

Use Case 19: Identifying regions that are undergoing epigenomic transitions during cell differentiation using the NIH Roadmap Epigenome data

---

American Society of Human Genetics  
Boston, MA 2013

Presented by the  
Bioinformatics Research Laboratory

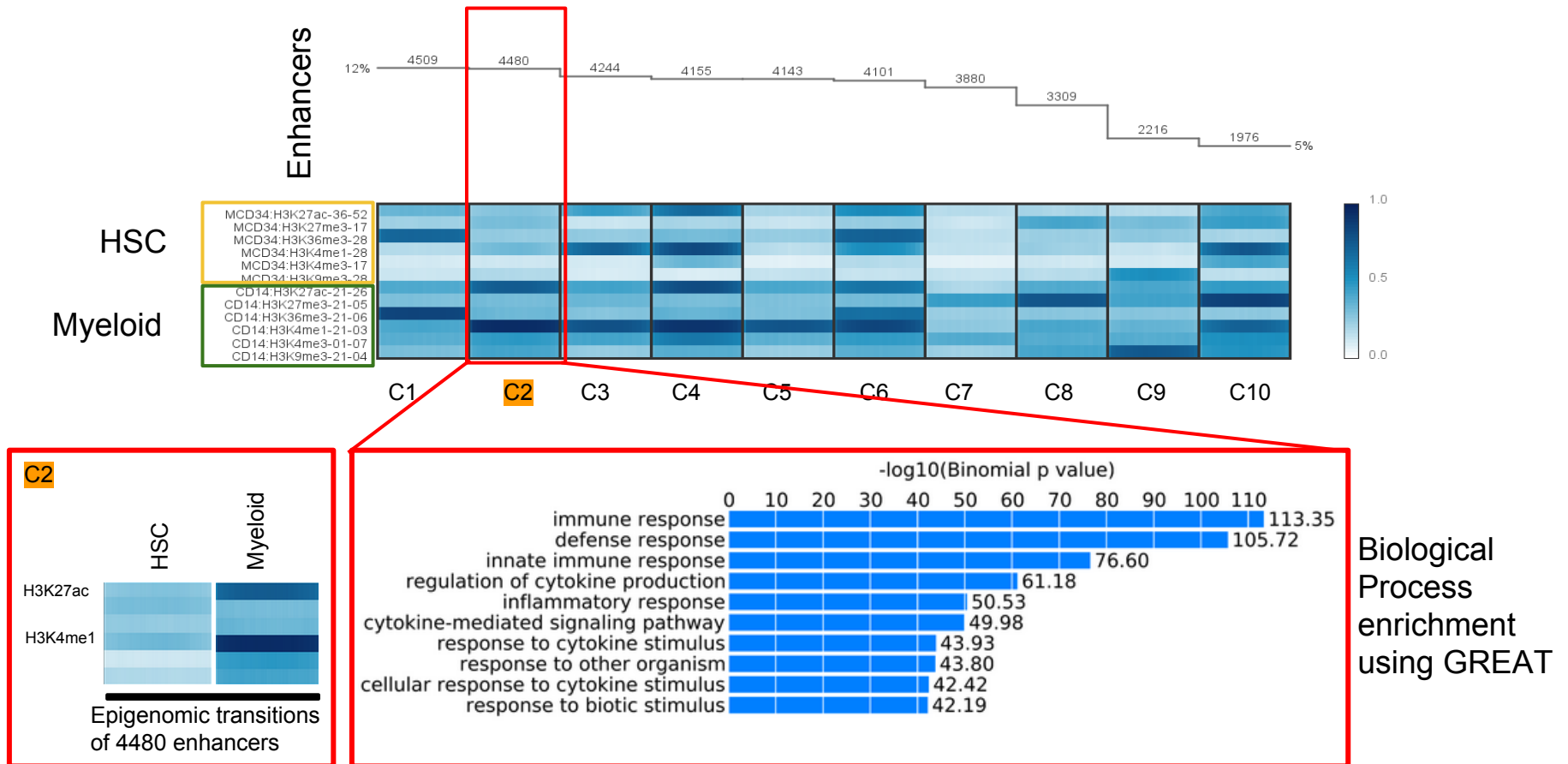
Baylor  
College of  
Medicine

## Summary of Use Case 19

**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 generated data outside of the NIH Epigenome Roadmap project. Computational analysis of existing data repositories, such as the Atlas, offers the opportunity for new biological discovery and insight that complements experimental approaches. We sought to use a computational based approach with the Human Epigenome Atlas to identify enhancers regulating myeloid-specific differentiation and the corresponding pathways that are regulated by the enhancers.

**Results:** Computational analysis of epigenomic marks and transcription factor binding patterns during cellular differentiation are highly coordinated.

# Summary of results



Cluster of enhancers that are undergoing epigenomic changes during myeloid cell differentiation from CD34 (HSC lineage) to CD14 (Myeloid lineage) were identified using Spark. Functional significance of biological process assessed through Spark do indicate importance of these regions during myeloid cell differentiation.

# 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 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: System/Network, Data, QC and Pre-processing, Genome, Transcriptome, Cistrome, Epigenome, Metagenome, Visualization, and Help. The 'System/Network' and 'Help' tabs are highlighted in green. Below the navigation bar, a welcome message reads 'Welcome to the Genboree Workbench! [Getting Started]'. The main content area is divided into three sections: 'Data Selector', 'Details', and 'Input Data' and 'Output Targets'.

**Data Selector:** Various Data Types (tracks, files, and ROIs (region of interests), etc)

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

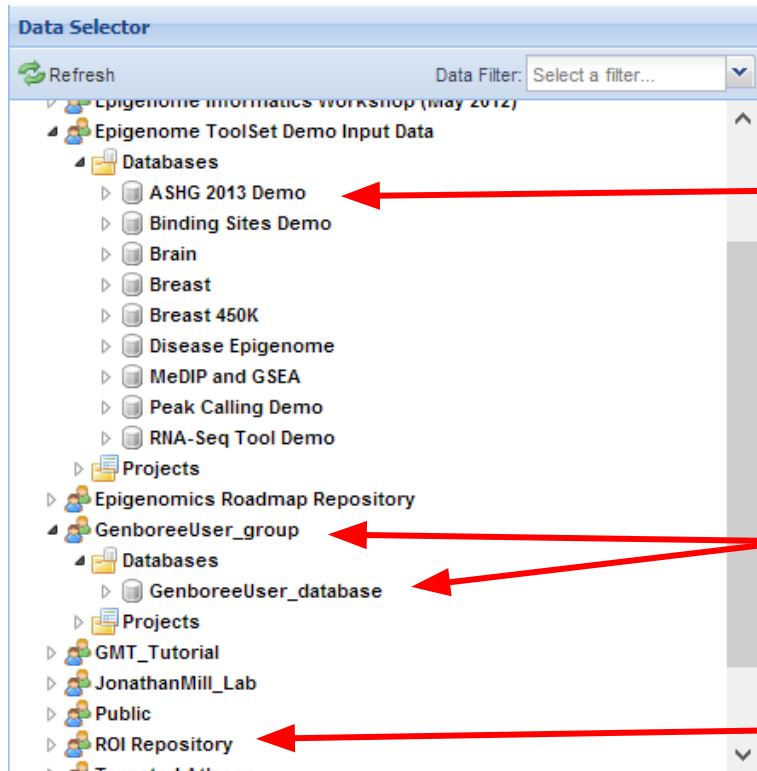
**Input Data:** Tells the tool to use this input data/file

**Output Targets:** Tells the tool where to deposit results

**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 used as a generic placeholder name for any Genboree user group name, thus “*GenboreeUser*” is **YOU**.
- Similarly, “*GenboreeUser\_database*” is used as a placeholder name for your database name. Therefore, as you go through the use cases, any place you see “*GenboreeUser\_group*” or “*GenboreeUser\_database*”, you should actually be interacting with your own group or database.



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

**GENBOREE**

BCM  
Baylor College of Medicine

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

Welcome to

Data Selector

Databases  
Entity Lists  
Entrypoints

Create Database  
Rename Database  
Delete Database

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.

Help: Tool Setting

**Help: Create Database**

⚠ 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.

**Input Data**

**Output Targets**

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

Bioinformatics Research Laboratory  
Baylor College of Medicine.

is available **free for academic use**.

HGSC  
HUMAN GENOME SEQUENCING CENTER



# 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"

**Tool Settings: Create Database**

**Target Group:**

**Reference Sequence:** Template: Human (hg19)  
Template: Human 3/12 (Hg15)  
Template: Human 3/12 (Hg16)  
Template: Human chr 12(Hg15)  
Template: Human chr 12(Hg16)

**Database Name:**   
**Description:**   
**Species:**   
**Version:**

**Submit** **Cancel**

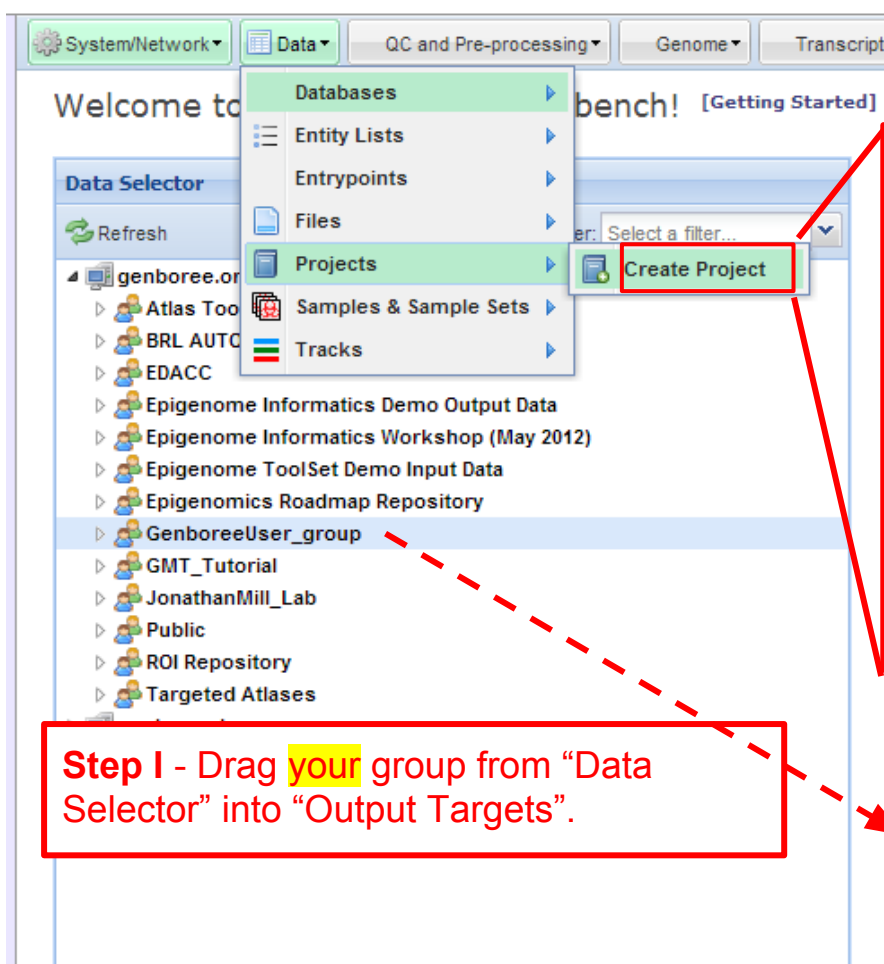
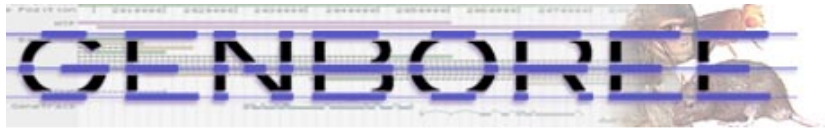
**Output Targets**

GenboreeUser\_group





# Steps for Creating a Project page



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

A screenshot of the "Create Project" form. The form has a "Tool Overview" section with a "Target Group:" field containing "GenboreeUser\_group". Below this is a "Settings" section with fields for "Project Name", "Unique Name", "Project Title", and "Project Description". The "Project Title" field contains "ASHG workshop" and the "Project Description" field contains "Result output from ASHG Wks". There are "Submit" and "Cancel" buttons at the bottom. A red box highlights the "Settings" section. A red arrow points from the "Unique Name" field to the "Output Targets" section below.

**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")

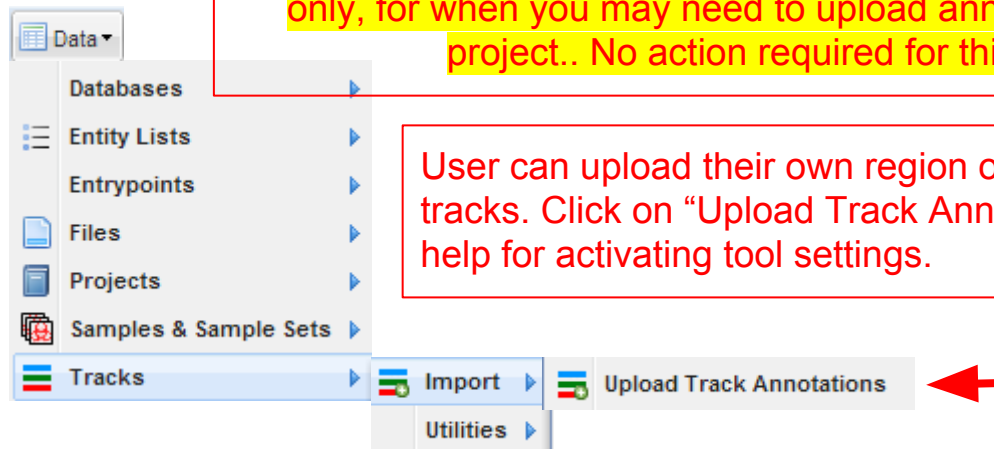
# Description of Regions of Interest (ROIs) Tracks

*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]

**Data Selector**

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

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

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

- View Track Grid
- View Sample Grid
- Tabular Annotation Viewer
- Launch UCSC Genome Browser

**Attribute**

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**

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

**Tool Settings**

## View Track Grid

**+ Tool Overview**

**Databases with tracks of interest:**

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

**Settings**

**X-axis attribute** eaAssayType ▼

**Y-axis attribute** eaSampleType ▼

**Page Title** Grid Viewer: Tracks from Relea

**Grid Title** Tracks from Release 9 Reposit

**X Label** eaAssayType

**Y Label** eaSampleType



**+ Advanced Settings:**

**Submit** **Cancel**

Step III - Select “eaAssay Type”

Step IV - Select “eaSample Type”

# Track Grid view of the data from Roadmap Epigenome Project

Releases
Informatics
Publications
Forums
Contributors

- [Data Access Policy](#)
- Data embargo period: from 04/15/2013 - 01/15/2014 or earlier as specified [here](#)
- Select cells by clicking and dragging, then use "View Selections" in the Selections menu
- Use "Save Selections" in the Selections menu to save selected (highlighted) cells in a group
- To see data authors, other metadata, and to download data, click a sample name in the filter
- Expression Array data may be downloaded [here](#)
- Human Epigenome Atlas releases are intended to be cumulative: e.g. Release 3 includes
- NOTE: Some pages may not be accessible over low bandwidth internet connections. This p

Tracks from Release 9 Repository

Filter rows:

Selections

Choose Databases

eaAssayType

eaSampleType

	Bisulfite-Seq	MeDIP-Seq	MFE-Seq	RRBS	DNAse Hypersensitivity	Digital Genomic Footprinting	mRNA-Seq	smRNA-Seq	ChIP-Seq Input	Histone H3K27me3	Histone H3K36me3	Histone H3K4me1	Histone H3K4me3	Histone H3K9ac	Histone H3K9me3	Histone H2AK5ac	Histone H2AK9ac	Histone H2A Z	Histone H2BK5ac	Histone H2BK12ac	Histone H2BK15ac	Histone H2BK20ac	Histone H2BK120ac	Histone H3K14ac	Histone H3K18ac	Histone H3K23ac	Histone H3K23me2	Histone H3K27ac	Histone H3K4ac	Histone H3K4me2	Histone H3K56ac	Histone H3K79me1	Histone H3K79me2	Histone H3K9me1	
Adipose Derived Mesenchymal Stem Cells									3	2	3	3	3	5	3																				
Adipose Nuclei									5	5	5	5	5	5	5												1								
Adipose Tissue	1					3		1																			1								
Adrenal Gland	1					2			2	4	2	2	1		1													2							
Adult Kidney				2					2		2	2	2	2	2													2							
Adult Liver	1					2			4	4	4	4	5	2	4												2								
Aorta	1					2			2	2	2	2	2		1												2								
Bladder						1		2			1	1															1								
Bone Marrow Derived Mesenchymal Stem Cell Cultured Cells				2						1	1	1	1	1	1												4								
Bone Marrow Derived Mesenchymal Stem Cells									4	3	3	3	3	3	3																				
Brain Angular Gyrus				1					2	1	2	2	2	1	2												2								
Brain Anterior Caudate				2					2	2	2	2	2	1	2												2								
Brain Cerebellum						1																													
Brain Cingulate Gyrus				1					2	1	2	2	2	1	2												2								
Brain Germinal Matrix	1	2				1	3	1	2	2	2	2	2		2																				
Brain Hippocampus Middle	2					2		3	3	3	3	3	3	1	3												3								

A Track/experiment or group of tracks (track-entity lists) can be selected and saved in your database by selecting "Selections" &gt; "Save Selections". However, for this use case track-entity lists have already been generated for you.

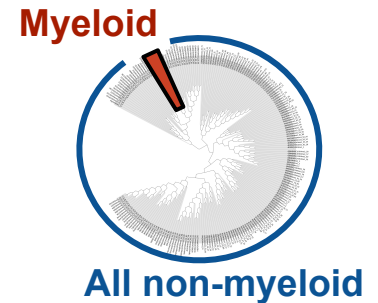
# Methodology Overview

Use Case 18

**Clustering/Heatmap:** select experimental tracks from the Human Epigenome Atlas to find myeloid cell lineage consisting of CD14 and CD15 cell types

Use Case 19  
Supplemental Slides

**LIMMA:** to find enhancer regions with differentially modified histone signals between two groups of data tracks - Myeloid vs non-myeloid



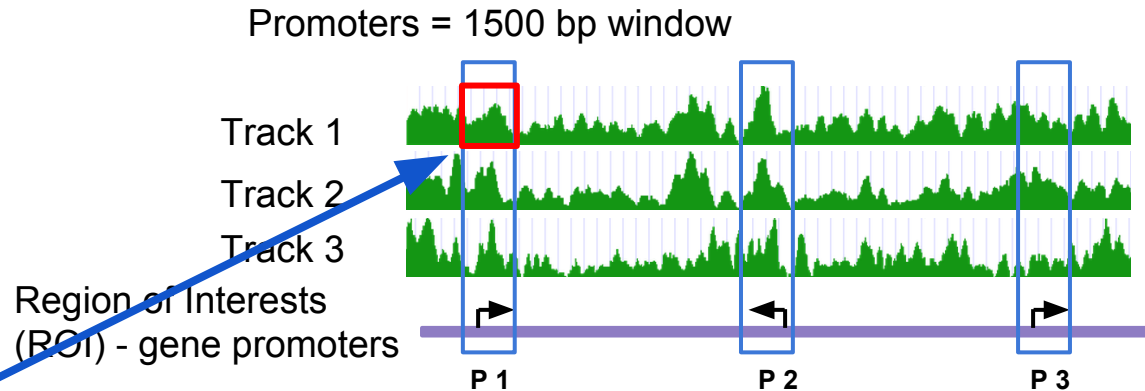
**Spark:** visualize epigenomic changes happening at LIMMA defined enhancer regions from hematopoietic stem cells to myeloid cell type by clustering regions with similar “epigenomic footprint” transitions

**GREAT:** assesses functional significance of cis-regulatory regions.

**HOMER:** de novo motif discovery.

# Methodology: Clustering/Heatmap

## I. Data Selection



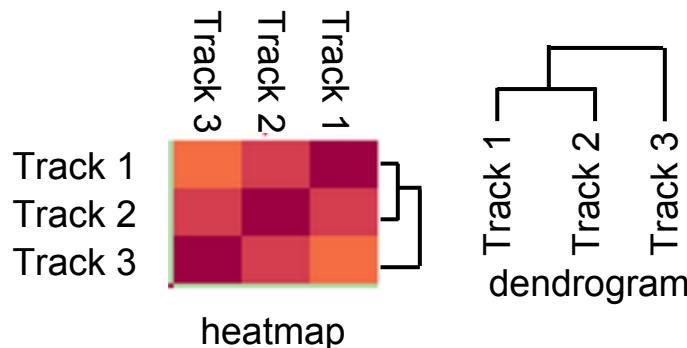
## II. Signal processing- Heatmap Tool calculates average signal for each track and ROI (promoter) combination

	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.

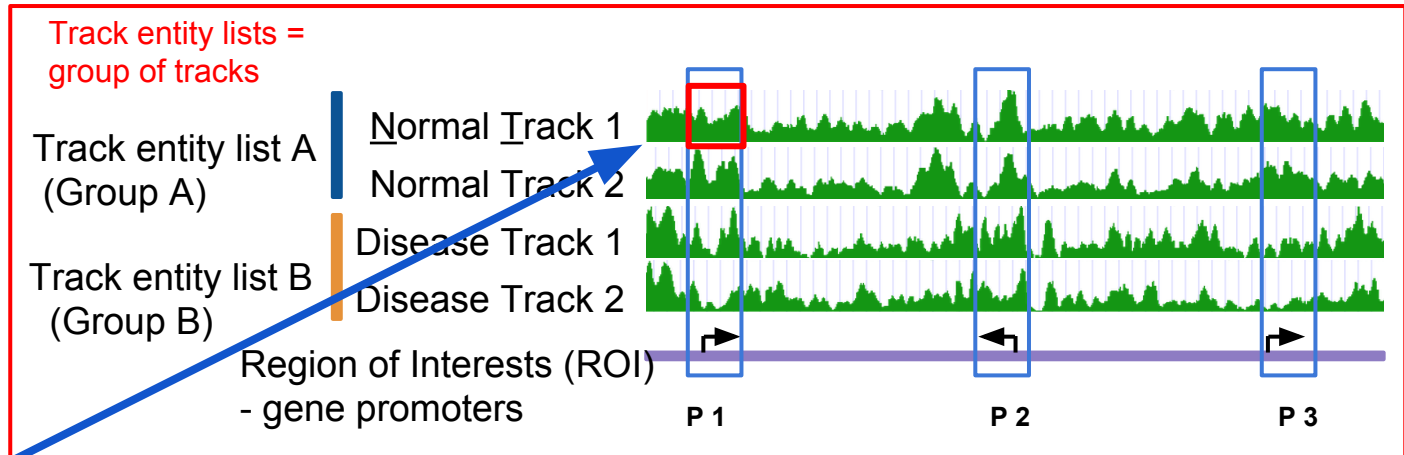


Note: Heatmap and dendrogram are shown as result. Genboree generates dendrogram separately, since if the output contains many tracks, row/column labels may be hard to visualize in heatmap.



# Methodology: LIMMA (Linear Model for Microarray Analysis)

## I. Data Selection



## II. Signal processing- Tool calculates average signal for each ROI and each track

	Group A		Group B	
	NT1	NT 2	DT1	DT2
P 1	0.8	0.7	0.3	0.2
P 2	0.65	0.6	0.7	0.5
P 3	0.8	0.8	0.2	0.15

Data are normalized and LIMMA tool compares average signal for each ROI (row) between two groups. Note: for limma to work, need at least two tracks in each group.

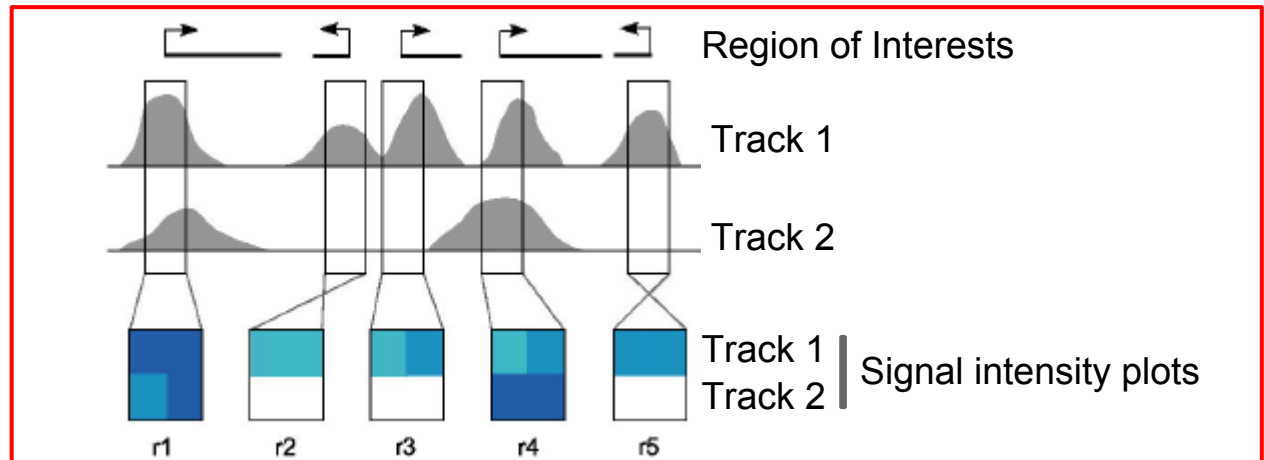
## III. Results output

ROIs (i.e. promoter, enhancers, etc) that exhibit significantly different signal between the groups compared, are provided to users as a region track which can be downloaded or used for downstream analysis

# Methodology: Spark

Spark allows user-guided k-means clustering to visualize epigenomic profiles on a genome-wide scale.

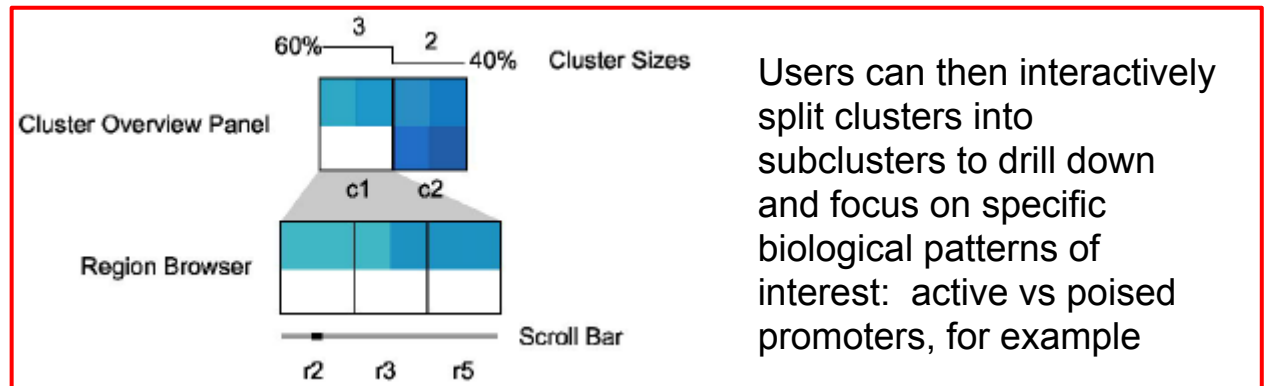
## I. Preprocessing (ROIs = r1 - r5)



## II. Clustering (C1, C2)



## III. Interactive Visualization



Watch Video

<http://sparkinsight.org/>

# Methodology: GREAT & HOMER

The patterns identified by Spark require further investigation to assess biological meaning

- GREAT<sup>2</sup> (Genomic Regions Enrichment of Annotations Tool) analysis
  - Assesses functional significance of cis-regulatory regions, here identified by Spark
  - Associates both proximal and distal binding events (Chip-Seq) with putative target genes
  - Uses gene annotation from several ontologies to associate the cis-regulatory regions with the annotations
  - Calculates statistical enrichments for associations between cis-regulatory regions and the annotations
  - Outputs annotation terms significantly associated with input cis-regulatory regions
- HOMER<sup>3</sup> analysis:
  - Utilizes a de novo motif discovery algorithm
  - Scores motifs by looking for motifs with differential enrichment between two sets of sequences
  - Enrichment measured using the cumulative hypergeometric distribution (or cumulative binomial distribution for large data sets)
  - Motifs with p-values below 1e-10 are typically reasonable candidates for further investigation. Motifs with a p-value greater than 1e-10 or even 1e-12 are likely false positives

<sup>2</sup>McLean et al. "GREAT improves functional interpretation of cis-regulatory regions". Nature Biotechnology, 28: 495-503 (2010).

<sup>3</sup>Heinz et al. "Simple Combinations of Lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities". Molecular Cell, 28: 576-589 (2010).

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**
          - Class: Enhancer
            - Limma:Myeloid\_comb**
          - Class: Gene
          - Class: Marker
          - Class: Sequence
        - Lists & Selections**
          - Lists of Tracks**
            - HL60\_Tracks
            - HSC\_Tracks
            - Immune\_HL60\_H3K4me1
            - Myeloid\_Tracks**
            - release9\_H3K27me3
            - release9\_H3K36me3
            - release9\_H3K4me1
            - release9\_H3K4me3
            - release9\_H3K9me3

### Details

Attribute	Value
Group	Epigenome ToolSet Demo Input Data
Database	ASHG 2013 Demo
Name	Limma:Myeloid_comb
Description	
BigBed	

### Input Data

- Myeloid\_Tracks
- HSC\_Tracks
- Limma:Myeloid\_comb

### Output Targets

- GenboreeUser\_database

**Step 1** - Drag "Myeloid\_Tracks", "HSC\_Tracks", and "Limma:Myeloid\_comb" from "ASHG 2013 Demo". If you have generated these by following supplemental slides, then drag them from your database.

**Step 2** - Drag your database in output targets

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

Welcome to the Genboree Webportal [Getting Started]

**Step 3 -Expand "Epigenome" --> "Analyze Signals" --> select "Cluster by Spark"**

**Data Select**

Refresh

Epig

Databases

ASHG 2013 Demo

All Annotations in Database

Tracks

Class: Enhancer

Limma:Myeloid\_comb

Class: Gene

Class: Marker

Class: Sequence

Lists & Selections

Lists of Tracks

HL60\_Tracks

HSC\_Tracks

Immune\_HL60\_H3K4me1

Myeloid\_Tracks

release9\_H3K27me3

release9\_H3K36me3

release9\_H3K4me1

release9\_H3K4me3

release9\_H3K9me3

filter...

Find Differences By Regression

Cluster by Spark

Compare by LIMMA

Details

Attribute

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

BigBed

**Input Data**

Myeloid\_Tracks

HSC\_Tracks

Limma:Myeloid\_comb

**Output Targets**

GenboreeUser\_database

Tool Settings
Cluster by Spark (Analyze Signals)

Tool Overview

Inputs:

Data Tracks/Files:

Limma:Myeloid\_comb

Group: Epigenome ToolSet Demo  
Input Data,  
Database: ASHG 2013 Demo  
Group: Epigenome ToolSet Demo  
Input Data,  
Database: ASHG 2013 Demo  
Group: Epigenome ToolSet Demo  
Input Data,  
Database: ASHG 2013 Demo

Output Database:

Database: GenboreeUser\_database
Group: GenboreeUser\_group

Spark Analysis Settings

Analysis Name
Spark\_Myeloid\_HSC-13-10-15-

Select ROI Track

MCD34:H3K4me3 17
MCD34:H3K9me3 28
Limma:Myeloid\_comb
CD14:H3K27ac 21 26
CD14:H3K27me3 21 05

Region Label
MyROIs

Statistics Type
global

# of Clusters
10

# of Bins:
20

Data Track Colors:

CD14:H3K27ac 21 26

blue

▼

CD14:H3K27me3 21 05

blue

▼

CD14:H3K36me3 21 06

blue

▼

CD14:H3K4me1 21 03

blue

▼

CD14:H3K4me3 01 07

blue

▼

CD14:H3K9me3 21 04

blue

▼

Limma:Myeloid\_comb

blue

▼

MCD34:H3K27ac 36 52

blue

▼

MCD34:H3K27me3 17

blue

▼

MCD34:H3K36me3 28

blue

▼

MCD34:H3K4me1 28

blue

▼

MCD34:H3K4me3 17

blue

▼

MCD34:H3K9me3 28

blue

▼

Submit
Cancel

**Step 4 - Analysis Name**  
"Spark\_Myeloid\_HSC"

**Step 5 - Select ROI (Region of Interest)**  
Track "Limma\_Myeloid\_comb"

**Step 6 - Enter the # of Clusters "10".**  
During visualization, User can change number of clusters based on this expert knowledge.

You will see this message upon successful submission of your Spark job:

Job Submission Status
Cluster by Spark (Analyze Signals)

Job Id: wbJob-spark-IkDrDs-1662

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

Status of the jobs submitted can be obtained through Job Summary

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 logos is a navigation bar with tabs for System/Network, Data, QC and Pre-processing, Genome, Transcriptome, Cistrome, Epigenome, Metagenome, Visualization, and Help. The left sidebar contains a menu with options: User Profile, Groups, Hosts, Jobs, Request Feature, Atlas Tools Access, BRL AUTO TEST, and FDACC. The 'Jobs' menu item is highlighted, and a red box around it has an arrow pointing to the 'Job Summary' option. The 'Job Summary' window is open, showing a 'Tool Overview' section with a 'Settings' tab. The settings include 'Start Date' (2013/7/18), 'End Date' (YYYY/MM/DD), 'Sort Order' (Newest first), and 'Group By' (None). Below these settings is an 'Advanced Settings' section. A red box around the 'Generate Report' button has an arrow pointing to it, with a text box below it saying 'Select "Generate Report" to see Job Summary'. The main content area on the right has sections for 'Details', 'Input Data', and 'Output Targets'. The 'Details' section is empty. The 'Input Data' and 'Output Targets' sections have icons for adding, removing, and refreshing data.

Genboree Workbench! [Getting Started]

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

User Profile Groups Hosts Jobs Request Feature

Atlas Tools Access BRL AUTO TEST FDACC

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

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.

BRL

HGSC HUMAN GENOME SEQUENCING CENTER



## You will receive an email with the following message when you Spark job has finished:

Your Spark job completed successfully.

### Job Summary:

JobID - wbJob-spark-AsqKIJ-9045  
Analysis Name -

### Inputs:

# of Data Tracks - 12  
ROI Track - Limma:Myeloid\_comb

### Outputs:

Output DB - UseCase  
Output Host - genboree.org

### Settings:

k - 10  
normType - exp  
numBins - 20  
regionLabel -  
statsType - global

### Additional Info:

To view your results in the Spark GUI:

- (a) download and unzip the results archive and then
- (b) launch Spark via Java Web Start and open the analysis folder.

Spark Java Web Start Link:

<http://www.bcgsc.ca/downloads/spark/current/start.inlp>

- The Genboree Team

### Result File Location in the Genboree Workbench:

(Direct links to files are at the end of this email)

Host: genboree.org  
Grp:  
Db:  
Files Area:  
\* Spark - Results/  
\*  
\*  
\*

### Result File URLs (click or paste in browser to access file):

FILE: Spark\_Myeloid\_HSC.zip

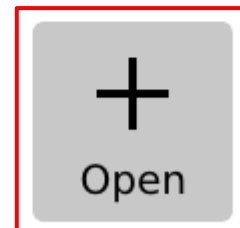
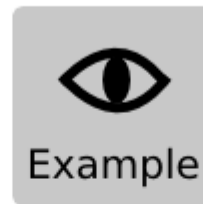
URL:

[http://genboree.org/java-bin/apiCaller.jsp?srcPath=http%3A%2F%2Fgenboree.org%2FREST%2Fv1%2Fgrp%2Fvamin\\_group%2Fdb%2FUseCase%2Ffile%2Fspark%2520-%2520Results%2FSpark\\_Myeloid\\_HSC%2FSpark\\_Myeloid\\_HSC.zip%2Fdata%3F&fileDownload=true&promptForLogin=true&errorFormal=html](http://genboree.org/java-bin/apiCaller.jsp?srcPath=http%3A%2F%2Fgenboree.org%2FREST%2Fv1%2Fgrp%2Fvamin_group%2Fdb%2FUseCase%2Ffile%2Fspark%2520-%2520Results%2FSpark_Myeloid_HSC%2FSpark_Myeloid_HSC.zip%2Fdata%3F&fileDownload=true&promptForLogin=true&errorFormal=html)

**Step 7 - Follow the steps to view the results in Spark GUI. Make sure Java is installed.**

**Step 8 - Download Spark Results and UNZIP the Folder**

# SPARK GUI

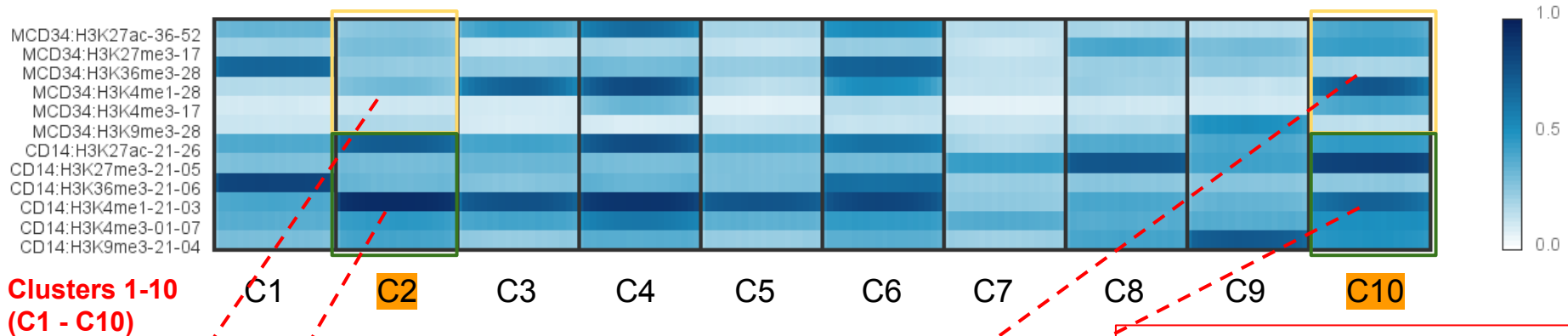
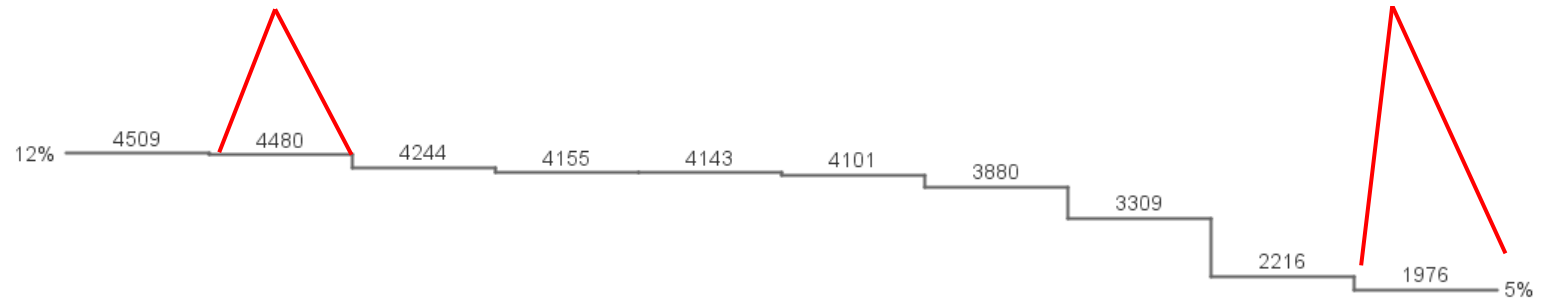


**Step 9** - Select Open and  
choose the unzipped folder

# SPARK output results

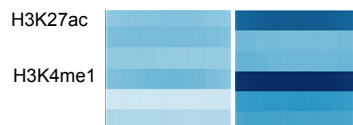
**4480 Enhancers (out of ~35000)  
have this profile of epigenomic marks.**

**1976 Enhancers (out of ~35000)  
have this profile of epigenomic marks.**



**Clustering (C2) reveals a shift from inactive epigenomic marks in human immune stem cells (MD34) to active epigenomic marks in myeloid cells (CD14).**

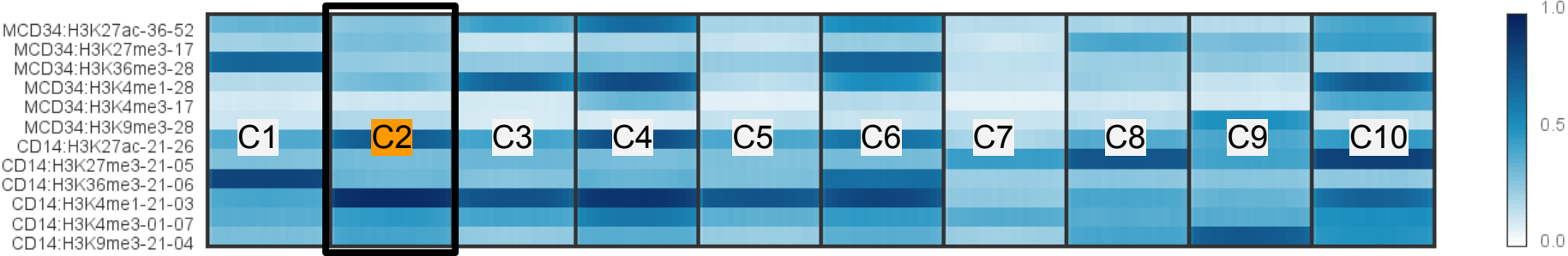
**Clustering (C10) reveals a shift from active epigenomic marks in human immune stem cells (MD34) to poised epigenomic marks in myeloid cells.(CD14)**



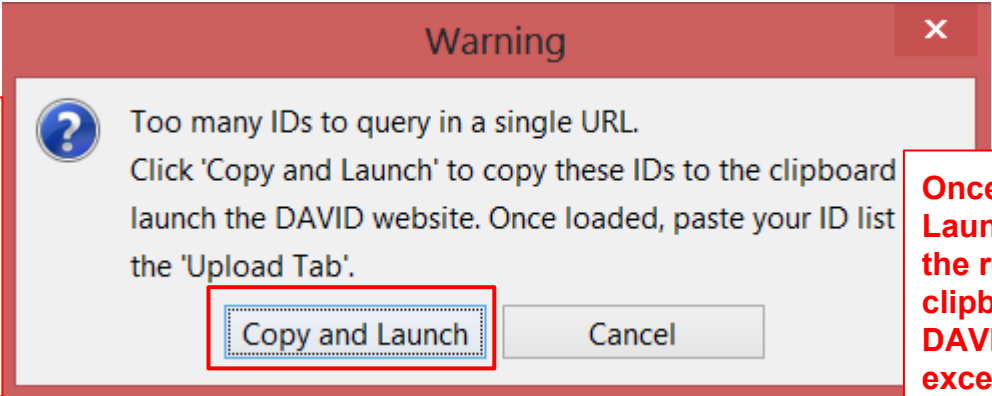
As an example, we will take C2 enhancers and perform Motif and GREAT analysis via Genboree.



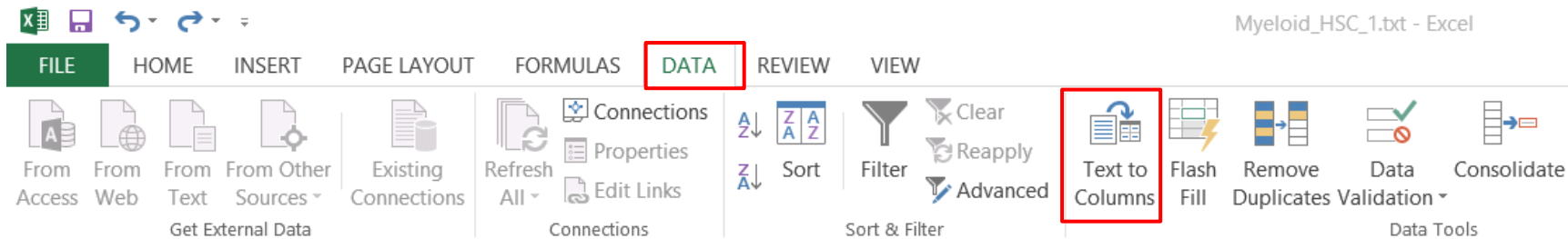
**Step 10 - Select C2 cluster and then select GO on top of the page**



**Step 11 - Select Copy and Launch. Web-browser will be launched. We will not use DAVID as regions are enhancers and not genes. We will do Motif and GREAT analysis via Genboree.**



**Once selected “Copy and Launch”. This action copies the region IDs to your clipboard and launches the DAVID website. Next, open excel and paste the region IDs (which exist on your clipboard) to make a BED file**



**Step 12 - Paste and Select the column. Under “Data”, select “Text to Columns”**

1	chr1:2073802-2074400
2	chr1:2081202-2083600
3	chr1:3581602-3583800
4	chr1:6804802-6805400
5	chr1:6816802-6817000
6	chr1:8164802-8165400
7	chr1:8186802-8188200
8	chr1:8284202-8285400
9	chr1:8285602-8285800
10	chr1:8949202-8951800
11	chr1:9152802-9153400
12	chr1:9291202-9294200
13	chr1:10049802-10050800
14	chr1:10051602-10052200
15	chr1:10052402-10054000
16	chr1:10268402-10270000
17	chr1:11027002-11028200
18	chr1:11785402-11786200
19	chr1:11874202-11874800
20	chr1:11911402-11911800
21	chr1:11953402-11954000
22	chr1:12256602-12257600

**Convert Text to Columns Wizard - Step 1 of 3**

The Text Wizard has determined that your data is Delimited.

If this is correct, choose Next, or choose the data type that best describes your data.

Original data type

Choose the file type that best describes your data:

☒ Delimited - Characters such as commas or tabs separate each field.

☐ Fixed width - Fields are aligned in columns with spaces between each field.

Preview of selected data:

1	chr1:2073802-2074400
2	chr1:2081202-2083600
3	chr1:3581602-3583800
4	chr1:6804802-6805400
5	chr1:6816802-6817000
6	chr1:8164802-8165400

**Step 13 - Select Delimited and select “Next”**

Buttons: Cancel, < Back, **Next >**, Finish

## Convert Text to Columns Wizard - Step 2 of 3



This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.

### Delimiters

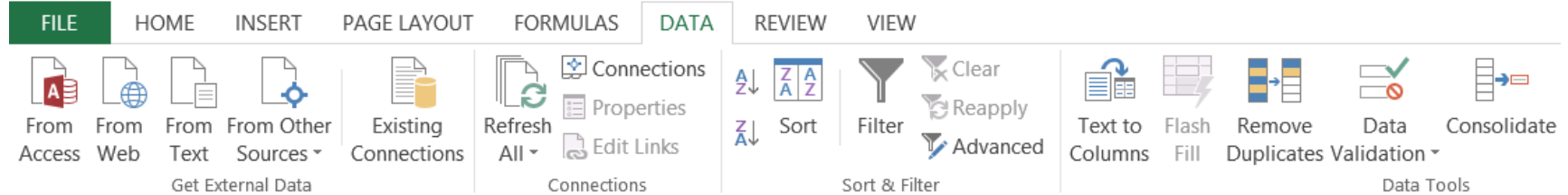
☒ Tab☐ Semicolon☐ Comma☐ Space☒ Other: ☐ Treat consecutive delimiters as oneText qualifier: 

### Data preview

chr1	2073802-2074400
chr1	2081202-2083600
chr1	3581602-3583800
chr1	6804802-6805400
chr1	6816802-6817000
chr1	8164802-8165400

**Step 14 - Select Delimiters as Others and enter a colon ":". Select "Finish".**

**Perform similar steps to separate Chromosome Start and Stop into different columns, but instead of entering a colon, you will enter hyphen.**



B1 : 2073802

	A	B	C
1	chr1	2073802	2074400
2	chr1	2081202	2083600
3	chr1	3581602	3583800

**Step 15 - Select columns B and C and format the cells by right clicking and choosing "Format Cells". Select category as "Number" with 0 decimal place**

13	chr1	10049802	10050800
14	chr1	10051602	10052200
15	chr1	10052402	10054000
16	chr1	10268402	10270000
17	chr1	11027002	11028200
18	chr1	11785402	11786200

**Step 16 - Save the file as "Spark -Myeloid\_HSC-C2" in Tab-delimited format**

**Format Cells**

Number Alignment Font Border Fill Protection

Category:

- General
- Number**
- Currency
- Accounting
- Date
- Time
- Percentage
- Fraction
- Scientific
- Text
- Special
- Custom

Sample: 2073802

Decimal places: 0

☐ Use 1000 Separator (,)

Negative numbers:

- 1234
- 1234
- (1234)
- (1234)

Number is used for general display of numbers. Currency and Accounting offer specialized formatting for monetary value.

OK Cancel



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

Welcome to Genboree! [Getting Started]

**Data Selector**

Refresh

genboree.org

- Atlas Tool
- BRL AUTO
- EDACC
- Epigenome Informatics Demo Output Data
- Epigenome Informatics Workshop (May 2012)
- Epigenome ToolSet Demo Input Data
- Epigenomics Roadmap Repository
- GenboreeUser\_group
  - Databases
    - GenboreeUser\_database
  - Projects
    - GMT\_Tutorial
    - JonathanMill\_Lab
    - Public
    - ROI Repository
    - Targeted Atlases
- genboree.bcgsc.ca
- genboree.cbrc.jp
- www.brain-research-lab.org

**Data**

- Databases
- Entity Lists
- Entrypoints
- Files
- Projects
- Samples & Sample Sets
- Tracks

**Import**

- Array Data
- Track Metadata
- Upload Track Annotations

**Details**

Attribute	Value
Group	Epigenome ToolSet Demo Input Data
Database	ASHG 2013 Demo
	Limma:Myeloid_comb

**Input Data**

**Output Targets**

GenboreeUser\_database

**Step 18 - Select "Upload Track Annotations"**

**Step 17 -To upload track annotations, drag your database**

Tool Settings

## Upload Track Annotations

**Tool Overview**

**Input Data:**

Data File: n/a [ None selected ]

**Output Location:**

Database: GenboreeUser\_database Group: GenboreeUser\_group

**Settings**

Select File  Spark-Myel...HSC\_C2.txt

Input Format  ▼

Track Class

Track Name  :

☒ Skip non-assembly chromosomes

☐ Skip out-of-range annotations

☒ 0 based and half open

☐ 1 based and fully closed

**Step 19** - Choose File "Spark-Myeloid\_HSC-C2.txt" to upload


**Step 20** - Select Bed

**Step 21** - Name Track Class as "Enhancer", Track Name as "Spark:Myeloid\_HSC\_C2"

You will see this message upon successful submission of your job (and you will have to wait for a success email):

Job Submission Status

## Upload Track Annotations

 **Job Id:** wbJob-uploadTrackAnnos-oDsEoK-0502

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.

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

## Data Selector

Refresh

Data Filter: Select a filter... ▾

- GenboreeUser\_database
  - All Annotations in Database
  - Tracks
    - Class: BranchSpecificROI
    - Class: Class
    - Class: Enhancer
      - Spark:HL60\_active\_C12
      - Spark:Myeloid\_HSC\_C2**
    - Class: Gene
    - Class: High Density Score Data
    - Class: MACS
    - Class: Marker
    - Class: Sequence
    - Class: SPARK
  - Lists & Selections
  - SampleSets
  - Samples
  - Files
  - Queries
  - Projects
    - GenboreeUser\_project
    - Use\_Case\_01\_GU

## Details

Attribute	Value
Group	GenboreeUser_group
Database	GenboreeUser_database
Name	Spark:Myeloid_HSC_C2
Description	
BigBed	none

## Input Data

- Spark:Myeloid\_HSC\_C2

**Step 22** -Drag Spark: Myeloid\_HSC\_C2 from your database in Input Data. Drag Your Database and Project Page in Output Targets

## Output Targets

- GenboreeUser\_database
- Use\_Case\_19\_GU

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

Welcome to the Genboree Workbench

**Find Motifs** **HOMER**

**Step 23 - Select "HOMER"**

**Data Selector**

Refresh Data Filter: S

- GenboreeUser\_database
  - All Annotations in Database
  - Tracks
    - Class: BranchSpecificROI
    - Class: Class
    - Class: Enhancer
      - Spark:HL60\_active\_C12
      - Spark:Myeloid\_HSC\_C2
    - Class: Gene
    - Class: High Density Score Data
    - Class: MACS
    - Class: Marker
    - Class: Sequence
    - Class: SPARK
  - Lists & Selections
  - SampleSets
  - Samples
  - Files
  - Queries
- Projects
  - GenboreeUser\_project
  - Use\_Case\_01\_GU

**Input Data**

Spark:Myeloid\_HSC\_C2

**Output Targets**

GenboreeUser\_database  
Use\_Case\_19\_GU

Tool Settings

# HOMER BETA

⊕ **Tool Overview**

**Input Track:**

**Genes/Peaks of Interest:** *Spark:Myeloid\_HSC\_C2* Group: *GenboreeUser\_group*, Database: *GenboreeUser\_database*

**Output Database/Project:**

**Database/Projects Of Interest:** *GenboreeUser\_database* Group: *GenboreeUser\_group*  
*Use\_Case\_19\_GU* Group: *GenboreeUser\_group*

**Settings**

**Analysis Name**

**Genome Version**

☒ Run against Genome  
☐ Run against Promoters

**Promoter Set**

⊕ **Basic Options**  
⊕ **Advanced Options**  
⊕ **Known Motif Options/Visualizations**

**Step 24 - You can change the Analysis Name or leave default name**

You will see this message upon successful submission of your job:

Job Submission Status

# HOMER BETA

✓ **Job Id:** *wbJob-homer-oJzHGs-1668*

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.

Welcome to the Genboree Workbench

## Data Selector

Refresh

Data Filter: S

- Analyze Structural Variants
- Find Motifs
- SNPs
- GREAT**

**Step 26 - Select "GREAT"**

- GenboreeUser\_database
  - All Annotations in Database
  - Tracks
    - Class: BranchSpecificROI
    - Class: Class
    - Class: Enhancer
      - Spark:HL60\_active\_C12
      - Spark:Myeloid\_HSC\_C2
    - Class: Gene
    - Class: High Density Score Data
    - Class: MACS
    - Class: Marker
    - Class: Sequence
    - Class: SPARK
  - Lists & Selections
  - SampleSets
  - Samples
  - Files
  - Queries
- Projects
  - GenboreeUser\_project
  - Use\_Case\_01\_GU

Group	GenboreeUser_group
Database	GenboreeUser_database
Name	Spark:Myeloid_HSC_C2
Description	
BigBed	none

## Input Data

- Spark:Myeloid\_HSC\_C2

**Step 25 -Drag Spark: Myeloid\_HSC\_C2 from your database in Input Data. Drag Your Database and Project Page in Output Targets**

## Output Targets

- GenboreeUser\_database
- Use\_Case\_19\_GU

Tool Settings

## GREAT BETA

+ Tool Overview

**Tracks of Interest:**

Track: Spark:Myeloid\_HSC\_C2 Group: GenboreeUser\_group,  
Database: GenboreeUser\_database

**Output Database/Project:**

Database/Projects Of Interest: GenboreeUser\_database Group: GenboreeUser\_group  
Use\_Case\_19\_GU Group: GenboreeUser\_group

**Settings**

Analysis Name GREAT-2013-10-15-15:24:48


**Submit** Cancel

**Step 27 - You can change the Analysis Name or leave default name**

You will see this message upon successful submission of your job:

Job Submission Status

## GREAT BETA

 **Job Id:** wbJob-great-EmC8IG-0940

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



You will receive an email with the following message when job is finished

Hello

Your GREAT job completed successfully.

Job Summary:

JobID - wbJob-great-AC73IN-8111

Additional Info:

Target Group:

Target Database

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

Result File Location in the Genboree Workbench:

<http://genboree.org/java-bin/project.jsp?projectName=Roadmap%20Epigenome%20Data%20Analysis>

## Project Page

### Project News:

2013/10/15:

Genboree User ran a GREAT (Genomic Regions Enrichment of Annotations Tool) job (wbJob-great-EmC8IG-0940). Click the link below to perform live analysis with GREAT:

- [Perform GREAT analysis](#)

2013/10/15:

Genboree User ran a HOMER job (wbJob-homer-oJzHGs-1668) and the results are available at the link below.

- [Link to Homer Results](#)

Will show significant enriched Motifs.

### GREAT - Genomic Regions Enrichment of Annotations Tool

Host	Group	Database	Track	Link for live analysis
genboree.org	GenboreeUser_group	GenboreeUser_database	Spark:Myeloid_HSC_C2	<a href="#">Click here</a>

GREAT version 2.0.2 current (04/03/2012 to now)

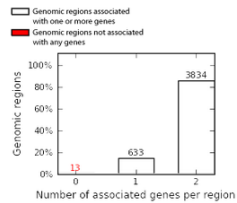
## Job Description

## Region-Gene Association Graphs

What do these graphs illustrate?

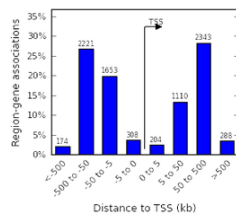
Number of associated genes per region

Download as PDF.



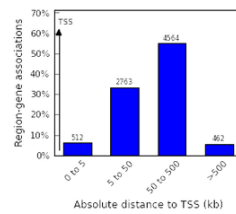
Binned by orientation and distance to TSS

Download as PDF.



Binned by absolute distance to TSS

Download as PDF.



## Global Controls

Global Export

Which data is exported by each option?

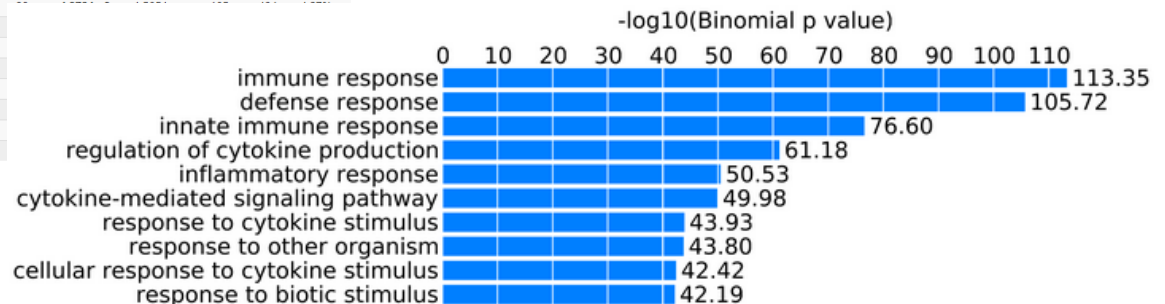
## GO Biological Process

## Visualize the table

Table controls: Export Shown top rows in this table: 20 Set Term annotation count: Min: 1 Max: Inf Set Visualize this table [select one]

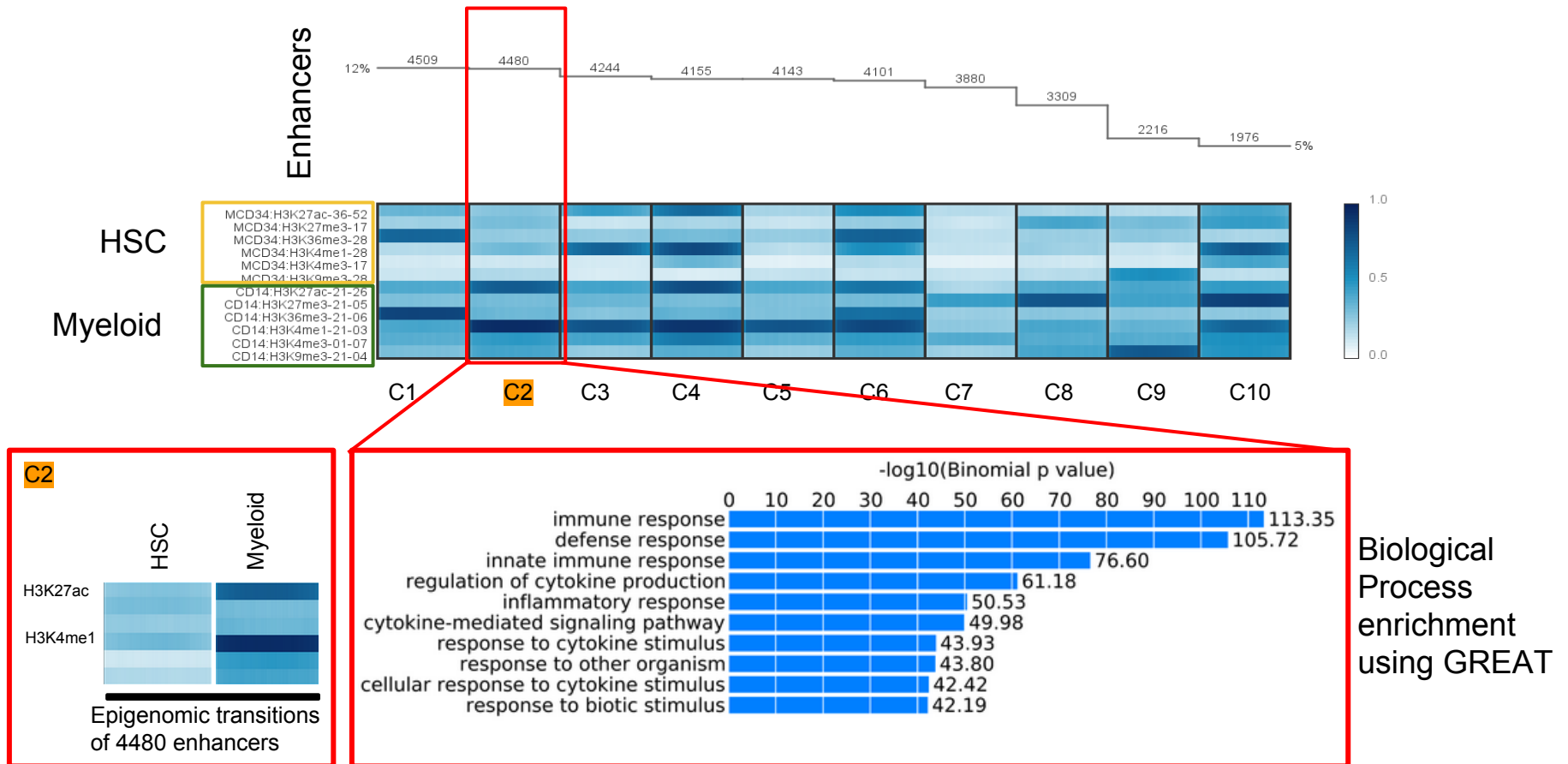
Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
immune response	1	4.4630e-114	3.9100e-110	2.5842	698	15.58%	3	1.1785e-16	1.6256	290	819	7.50%
defense response	2	1.9063e-106	8.3504e-103	2.4230	729	16.27%	9	8.0547e-14	1.5425	295	878	7.63%
innate immune response	4	2.5040e-77	5.4843e-74	2.8635	410	9.15%	27	5.9414e-9	1.6475	150	418	3.88%
regulation of cytokine production	6	6.6444e-62	9.7019e-59	2.7843	344	7.68%	44	6.1946e-8	1.7069	116	312	3.00%
inflammatory response	10	2.9646e-51	2.5973e-48	2.5911	321	7.17%	30	1.7891e-8	1.7079	125	336	3.23%
cytokine-mediated signaling pathway	11	1.0464e-50	8.3345e-48	3.2113	228	5.09%	280	4.8656e-3	1.4771	74	230	1.91%
response to cytokine stimulus	13	1.1666e-44	7.8621e-42	2.2922	351	7.83%	57	4.6489e-7	1.5995	131	376	3.39%
response to other organism	14	1.5904e-44	9.9522e-42	2.1747	390	8.71%	122	2.9223e-5	1.4394	153	488	3.96%
cellular response to cytokine stimulus	15	3.7600e-43	2.1961e-40	2.6430	261	5.83%	184	3.6379e-4	1.5193	92	278	2.38%
response to biotic stimulus	16	6.5072e-43	3.5631e-40	2.1159	399	8.91%	144	6.1341e-5	1.4112	158	514	4.09%
positive regulation of cytokine production	18	1.6859e-41	8.2057e-39	3.0721	198	4.42%	164	1.5068e-4	1.7475	59	155	1.53%
regulation of defense response	20	1.3431e-39	5.8833e-37	2.2239	331	7.39%	43	6.2706e-8	1.6468	132	368	3.42%
regulation of immune response	21	9.4602e-39	3.9467e-36	2.0441	390	8.71%						
response to bacterium	22	9.7107e-38	3.8671e-35	2.3380	285	6.36%						
regulation of innate immune response	28	1.2666e-31	3.9632e-29	2.4648	215	4.80%						
immune effector process	29	1.6749e-30	5.0598e-28	2.7405	172	3.84%						
response to lipopolysaccharide	31	6.9457e-28	1.9629e-25	2.3661	204	4.55%						
response to molecule of bacterial origin	33	2.4502e-27	6.5049e-25	2.2971	212	4.73%						
leukocyte migration	34	2.8380e-27	7.3130e-25	2.4672	184	4.11%						
leukocyte chemotaxis	35	8.1503e-27	2.0401e-24	3.7894	95	2.12%						

The test set of 4,480 genomic regions picked 3,865 (22%) of all 17,744 genes.  
GO Biological Process has 8,751 terms covering 14,760 (83%) of all 17,744 genes, and 697,512 term - gene associations.  
8,761 ontology terms (100%) were tested using an annotation count range of [1, Inf].



The results show that the regions that are undergoing epigenomic changes during myeloid differentiation are involved in cytokine-regulated biological processes. This clearly does support our knowledge of myeloid differentiation being highly regulated by various cytokines. ([Oncogene, 2000; 19\(21\):2511-22](#))

# Summary of results



Cluster of enhancers that are undergoing epigenomic changes during myeloid cell differentiation from CD34 (HSC lineage) to CD14 (Myeloid lineage) were identified using Spark. Functional significance of biological process assessed through Spark do indicate importance of these regions during myeloid cell differentiation.

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, a navigation bar contains tabs for 'System/Network', 'Data', 'QC and Pre-processing', 'Genome', 'Transcriptome', 'Cistrome', 'Epigenome', 'Metagenome', 'Visualization', and 'Help'. The 'System/Network' tab is active, showing a sidebar with a tree view of resources. The 'Request Feature' option is highlighted in the sidebar, and a red box around it points to a 'Request Feature' dialog box. The dialog box has a title bar 'Request Feature' and a 'Settings' tab. It contains fields for 'User Name' (Genboree User), 'User Email' (andrewj@bcm.edu), and a 'Message' text area. 'Submit' and 'Cancel' buttons are at the bottom. In the background, the 'Details' table is visible with columns 'Attribute' and 'Value', showing 'View Link' and 'Link to Project' with values 'GenboreeUser\_group' and 'Use\_Case\_18\_GU'.

GENBOREE

BCM  
Baylor College of Medicine

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

Genboree Workbench! [Getting Started]

User Profile Groups Hosts Jobs Request Feature

Atlas Tools Access BRL AUTO TEST EDACC Epigenome Informatics Demo Epigenome Informatics Work Epigenome ToolSet Demo In Epigenomics Roadmap Repo GenboreeUser\_group GMT\_Tutorial JonathanMill\_Lab Public ROI Repository Targeted Atlases genboree.bcgsc.ca genboree.cbrc.jp www.brain-research-lab.org

Data Filter: Select a filter...

Details

Attribute	Value
View Link	Link to Project
	GenboreeUser_group
	Use_Case_18_GU

Tool Settings

Request Feature

Tool Overview

Settings

User Name Genboree User

User Email andrewj@bcm.edu

Message

Submit Cancel

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.

BRL

HGSC  
HUMAN GENOME SEQUENCING CENTER