

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

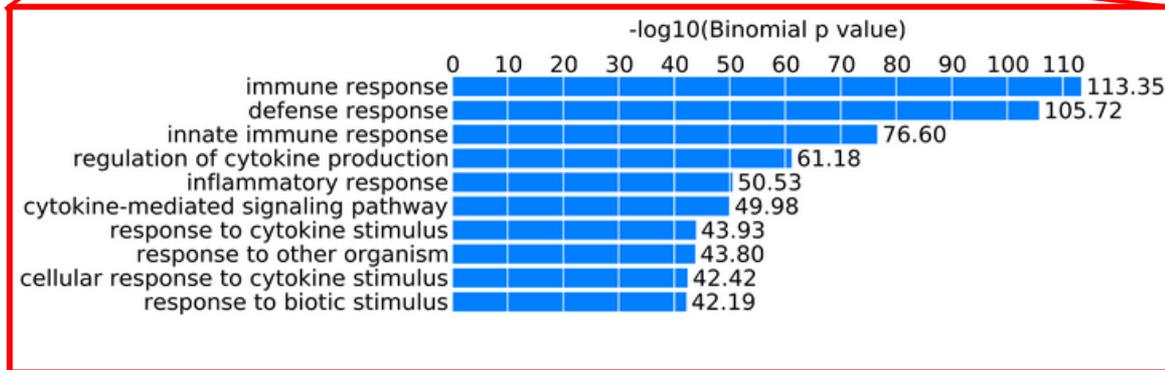
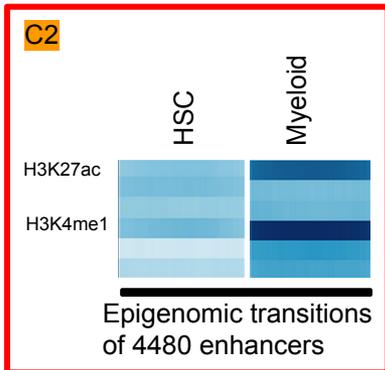
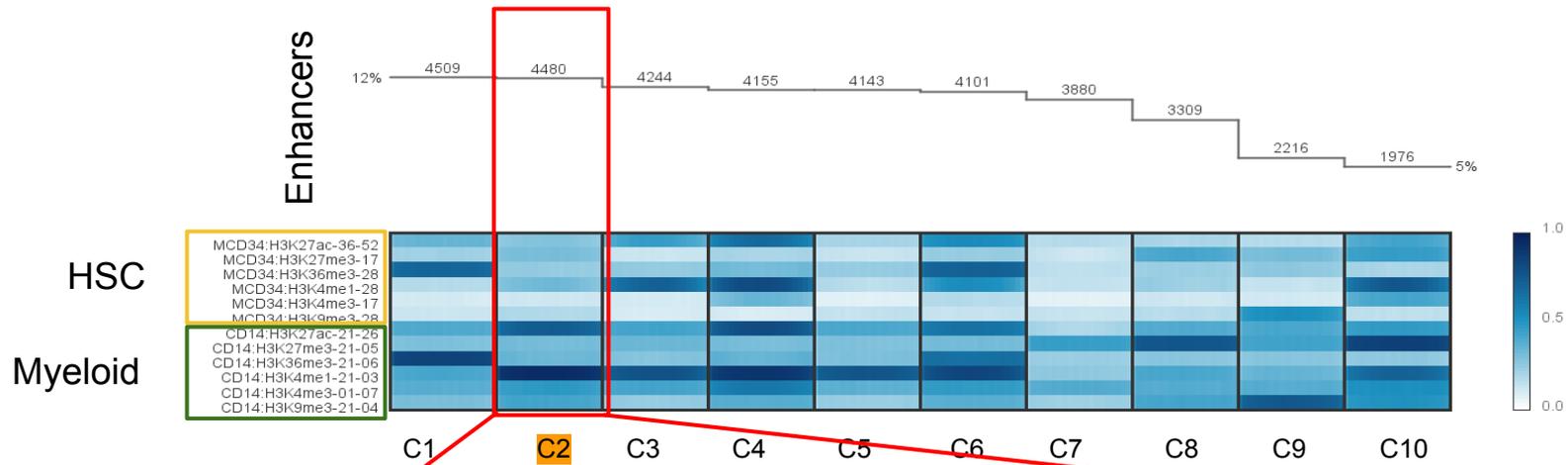
Baylor
College of
Medicine

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



Biological Process enrichment using GREAT

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 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:

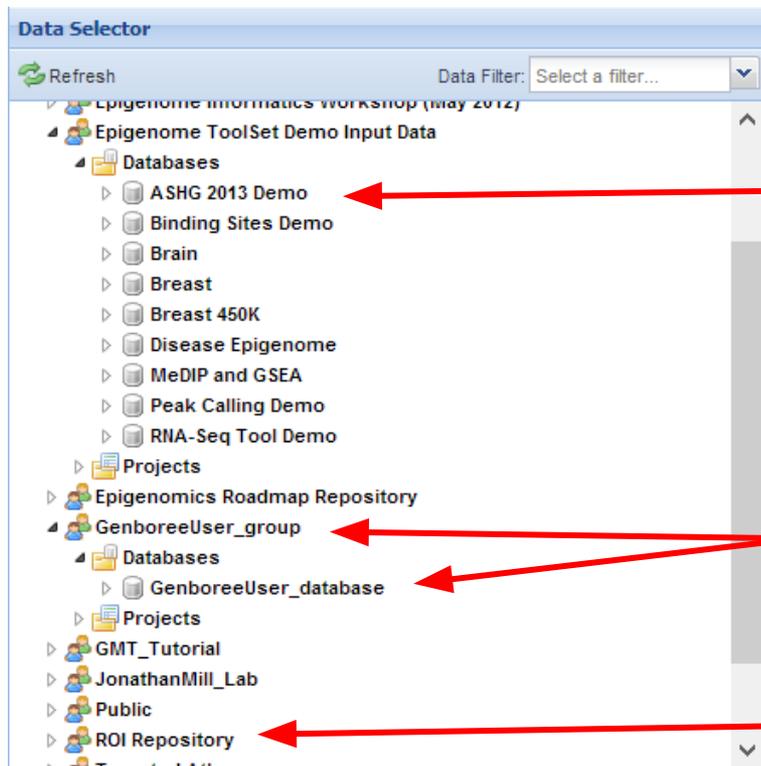
- Data Selector:** A tree view showing data sources from genboree.org, genboree.bcgsc.ca, genboree.cbrc.jp, and www.brain-research-lab.org. A callout box explains: **Data Selector:** Various Data Types (tracks, files, and ROIs (region of interests), etc).
- Details:** A table showing information for 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 sections in the interface.

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



The screenshot shows the GENBOREE Workbench interface. At the top, there are navigation tabs: System/Network, Data, QC and Pre-processing, Genome, Transcriptome, Cistrome, Epigenome, Metagenome, Visualization, and Help. The 'Data' tab is active, showing a 'Data Selector' panel with options: Databases, Entity Lists, and Entrypoints. A red box highlights the 'Databases' menu, which is open, showing options: Create Database, Rename Database, and Delete Database. A red callout box points to the 'Create Database' option with the text: "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." Below this, a 'Help: Create Database' dialog box is open. It contains a warning icon and text: "This tool is a recent addition. Please contact genboree_admin@genboree.org with questions or comments, or for help using it on your own data." Below that, it says: "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." The dialog has sections for 'Output Targets' and 'Tool-Specific Settings'. The 'Output Targets' section has instructions: "Drag 1 destination group into 'Output Targets'. The new database will be created in this group." and "Output type(s): Group min: 1 ; max: 1". The 'Tool-Specific Settings' section lists: Reference Sequence, Database Name, Description, Species, Version, and Submit. A red callout box points to the 'Output Targets' section in the main interface with the text: "To create a database, you need to drag a Group into 'Output Targets'". The 'Output Targets' section in the main interface is currently empty. At the bottom right, there is a logo for the Human Genome Sequencing Center (HGSC) and text: "Bioinformatics Research Laboratory of Medicine. is available free for academic use."

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 'Welcome to' message and a 'Data Selector' panel. The 'Data Selector' panel has a 'Refresh' button and a tree view of groups. The 'GenboreeUser_group' is selected. A 'Create Project' button is visible in the 'Data Selector' panel. A red box highlights the 'Create Project' button and the 'GenboreeUser_group' in the tree view.



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

The 'Create Project' form is shown. It has a 'Tool Overview' section with a 'Target Group' field containing '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 of the form are 'Submit' and 'Cancel' buttons. A red box highlights the 'Settings' section and the 'Unique Name' checkbox.

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")

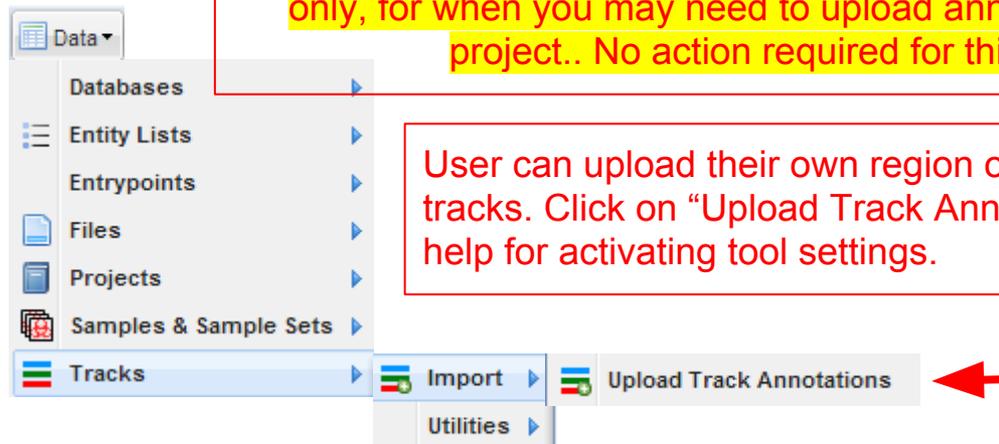
Description of Regions of Interest (ROIs) Tracks

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/), 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).¹

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:



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. The window is divided into sections: 'Tool Overview' and 'Settings'. Under 'Tool Overview', there is a section 'Databases with tracks of interest:' containing a 'Database:' field with the value 'Release 9 Repository' and a 'Group:' field with the value 'Epigenomics Roadmap Repository'. Under 'Settings', there are several fields: 'X-axis attribute' with a dropdown menu showing 'eaAssayType', 'Y-axis attribute' with a dropdown menu showing 'eaSampleType', 'Page Title' with the value 'Grid Viewer: Tracks from Relea', 'Grid Title' with the value 'Tracks from Release 9 Reposit', 'X Label' with the value 'eaAssayType', and 'Y Label' with the value 'eaSampleType'. At the bottom, there are 'Submit' and 'Cancel' buttons. Two red boxes with arrows point to the dropdown menus for 'X-axis attribute' and 'Y-axis attribute', with labels 'Step III - Select “eaAssay Type”' and 'Step IV - Select “eaSample Type”' respectively. The 'Submit' button is also highlighted with a red box.

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”

Methodology Overview

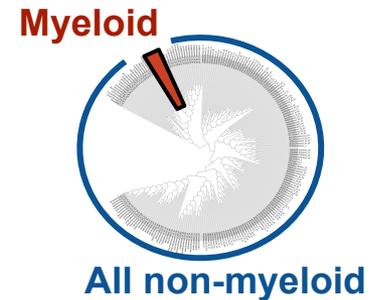
Use Case 18

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

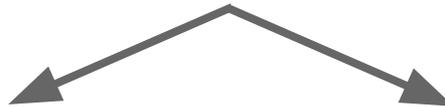


Use Case 19
Supplemental Slides

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

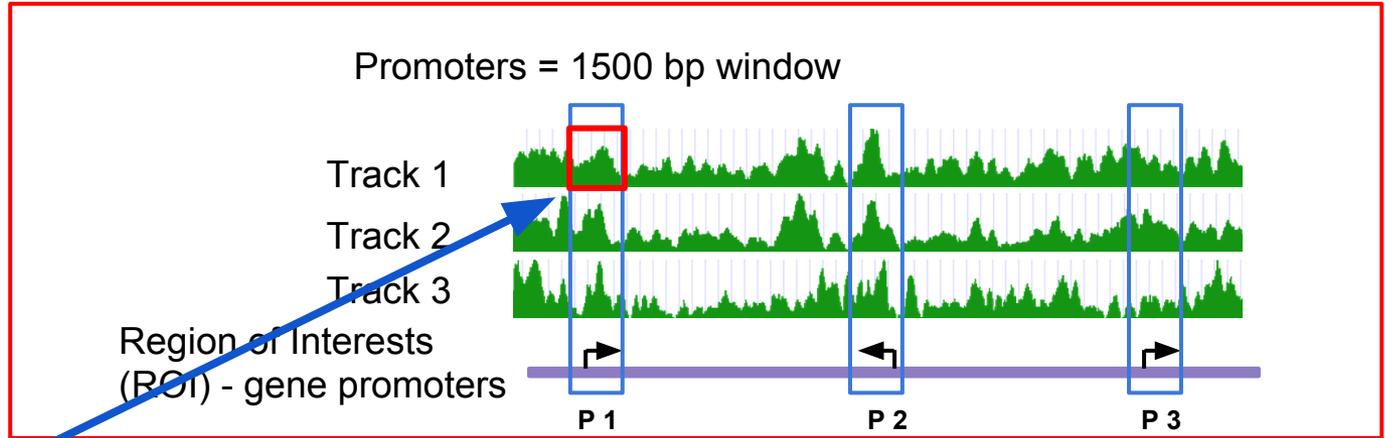


GREAT: assesses functional significance of cis-regulatory regions.

HOMER: de novo motif discovery.

Methodology: Clustering/Heatmap

I. Data Selection



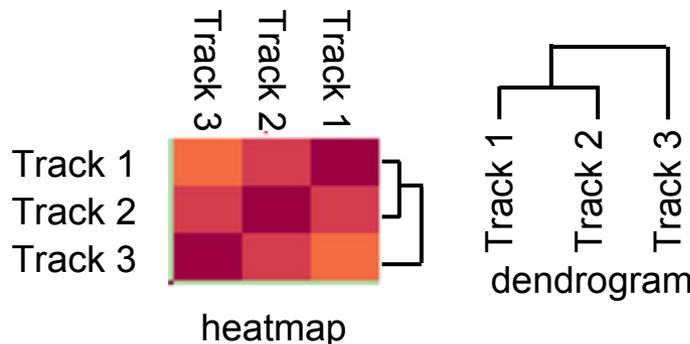
II. Signal processing- Heatmap Tool calculates average signal for each track and ROI (promoter) combination

	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

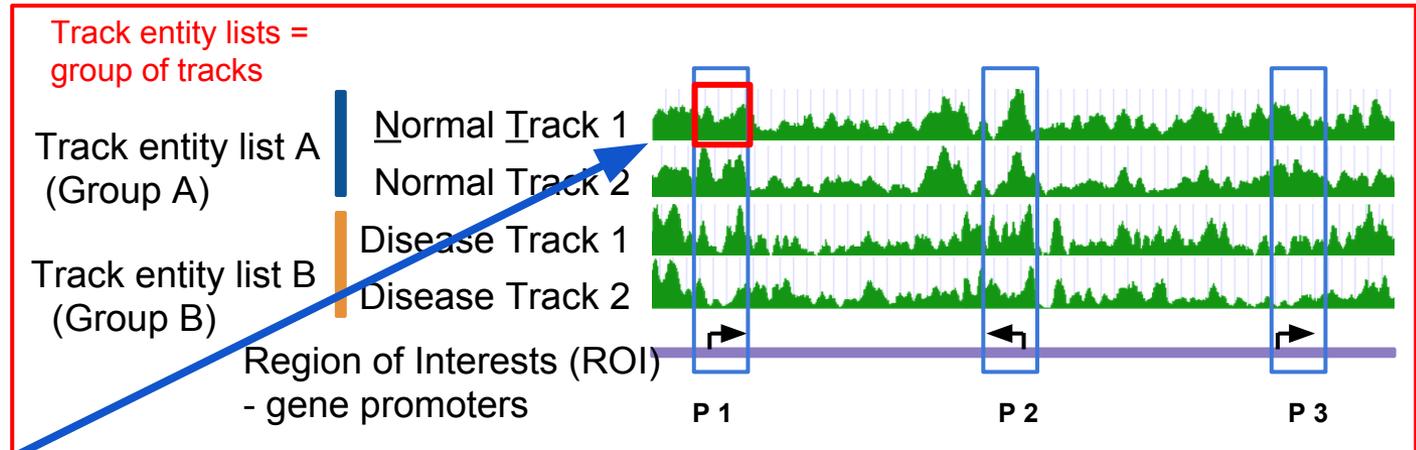
Correlation values are dynamically scaled and represented in heatmap.



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.

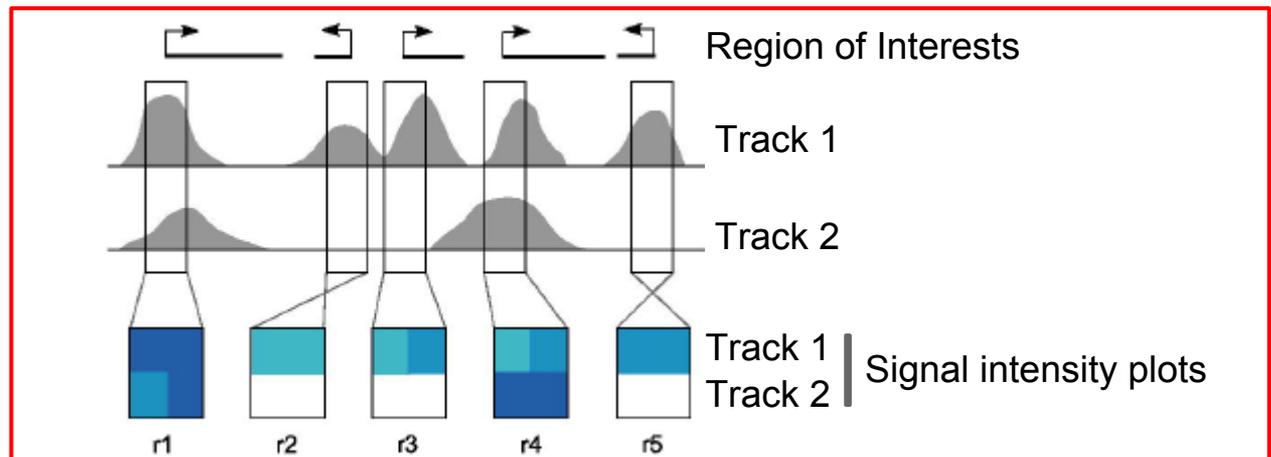
III. Results output

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.

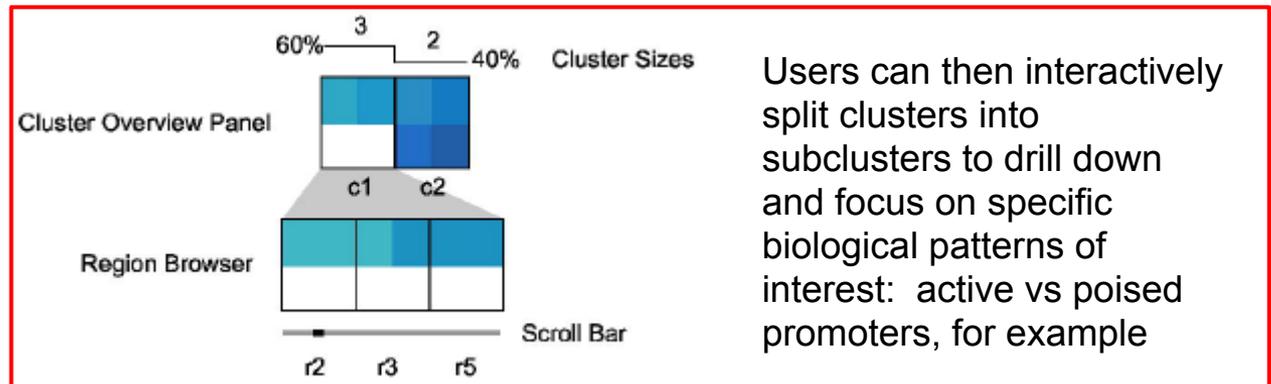
I. Preprocessing (ROIs = r1 - r5)



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

- GREAT² (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
- HOMER³ 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

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

³Heinz 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 (July 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"

- 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

Find Differences By Regression

Cluster by Spark

Compare by LIMMA

Input Data

- Myeloid_Tracks
- HSC_Tracks
- Limma:Myeloid_comb

Output Targets

- GenboreeUser_database

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_H3K4me3
 - release9_H3K9me3

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

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)

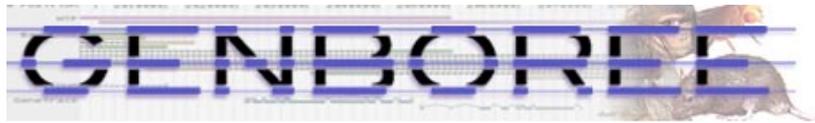
Job Id: wbJob-spark-1kDrDs-1662

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 for assistance.

Status of the jobs submitted can be obtained through Job Summary



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

User Profile | Groups | Hosts | **Jobs** | Request Feature

Job Summary | Data Filter: Select a filter...

Atlas Tools Access | BRL AUTO TEST | FDACC

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

Details

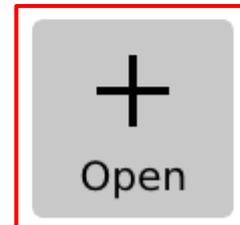
Attribute	Value
-----------	-------

Input Data

Output Targets

Select "Generate Report" to see Job Summary

SPARK GUI

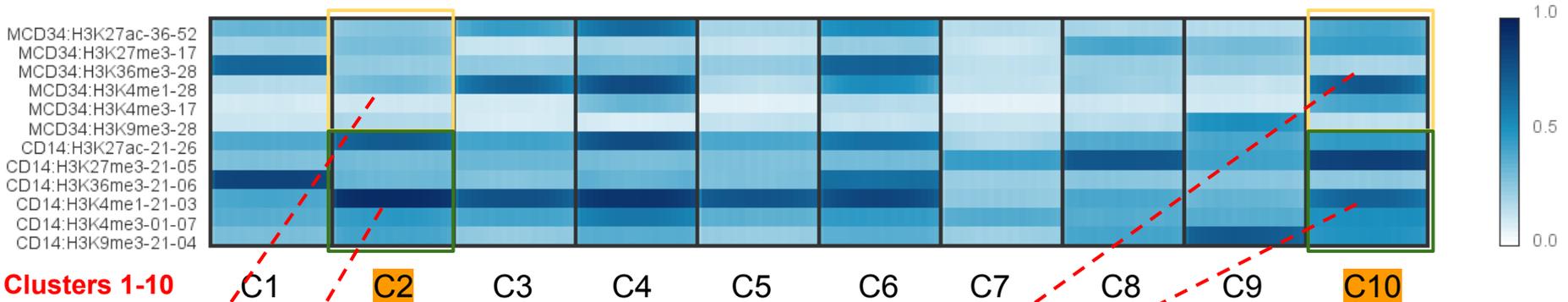
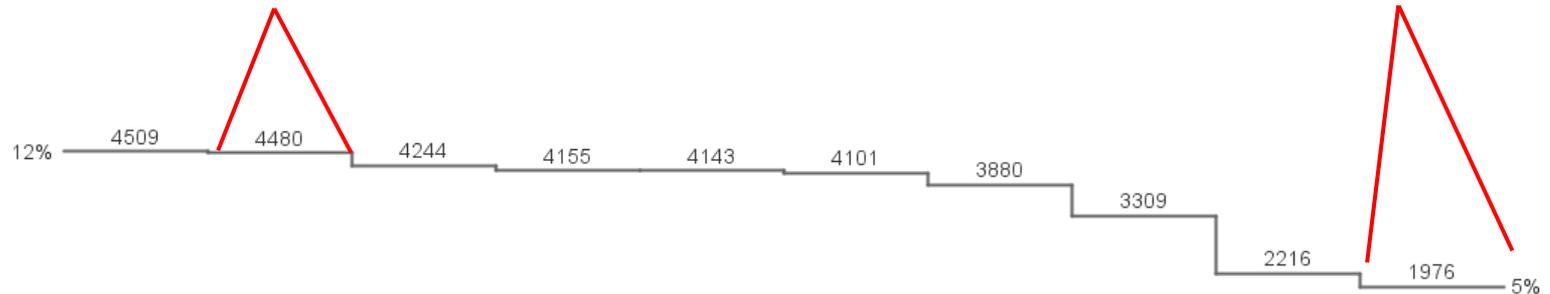


Step 9 - Select Open and choose the unzipped folder

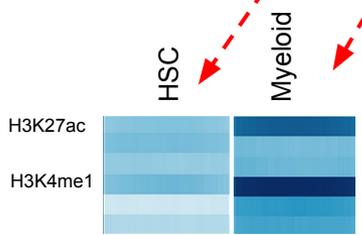
SPARK output results

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

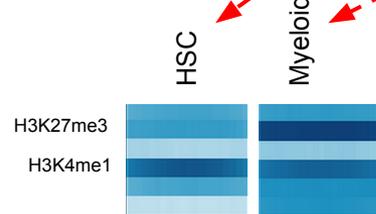
1976 Enhancers (out of ~35000) have this profile of epigenomic marks.



Clusters 1-10 (C1 - C10)

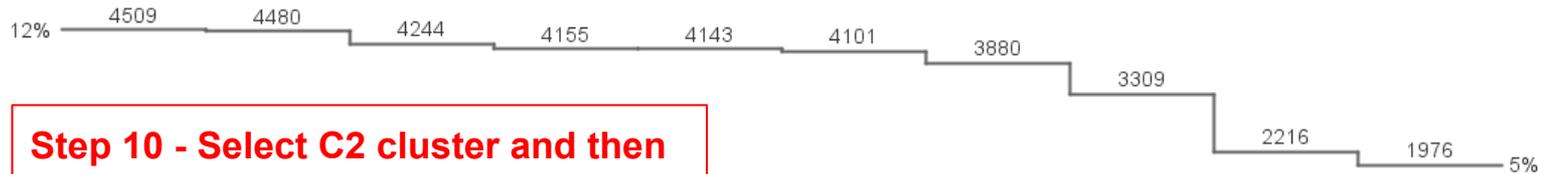
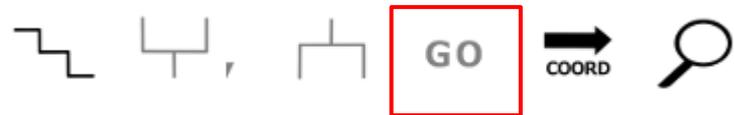


Clustering (C2) reveals a shift from inactive epigenomic marks in human immune stem cells (MD34) to active epigenomic marks in myeloid cells (CD14).

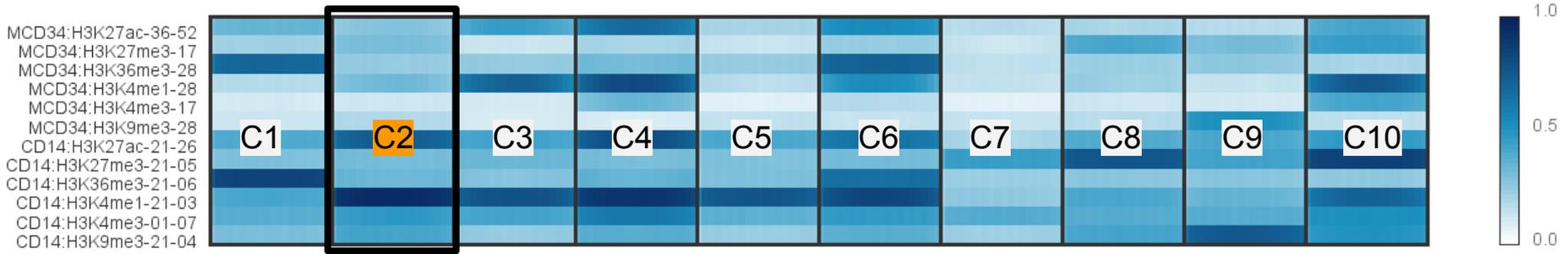


Clustering (C10) reveals a shift from active epigenomic marks in human immune stem cells (MD34) to poised epigenomic marks in myeloid cells.(CD14)

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.

Warning

?

Too many IDs to query in a single URL.
Click 'Copy and Launch' to copy these IDs to the clipboard launch the DAVID website. Once loaded, paste your ID list the 'Upload Tab'.

Copy and Launch

Cancel

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

Excel ribbon showing the **DATA** tab selected. The **Text to Columns** button is highlighted with a red box. Other visible buttons include From Access, From Web, From Text, From Other Sources, Existing Connections, Refresh All, Properties, Edit Links, Connections, Sort, Filter, Clear, Reapply, Advanced, Flash Fill, Remove Duplicates, Data Validation, and Consolidate.

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 one

Text qualifier:

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

Data preview

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

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

Cancel

< Back

Next >

Finish



FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW

From Access From Web From Text From Other Sources Existing Connections Refresh All Connections Properties Edit Links Sort Filter Sort & Filter Clear Reapply Advanced Text to Columns Flash Fill Remove Duplicates Data Validation Consolidate Data Tools

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

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

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

- Import
- Utilities

- Array Data
- Track Metadata
- Upload Track Annotations

Step 18 - Select "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 17 -To upload track annotations, drag your database



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**.



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 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_GII

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

Welcome to the Genboree Workbench

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

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

Group	Value
Database	GenboreeUser_database
Name	Spark:Myeloid_HSC_C2
Description	
BigBed	none

Input Data

↑ ↓ ✕ 🗑

- Spark:Myeloid_HSC_C2

Output Targets

↑ ↓ ✕ 🗑

- GenboreeUser_database
- Use_Case_19_GU



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**.



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 for assistance.

Welcome to the Genboree Workbench

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
 - Sample Sets
 - Samples
 - Files
 - Queries
- Projects
 - GenboreeUser_project
 - Use_Case_01_GU

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

Step 26 - Select "GREAT"

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 *GenboreeUser_database* Group: *GenboreeUser_group*
Of Interest: *Use_Case_19_GU* Group: *GenboreeUser_group*

Settings

Analysis Name

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 for assistance.

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:

- **Peform GREAT anlalysis**

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

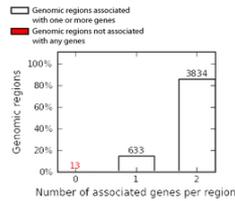
Host	Group	Database	Track	Link for live analysis
genboree.org	GenboreeUser_group	GenboreeUser_database	Spark:Myeloid_HSC_C2	Click here

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

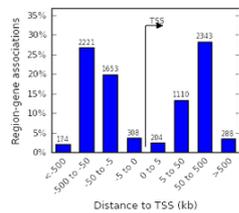
Job Description
Region-Gene Association Graphs

What do these graphs illustrate?

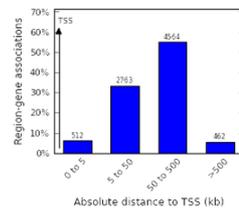
Number of associated genes per region
 Download as PDF.



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

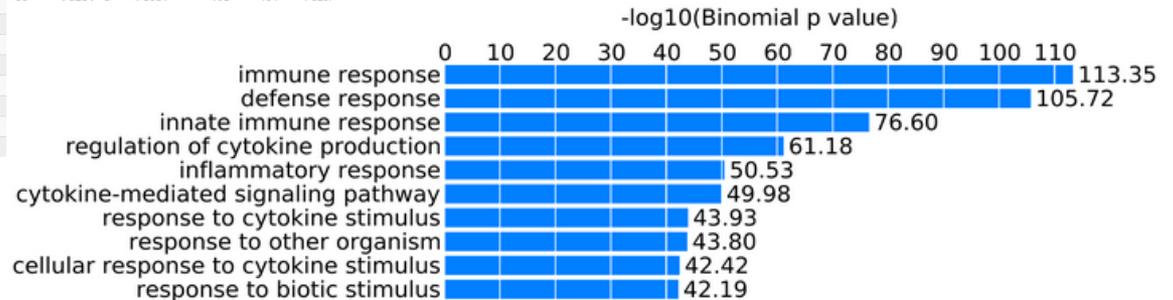
Visualize the table

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

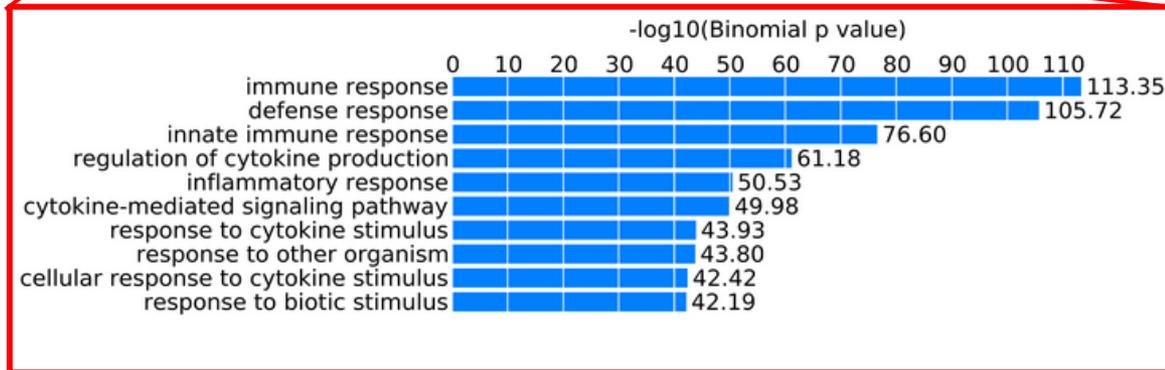
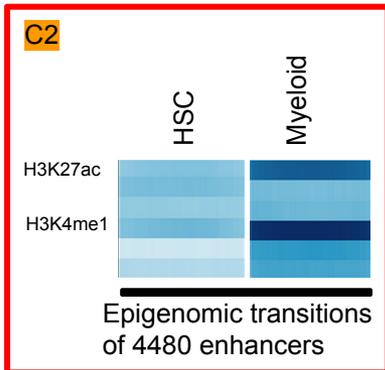
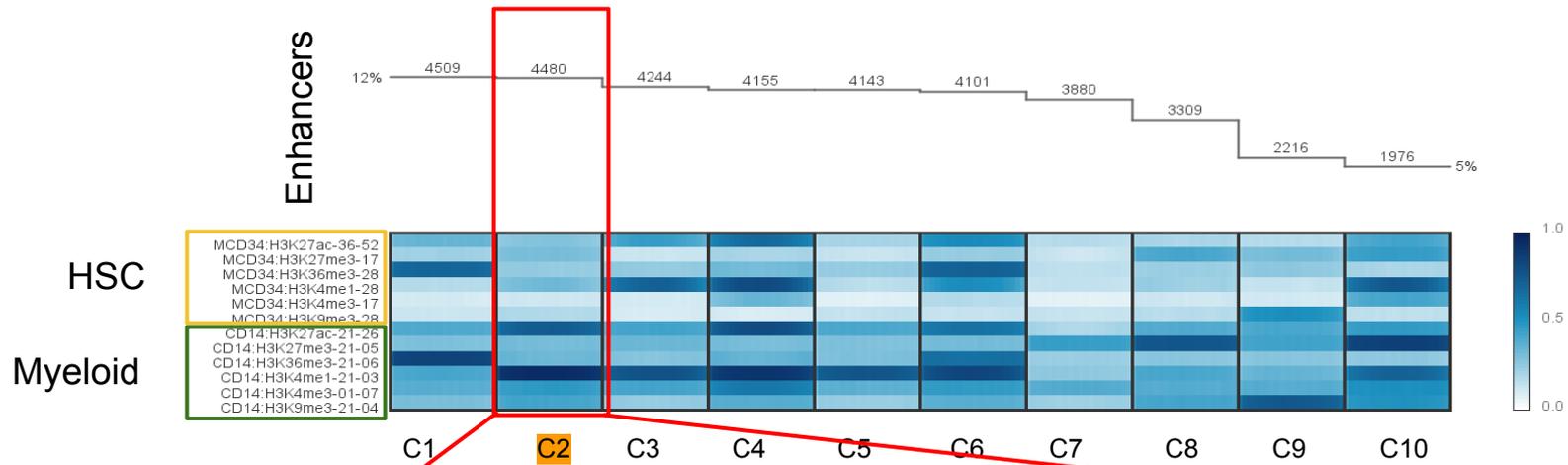
Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
immune response	1	4.4630e-114	3.9100e-110	2.5842	698	15.58%	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 8,761 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



Biological Process enrichment using GREAT

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 visible 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 of categories: 'User Profile', 'Groups', 'Hosts', 'Jobs', and 'Request Feature' (which is highlighted with a red box). Below 'Request Feature' is a list of tool categories including 'Atlas Tools Access', 'BRL AUTO TEST', 'EDACC', 'Epigenome Informatics Dem...', 'Epigenome Informatics Wor...', 'Epigenome ToolSet Demo In...', 'Epigenomics Road map Repo...', 'GenboreeUser_group', 'GMT_Tutorial', 'JonathanMill_Lab', 'Public', 'ROI Repository', 'Targeted Atlases', 'genboree.bcgsc.ca', 'genboree.cbrc.jp', and 'www.brain-research-lab.org'. The main content area shows a 'Genboree Workbench! [Getting Started]' header, a 'Data Filter: Select a filter...' dropdown, and a 'Details' table with columns 'Attribute' and 'Value'. The 'Details' table contains rows for 'View Link' (Link to Project), 'GenboreeUser_group', and 'Use_Case_18_GU'. A 'Request Feature' dialog box is open in the foreground, containing a 'Settings' tab, 'User Name' (Genboree User), 'User Email' (andrewj@bcm.edu), a 'Message' text area, and 'Submit' and 'Cancel' buttons. At the bottom of the page, there is a logo for 'BRL' (Baylor Research Laboratory) on the left, a footer stating 'Genboree is built & maintained by the Bioinformatics Research Laboratory at Baylor College of Medicine.' and 'Genboree is a hosted service. Code is available free for academic use.', and the 'HGSC HUMAN GENOME SEQUENCING CENTER' logo on the right.