponding interface elements. The 'System/Network' and 'Help' tabs are highlighted in green.

**Important:** Toolset Menu turns **GREEN** when "Input Data" and "Output Targets" are properly populated for a tool to run. Please note that "System/Network" and "Help" options are always green since "User Profile", "Jobs", and "Request Feature" are always available for use and do not need "Input Data" and "Output Targets" to be populated.

# Preparation Prior to Starting the Use Case

- “**GenboreeUser\_group**” is a name template for an automatically created Genboree user group **for you** where “**GenboreeUser**” is **your user name**.
- Similarly, “**GenboreeUser\_database**” is a name template for your database.
- Of course, you may create many more databases and may create and be member of many other groups.



Under “Epigenome Toolset Demo Input Data” you will find “ASHG 2013 Demo” database, where we have provided you with sample data to try out the use cases

When making screenshots for providing instructions we have used “GenboreeUser\_group” and “GenboreeUser\_database” as output targets, however, you will use your own group and database. Following slides will show you how to create database and project

ROI Repository contains database of annotated regions (eg. Gencode and Refseq annotations)

# Display Tool Setting “Help” dialogue box in the Workbench



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

Welcome to

Data Selector

Databases Create Database  
Entity Lists Rename Database  
Entrypoints Delete Database

Details

Attribute	Value
-----------	-------

Input Data

Output Targets

Help: Tool Settings

### Help: Create Database

Configure Tool

**⚠** This tool is a recent addition. Please contact [genboree\\_admin@genboree.org](mailto:genboree_admin@genboree.org) with questions or comments, or for help using it on your own data.

This tool will create/add a new database in the target Group. Note that the database to be created should NOT already exist in the group.

**Output Targets**

**Instructions:**

- Drag 1 destination group into "Output Targets". The new database will be created in this group.

**Output type(s):**

- Group  
min: 1 ; max: 1

**Can be empty?** NO

**Tool-Specific Settings**

**Settings:**

- **Reference Sequence**  
Select the genome assembly the database will use.
- **Database Name**  
The name of the database to be created/added. [REQUIRED]
- **Description**  
A line or two describing the database. [OPTIONAL]
- **Species**  
Automatically selected based on Reference Sequence. [OPTIONAL]
- **Version**  
Automatically selected based on Reference Sequence. [OPTIONAL]
- **Submit**  
Once you've reviewed the name of the new database to be created.

Bioinformatics Research Laboratory  
of Medicine.

available free for academic use.

HGSC  
HUMAN GENOME SEQUENCING CENTER

A grey background (not green) means that the tool is not active. Clicking a non-active tool displays the help text that includes instruction for how to activate the tool.

To create a database, you need to drag a **Group** into "Output Targets".



# Steps for Creating a Database

**Step I** - Drag **your** group from “Data Selector” into “Output Targets”. GenboreeUser\_group and GenboreeUser\_database is placeholder for your group and databases

**Step II** - Click “Data” => “Databases” => “Create Database”. “Create Database” tool **IS** active, since it requires a “Group” to be in “Output Targets”. Select “Create Database” for tool settings.

**Step III** - Select “Template: Human (hg19)”

**Step IV** - Type database name (i.e. “GenboreeUser\_database” and click “Submit”





# Steps for Creating a Project page

The screenshot shows the GENBOREE web interface. At the top, there is a navigation bar with tabs for 'System/Network', 'Data', 'QC and Pre-processing', 'Genome', and 'Transcript'. Below this is a 'Data Selector' panel on the left with a 'Refresh' button and a tree view of groups. The 'GenboreeUser\_group' is selected. A 'Create Project' button is highlighted in the 'Data Selector' tree. A red box highlights the 'Create Project' button and the 'GenboreeUser\_group' in the tree.



**Step II** - “Create Project” tool **IS** active, since it requires a “Group” to be in “Output Targets”. Select “Create Project” for tool settings.

The screenshot shows the 'Create Project' tool settings page. It has a 'Tool Overview' section with a 'Target Group' field set to 'GenboreeUser\_group'. Below this is a 'Settings' section with fields for 'Project Name', 'Project Title' (containing 'ASHG workshop'), and 'Project Description' (containing 'Result output from ASHG Wks'). There is a 'Unique Name' checkbox which is checked. At the bottom are 'Submit' and 'Cancel' buttons.

**Step I** - Drag **your** group from “Data Selector” into “Output Targets”.

**Step IV** - Select “Refresh” in “Data Selector” to view your newly created database and project page in your group.

**Step III** - Type Project Name, Title, and Desc. and click “Submit”.

**NOTE:** Project name has to be unique for all Genboree users, so you could do something like ‘Use\_case\_18’ + your initials (i.e. “Use\_case\_18-abc”)

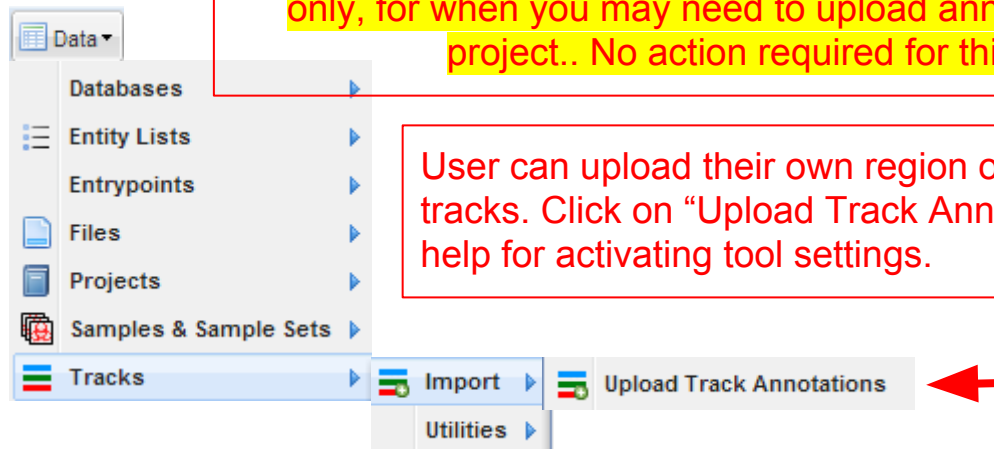
# Description of Regions of Interest (ROIs) Tracks

Source of ROIs that will be used in this analysis:

- **Track “GeneProteinCoding\_promoter”**: This track contains promoters of protein coding genes that were defined using Gencode V10 annotations ([www.gencodegenes.org/](http://www.gencodegenes.org/)), with transcription start sites (TSS) +/- 1500bp. The track contains 20,007 promoters from human genome build Hg19.
- **Track “GeneLincRNA\_promoter”**: This track contains promoters of lincRNAs that were defined using Gencode V10 annotations, with transcription start site (TSS) +/- 1500bp. the track contains 5,484 promoters from human genome build Hg19.
- **Track “ChromHMM:Enhancers”**: This track contains enhancers obtained from Manolis Kellis. Enhancers here are defined by ChromHMM using the NIH Roadmap Consortium data ([www.epigenomeatlas.org](http://www.epigenomeatlas.org)).<sup>1</sup>

Please note: Upload of annotations is provided for your information only, for when you may need to upload annotations for your own project.. No action required for this use case.

Upload your own  
Track Annotations:



User can upload their own region of interests tracks. Click on “Upload Track Annotations” to see help for activating tool settings.

1. Ernst, J. & Kellis, M. “ChromHMM: automating chromatin-state discovery and characterization”. *Nat. Methods* **9**, 215–216 (2012).

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

Welcome to the Genboree Workbench! [Getting Started]

**Step I - Drag "Release 9 Repository" database from "Epigenomics Roadmap Repository" to "Input Data"**

**Step II - Select "View Track Grid"**

Attribute	Value
Group	Epigenomics Roadmap Repository
Role	public
Name	Release 9 Repository
Description	Release 9 Repository
Species	Homo sapiens

**Input Data**

- Release 9 Repository

**Output Targets**

Refresh Data Filter: Select a filter...

- genboree.org
  - Atlas Tools Access
  - BRL AUTO TEST
  - EDACC
  - Epigenome Inform
  - Epigenome Inform
  - Epigenome Tools
  - Epigenomics Roadmap Repository**
    - Databases
      - Data Freeze 1 - Full Repo
      - Data Freeze 2 Repository
      - Release 5 Repository
      - Release 6 Repository
      - Release 7 Repository
      - Release 8 Repository
      - Release 9 Repository
    - Projects
      - GenboreeUser\_group
      - GMT\_Tutorial
      - JonathanMill\_Lab
      - Public
      - ROI Repository

# Select how you want the tracks displayed in the “View Track Grid” tool.

The screenshot shows the 'View Track Grid' tool settings window. The window title is 'Tool Settings' and the main title is 'View Track Grid'. There is a help icon (?) in the top right corner.

**Tool Overview**

**Databases with tracks of interest:**

Database: *Release 9 Repository*      Group: *Epigenomics Roadmap Repository*

**Settings**

X-axis attribute:  (dropdown arrow) ← Step III - Select “eaAssay Type”

Y-axis attribute:  (dropdown arrow) ← Step IV - Select “eaSample Type”

Page Title:

Grid Title:

X Label:

Y Label:

**Advanced Settings:**



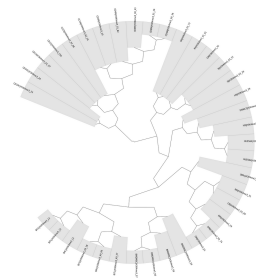
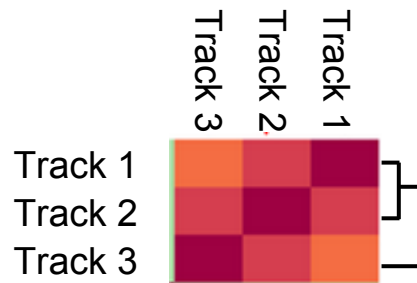
# Methodology Overview

H3K4me3 Atlas signal projected over 20,007 protein coding gene promoters (Gencode defined)

H3K4me1 Atlas signal projected over 5,484 lincRNA gene promoters (Gencode defined)

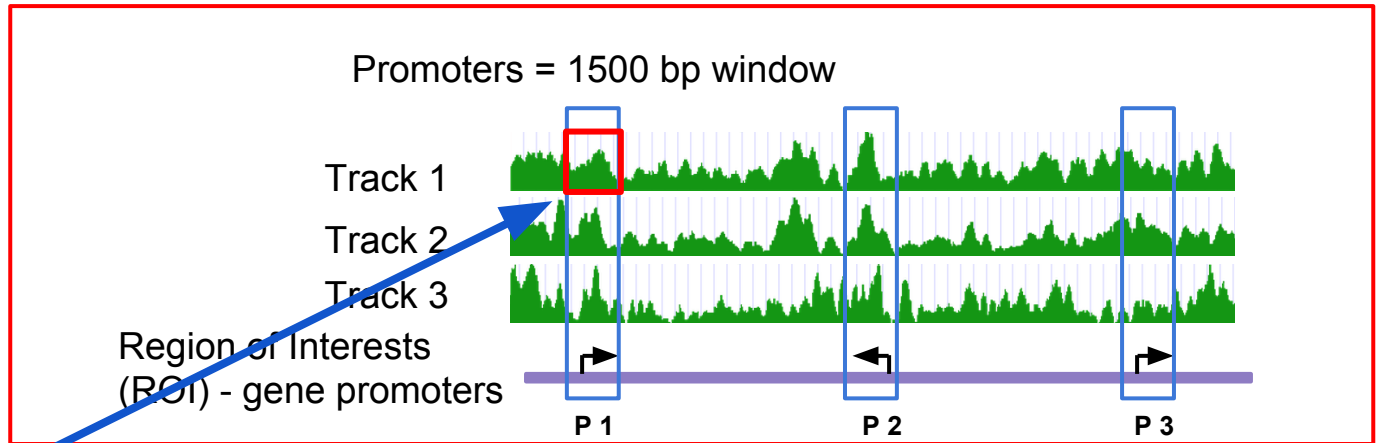


**Cluster analysis (Pearson Correlation)  
and display as Heatmap and Newick file**



# Methodology: Clustering/Heatmap

## I. Data Selection



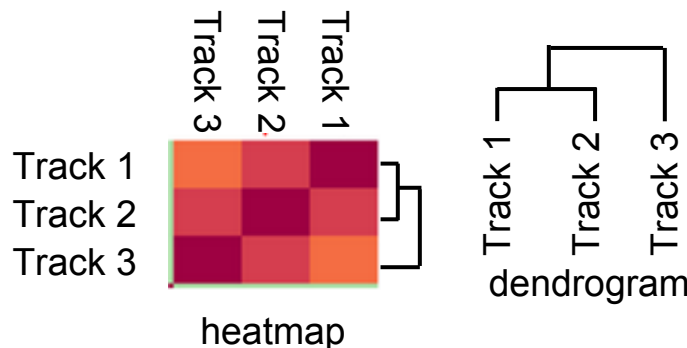
II. Signal processing-  
Heatmap Tool calculates  
average signal for each  
track and ROI (promoter)  
combination

	Track 1	Track 2	Track 3
P 1	0.8	0.7	0.3
P 2	0.7	0.6	0.7
P 3	0.6	0.8	0.2

Data matrix (3 x 3 shown here) with average signal is normalized and a correlation matrix generated. Correlation matrix is used to calculate distance measure and hierarchical clustering to group samples based on similarity to plot dendrogram.

## III. Visualization of results

Correlation values are dynamically scaled and represented in heatmap.



Note: Heatmap and dendrogram are shown as result. Genboree generates dendrogram separately, since if the output contains many tracks, row/column labels may be hard to visualize in heatmap.



## Rationale for Selection of H3K4me3 Histone Mark

In this use case, we will use H3K4me3 signal tracks of various cell types and tissues profiled in the NIH Roadmap Epigenomics project. The H3K4me3 mark is associated with active promoters. We sought to ask how the various cell types/tissues in the Human Epigenome Atlas cluster using a H3K4me3-promoter combination. Here, the promoters are defined as the transcription start site (TSS) +/- 1500 base pairs using Gencode annotations.

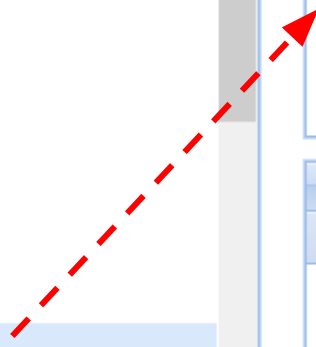
Welcome to the Genboree Workbench! [\[Getting Started\]](#)

### Data Selector

Refresh Data Filter: Select a filter...

- Epigenomics Roadmap Repository
- GenboreeUser\_group
- GMT\_Tutorial
- JonathanMill\_Lab
- Public
- ROI Repository**
  - Databases
    - ROI Repository - hg18
    - ROI Repository - hg19**
      - All Annotations in Database
      - Tracks**
        - Class: Affymetrix
        - Class: Agilent
        - Class: ENCODE
        - Class: ENCODE - T.f. Binding Sites
        - Class: Enhancer
        - Class: GC
        - Class: Gencode**
          - Gene:LincRNA\_promoter
          - Gene:ProteinCoding\_promoter**
        - Class: Gene
        - Class: Gene Model

**Step1 - Drag Promoter Track to Input Data**



### Details

Attribute	Value
Group	ROI Repository
Database	ROI Repository - hg19
Name	Gene:ProteinCoding_promoter
Description	
BigBed	none

### Input Data

↑ ↓ ✕ 🎨

- Gene:ProteinCoding\_promoter**

### Output Targets

↑ ↓ ✕ 🎨

Welcome to the [...]

**Step 2 - Drag**  
"release9\_H3K4me3\_subset"  
tracks to Input Data

We are providing the heatmap tool with two types of tracks: data tracks and annotation tracks. The signal intensity from the "Release9\_H3K4me3\_subset", 37 data tracks, are projected over the "Gene: ProteinCoding\_promoter" annotation tracks. The Pearson correlation metric is calculated and the similarity matrix is presented visually as a heatmap.

Data Selector

Refresh

- Epigenome ToolSet Demo Input Data
  - Databases
    - ASHG 2013 Demo
      - All Annotations in Database
      - Tracks
      - Lists & Selections
        - Lists of Tracks
          - HL60\_Tracks
          - HSC\_Tracks
          - Immune\_HL60\_H3K4me1
          - Myeloid\_Tracks
          - release9\_H3K27me3
          - release9\_H3K36me3
          - release9\_H3K4me1
          - release9\_H3K4me1\_subset
          - release9\_H3K4me3
          - release9\_H3K4me3\_subset
          - release9\_H3K9me3

- Sample Sets
- Samples
- Files
- Queries

Input Data

↑ ↓ ✕ 🏠

- Gene:ProteinCoding\_promoter
- release9\_H3K4me3\_subset

Output Targets

↑ ↓ ✕ 🏠

Welcome to the Genboree Workbench! [Getting Started]

### Data Selector

Refresh

- Peak Calling Dem
- RNA-Seq Tool De
- Projects
- Epigenomics Roadmap
- GenboreeUser\_group
  - Databases
  - GenboreeUser\_d
  - Projects
    - GenboreeUser\_project
    - Use\_Case\_01\_GU
    - Use\_Case\_02\_GU
    - Use\_Case\_05\_GU
    - Use\_Case\_07\_GU
    - Use\_Case\_09\_GU
    - Use\_Case\_12\_GU
    - Use\_Case\_13\_GU
    - Use\_Case\_14\_GU
    - Use\_Case\_18\_GU
    - Use\_Case\_19\_GU
    - Use\_Case\_20\_GU
    - Use\_Case\_21\_GU
- GMT Tutorial

**Step 3** - Drag YOUR Database and Project into Output Targets. This is the database and project where your results will be deposited. The Epigenome menu bar will turn green. Click on "Compute Similarity Matrix (heatmap)".

- Random Forest
- QIIME
- QC
- Search for Similar Signals by Correlation
- Analyze Signals
- Compute Similarity Matrix (heatmap)**
- Create Track Lists from Newick Tree
- Slice Epigenomic Data
- Analyze Signals in the Context of Epigenome Atlas

Gene:ProteinCoding\_promoter  
release9\_H3K4me3\_subset

### Output Targets

GenboreeUser\_database  
Use\_Case\_18\_GU

Check that the “Input Files Directory and “Output Database” and “Project” are correct (based on what you named them). The default parameters will be used unless noted otherwise.

A default “Analysis Name” is generated by Genboree. It is recommended that all text and the time stamp be kept, and that you append some unique text to the beginning to help you distinguish different jobs run from the same tool.

**Step 4** - A default “Analysis Name” is generated by Genboree. It is recommended that all text and the time stamp be kept, and that you append some unique text to the beginning to help you distinguish different jobs run from the same tool.

**Step 5** - Select “Pearson’s Correlation ” as distance function and Average as hierarchical clustering function

**Step 6** - Expand “No Data Regions” and select “If BOTH tracks have no data for that region”. Click “Submit”.

Tool Overview

**Input Entity Lists(s)/ROI-Track:**

Items: Gene:ProteinCoding\_promoter (Track)  
release9\_H3K4me3\_subset (Track Entity List)

**Output Database/Project:**

Database/Projects Of Interest: GenboreeUser\_database Group: GenboreeUser\_group  
Use\_Case\_18\_GU Group: GenboreeUser\_group

labelLeave data matrix unchanged

**Epigenomic Experiment Heatmap Tool**

Analysis Name: EpigenomeExpHeatmap2013-1

Normalization: Quantile

Aggregating Function: Avg

Distance Function: Pearson's Correlation

Hierarchical Clustering Function: Average

Key:

Key Size: 0.75

Height: 8

Width: 10

Trace: None

Density: Histogram

Dendrograms to display on heatmap: Both

**No Data Regions**

No Data Value: 0

Remove No Data Regions?

If EITHER track has no data for that region:

If BOTH tracks have no data for that region:

Submit Cancel

You will see the message below upon successful submission of your heatmap job:

Job Submission Status

**Compute Similarity Matrix (heatmap) BETA**

Job Id: wbJob-epigenomicsHeatmap-WZhw3D-1492

Your job has been successfully submitted. You will be notified by email when your job has completed.

You can track the progress of your job using the 'Job Summary' tool under System/Network/Jobs.

If you have questions, please contact [genboree\\_admin@genboree.org](mailto:genboree_admin@genboree.org) for assistance.

OK

Running time of the job will vary based on the data tracks you choose for analysis. Job you submitted will take around 10 mins to complete once it starts running. You can status of the job through Job Summary

The screenshot displays the GENBOREE web interface. At the top, the GENBOREE logo is on the left, and the BCM (Baylor College of Medicine) logo is on the right. Below the logo is a navigation bar with tabs for System/Network, Data, QC and Pre-processing, Genome, Transcriptome, Cistrome, Epigenome, Metagenome, Visualization, and Help. A left sidebar contains a menu with options: User Profile, Groups, Hosts, Jobs, and Request Feature. The 'Jobs' menu is expanded, showing 'Job Summary' highlighted. A red box highlights the 'Job Summary' menu item, with an arrow pointing to a 'Job Summary' window. This window has a 'Settings' tab and contains the following fields: Start Date (2013/7/18), End Date (YYYY/MM/DD), Sort Order (Newest first), and Group By (None). Below these fields are 'Advanced Settings' and two buttons: 'Generate Report' and 'Cancel'. A red box highlights the 'Generate Report' button, with an arrow pointing to a text box that says 'Select "Generate Report" to see Job Summary'. The main content area of the interface includes a 'Details' table, 'Input Data' section with icons for up/down arrows, a red X, and a pencil, and an 'Output Targets' section with similar icons.

**GENBOREE**

BCM  
Baylor College of Medicine

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

Job Summary [Getting Started]

Job Summary

Tool Settings

**Job Summary**

Tool Overview

Settings

Start Date 2013/7/18

End Date YYYY/MM/DD

Sort Order Newest first

Group By None

Advanced Settings:

Generate Report Cancel

Select "Generate Report" to see Job Summary

Details

Attribute	Value
-----------	-------

Input Data

Output Targets

BRL

Genboree is built & maintained by the **Bioinformatics Research Laboratory** at **Baylor College of Medicine**.

Genboree is a hosted service. Code is available **free for academic use**.

HGSC  
HUMAN GENOME SEQUENCING CENTER

# You will get the following e-mail message when your job is completed

```
Hello`

Your Compute Similarity Matrix (heatmap) job completed successfully.

Job Summary:
JobID - wbJob-epigenomicsHeatmap-e3CshO-5730
Analysis Name - EpigenomeExpHeatmap2013-10-09-00:42:01
Inputs:
1. Entitylist -
2. Trk -
3. Entitylist -
Outputs:
1. Db -
2. Prj -
Settings:
analysisName - EpigenomeExpHeatmap2013-10-09-00:42:01
dendograms - both
density - histogram
distfun - cor
hclustfun - av
height - 8
key - TRUE
keySize - 0.75
naGroup - 10
normalization - c
removeNoDataReg
replaceNAValue
spanAggFunction
trace - none
width - 10

- The Genboree Team

Result File Location in the Genboree Workbench:
http://genboree.org/java-bin/project.jsp?projectName=Roadmap%20Epigenome%20Data%20Analysis
```

**Clicking on the link will take you to the project page containing your results**

2013/10/17: Genboree User ran Epigenomic Heatmap Tool (EpigenomeExpHeatmap2013-10-17-14 51 14) and the results are available at the link below.

- **Study Name:** EpigenomeExpHeatmap2013-10-17-14 51 14
- **Link to results**

**Click on the "Link to results" in your Project page**

## Project Page



# Link to Results on Your Project Page

## Table of Content: Epigenomic HeatMap

Study Name: Promoter H3K4me3 2013-10-13-16 37 22

User: Genboree User

Date: 2013/10/13 19:37 CDT

### Epigenomic HeatMap Plots

[Heatmap](#)  
[Correlation plot](#)

Click on the image links for any of the Newick Tree Visualizations to see circular tree of clustered epigenomes profiled in the NIH Roadmap Epigenome Project or the Heatmap

### Newick Tree Visualizations

#### Equal Branch Lengths

Rows [\[PNG\]](#) [\[SVG\]](#)

Columns [\[PNG\]](#) [\[SVG\]](#)

#### Scaled Branch Lengths

Rows [\[PNG\]](#) [\[SVG\]](#)

Columns [\[PNG\]](#) [\[SVG\]](#)

#### Natural Log Scaled Branch Lengths

Rows [\[PNG\]](#) [\[SVG\]](#)

Columns [\[PNG\]](#) [\[SVG\]](#)

#### Log10 Scaled Branch Lengths

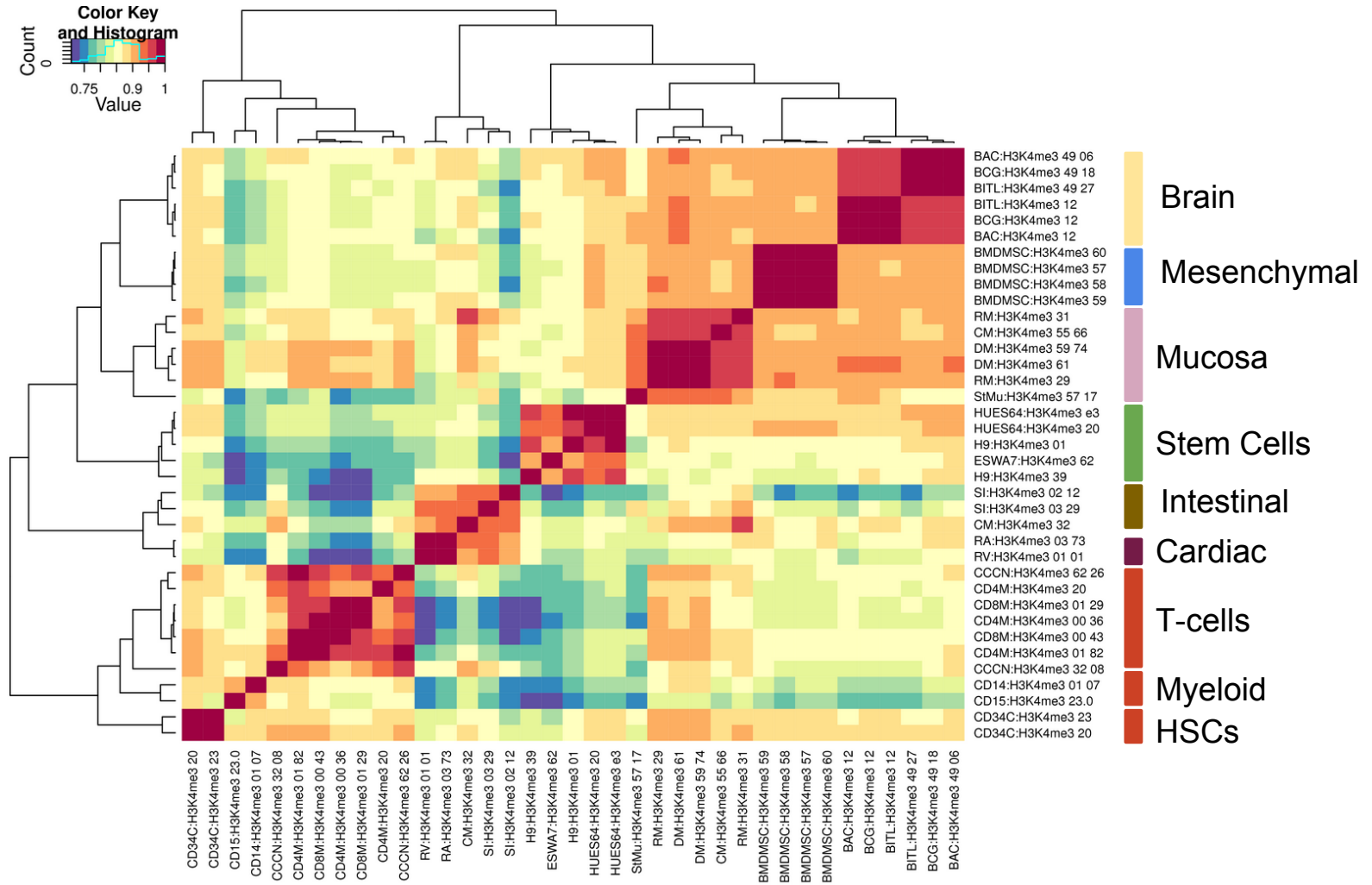
Rows [\[PNG\]](#) [\[SVG\]](#)

Columns [\[PNG\]](#) [\[SVG\]](#)

“Equal Branch Lengths”: Rows and Columns of the dendrogram will be the same, as pairwise comparison of tracks was performed. However, user can choose comparison of different track sets, and this will generate separate Rows and Columns tree.

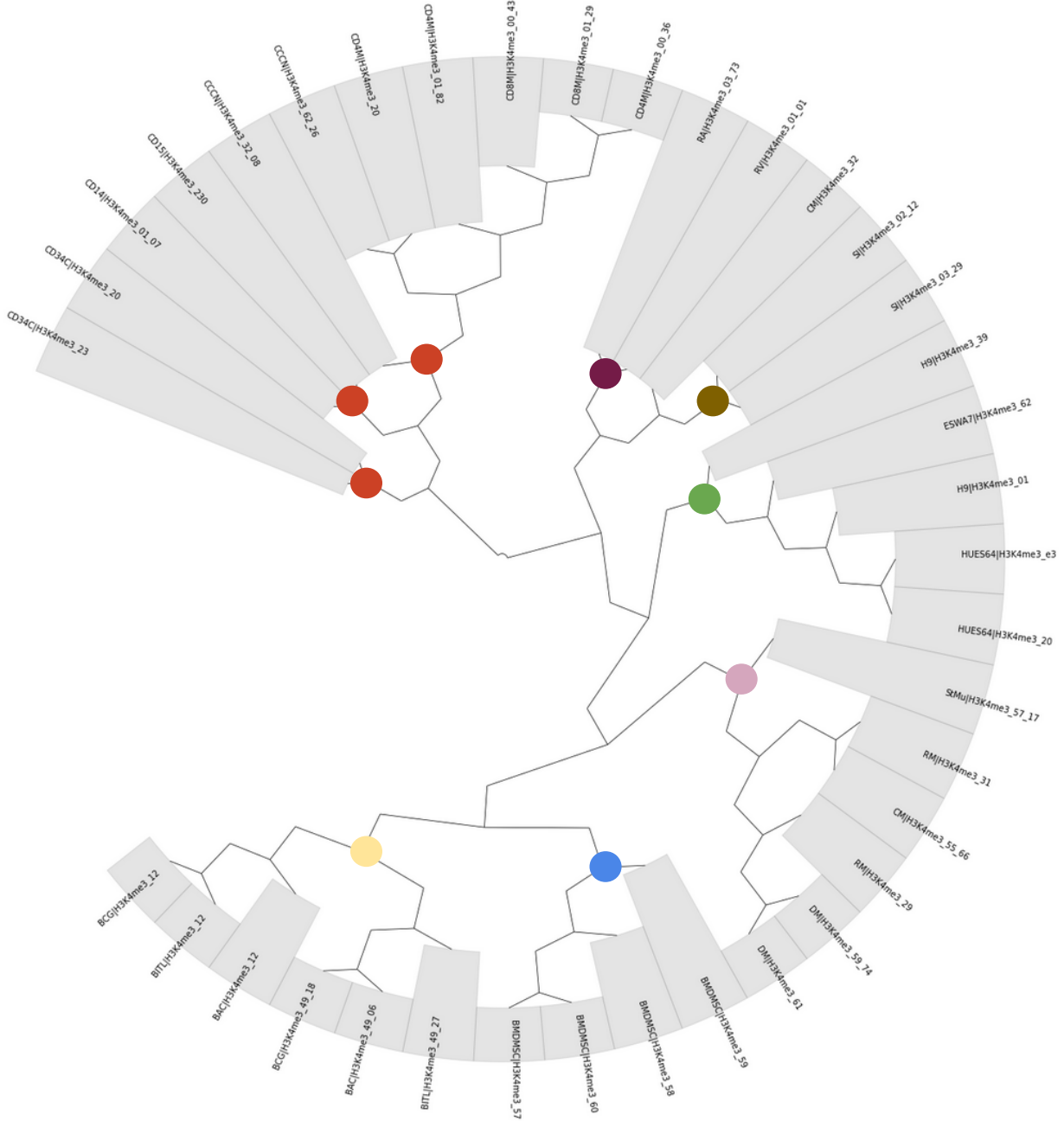
“Scaled, Natural Log Scaled, and Log10 Scaled Branch Length”: Different scaling of branch lengths are used here, since in some scenarios scaled branch length may provide better visualization

# H3K4me3 signal at protein-coding promoter region can distinguish different cell types/tissues of origin



# Clustering Results Presented in a Circular Dendrogram

- Brain
- Mesenchymal
- Mucosa
- Stem Cells
- Intestinal
- Cardiac
- T-cells
- Myeloid
- HSCs



## **Rationale for Selection of H3K4me1 Histone Mark**

Next, we will like to see how these different cell types/tissues cluster over different histone modification signal and region of interests. Here we use H3K4me1 signal tracks profiled in the NIH Roadmap Epigenome project and lincRNA promoter. H3K4me1 mark is associated with active enhancers. The lincRNA promoters were defined using Gencode annotations.

Welcome to the Genboree Workbench! [Getting Started]

**Data Selector**

Refresh Data Filter: Select a filter...

- Epigenome Informatics Workshop (May 2012)
  - Epigenome ToolSet Demo Input Data**
    - Databases
      - ASHG 2013 Demo
        - All Annotations in Database
        - Tracks
        - Lists & Selections
          - Lists of Tracks**
            - HL60\_Tracks
            - HSC\_Tracks
            - Immune\_HL60\_H3K4me1
            - Myeloid\_Tracks
            - release9\_H3K27me3
            - release9\_H3K36me3
            - release9\_H3K4me1
            - release9\_H3K4me1\_subset**
            - release9\_H3K4me3
            - release9\_H3K4me3\_subset
            - release9\_H3K9me3

**Details**

Attribute

Track 1

Track 2

Track 3

Track 4

Track 5

**Step 9**

- Random Forest
- QIIME
- QC
- Search for Similar Signals by Correlation
- Analyze Signals
- Compute Similarity Matrix (heatmap)**
- Create Track Lists from Newick Tree
- Slice Epigenomic Data
- Analyze Signals in the Context of Epigenome Atlas

**Input Data**

Gene:LincRNA\_promoter

release9\_H3K4me1\_subset

**Output Targets**

GenboreeUser\_database

Use\_Case\_18\_GU

**Step 7** - Remove the previous track entity list and ROI from "Input Data"

Drag "Gene:LincRNA\_promoter" track from ROI Repository. Drag "release9\_H3K4me1\_subset" track entity list to Input Data

**Step 8** - Drag your database and project page to Output Targets



**Tool Settings**

## Compute Similarity Matrix (heatmap) BETA

**Tool Overview**

**Input Entity Lists(s)/ROI-Track:**

Items: *Gene:LincRNA\_promoter (Track)*  
*release9\_H3K4me1\_subset (Track Entity List)*

**Output Database/Project:**

Database/Projects Of Interest: *GenboreeUser\_database* Group: *GenboreeUser\_group*  
*Use\_Case\_18\_GU* Group: *GenboreeUser\_group*

labelLeave data matrix unchanged

**Epigenomic Experiment Heatmap Tool**

Analysis Name:

Normalization:

Aggregating Function:

**Distance Function:**

**Hierarchical Clustering Function:**

Key:

Key Size:

Height:

Width:

Trace:

Density:

Dendrograms to display on heatmap:

**No Data Regions**

No Data Value:

Remove No Data Regions?

If EITHER track has no data for that region

**If BOTH tracks have no data for that region**

**Step 10** - A default "Analysis Name" is generated by Genboree. It is recommended that all text and the time stamp be kept, and that you append some unique text to the beginning to help you distinguish different jobs run from the same tool.

**Step 11** - Select Pearson's Correlation (Absolute Value) as distance function and Average as hierarchical clustering function

**Step 12** - Select to remove data if both tracks have no data for that region

**You will see the message below upon successful submission of your heatmap job:**

**Job Submission Status**

## Compute Similarity Matrix (heatmap) BETA

**Job Id:** *wbJob-epigenomicsHeatmap-N2rF7v-1479*

Your job has been successfully submitted. You will be notified by email when your job has completed.

You can track the progress of your job using the 'Job Summary' tool under System/Network/Jobs.

If you have questions, please contact [genboree\\_admin@genboree.org](mailto:genboree_admin@genboree.org) for assistance.

# You will get following e-mail message when job is completed

Hello

Your Compute Similarity Matrix (heatmap) job completed successfully.

Job Summary:  
JobID - wbJob-epigenomicsHeatmap-e3CshO-5730  
Analysis Name - EpigenomeExpHeatmap2013-10-09-00:42:01

Inputs:  
1. Entitylist -  
2. Trk -  
3. Entitylist -

Outputs:  
1. Db -  
2. Prj -

Settings:  
analysisName - EpigenomeExpHeatmap2013-10-09-00:42:01  
dendograms - both  
density - histogram  
distfun - cor  
hclustfun - ave  
height - 8  
key - TRUE  
keySize - 0.75  
naGroup - 100  
normalization - d  
removeNoDataReg  
replaceNAValue  
spanAggFunction  
trace - none  
width - 10

- The Genboree Team

Result File Location in the Genboree Workbench:  
<http://genboree.org/java-bin/project.jsp?projectName=Roadmap%20Epigenome%20Data%20Analysis>

**Clicking on the link will take you to the project page containing your results**

## Project Page

2013/10/17:

Genboree User ran Epigenomic Heatmap Tool (EpigenomeExpHeatmap2013-10-17-14 51 14) and the results are available at the link below.

- **Study Name:** EpigenomeExpHeatmap2013-10-17-14 51 14
- **Link to results**

**Click on the "Link to results" in your Project page**



## Table of Content: Epigenomic HeatMap

**Study Name:** lincRNA H3K4me1 2013-10-13-17 00 03

**User:** Genboree User

**Date:** 2013/10/13 18:10 CDT

---

### Epigenomic HeatMap Plots

[Heatmap](#)

[Correlation plot](#)

---

### Newick Tree Visualizations

#### Equal Branch Lengths

Rows [PNG](#) [SVG](#)

Columns [PNG](#) [SVG](#)

#### Scaled Branch Lengths

Rows [PNG](#) [SVG](#)

Columns [PNG](#) [SVG](#)

#### Natural Log Scaled Branch Lengths

Rows [PNG](#) [SVG](#)

Columns [PNG](#) [SVG](#)

#### Log10 Scaled Branch Lengths

Rows [PNG](#) [SVG](#)

Columns [PNG](#) [SVG](#)

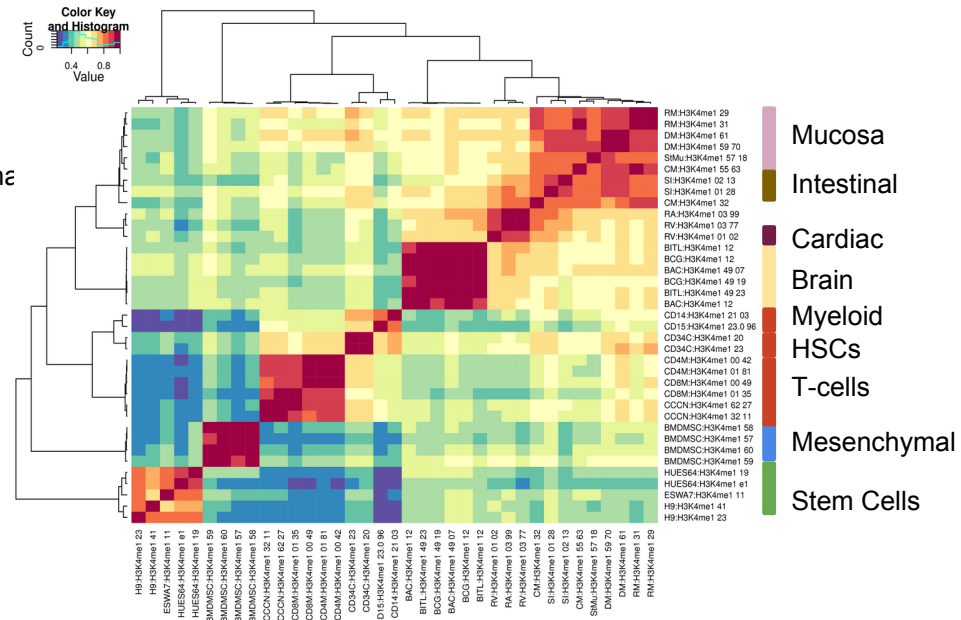
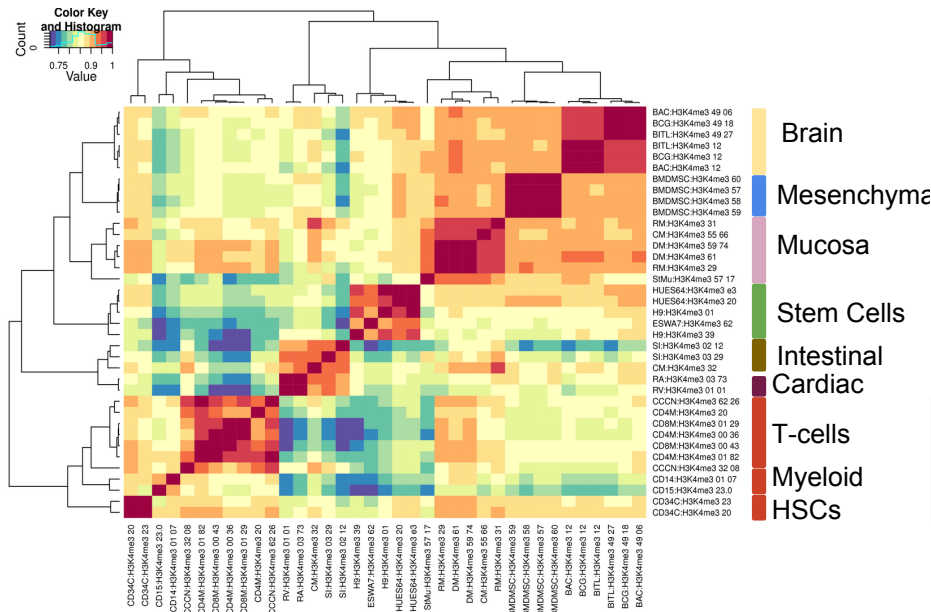
Click here to see circular tree of clustered epigenomes profiled in the NIH Roadmap Epigenome Project



# Summary of Results

H3K4me3 signal over protein coding gene promoters on the NIH Roadmap Epigenome data

H3K4me1 signal over lincRNA gene promoters on the NIH Roadmap Epigenome data



Cluster analysis of the epigenomes profiled in the NIH Roadmap Epigenome project suggests that lincRNAs play an important role in maintaining cellular identity.

If you are interested in learning about how Genboree tools can help identify regions that are specific for given lineage (Use case 19 supplementary slides) and how you can study regions that are undergoing epigenomic changes during cell differentiation, go on to use case 19.

Help us improve Genboree. Please provide a comment or request feature.

The screenshot displays the Genboree Workbench interface. At the top, the 'GENBOREE' logo is on the left, and the 'BCM Baylor College of Medicine' logo is on the right. Below the logo is a navigation bar with tabs for 'System/Network', 'Data', 'QC and Pre-processing', 'Genome', 'Transcriptome', 'Cistrome', 'Epigenome', 'Metagenome', 'Visualization', and 'Help'. A sidebar on the left contains a tree view with categories like 'User Profile', 'Groups', 'Hosts', 'Jobs', and 'Request Feature' (which is highlighted with a red box). The main content area shows a 'Genboree Workbench! [Getting Started]' header and a 'Data Filter' dropdown. A 'Details' table is visible on the right. A 'Request Feature' dialog box is open in the foreground, containing a 'Settings' section with input fields for 'User Name' (Genboree User), 'User Email' (andrewj@bcm.edu), and a 'Message' text area. 'Submit' and 'Cancel' buttons are at the bottom of the dialog. At the bottom of the page, there is a footer with the BRL logo, text stating 'Genboree is built & maintained by the Bioinformatics Research Laboratory at Baylor College of Medicine.', and the HGSC logo.

Genboree is built & maintained by the Bioinformatics Research Laboratory at Baylor College of Medicine.

Genboree is a hosted service. Code is available free for academic use.