Use Case 20: Exploring the role of epigenetics in myeloid cancer using reference epigenomes from the Human Epigenome Atlas

American Society of Human Genetics
Boston, MA

October 22, 2013

Presented by the
Bioinformatics Research Laboratory

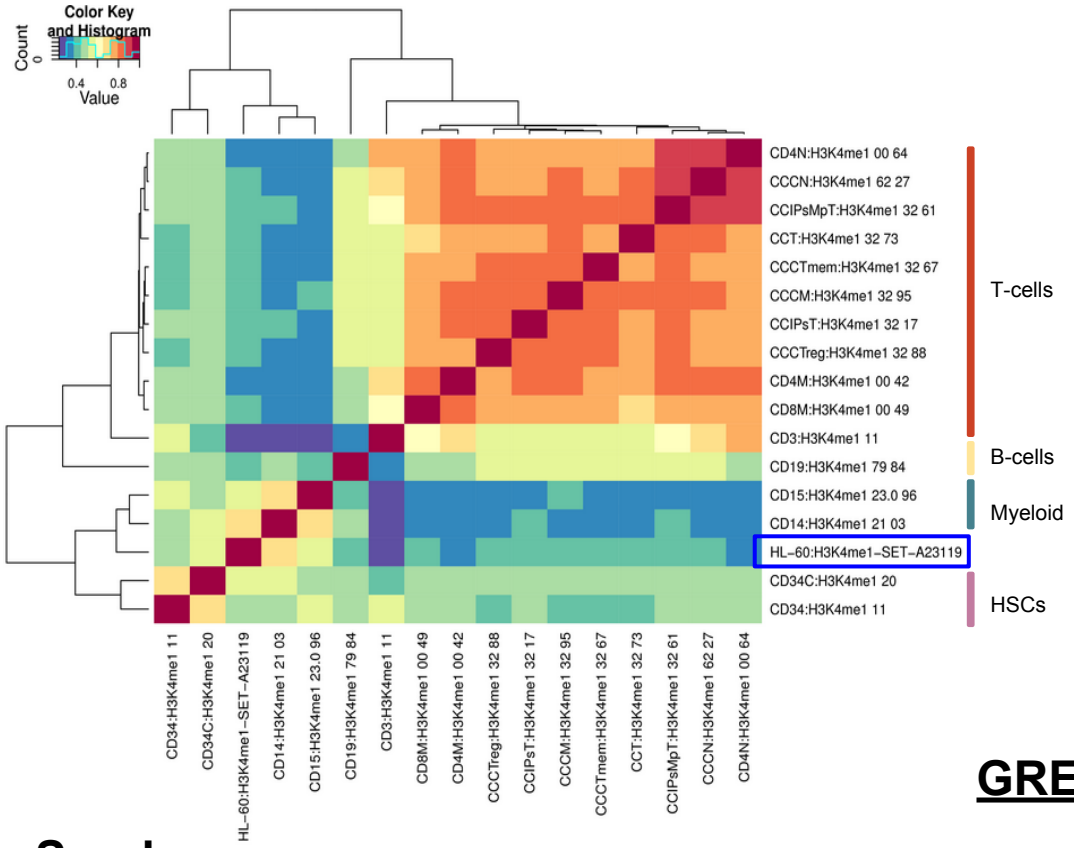
Baylor College of Medicine

Summary of Use Case 20

Background: Studying regions of epigenomic changes in normal versus disease state can help determine pathways that are involved in disease progression. In cancer, genomic locus are altered with copy number changes. Such changes can result in gain or loss of driver or suppressor genes. These changes not only alters the genome but also the epigenome. Studying changes in the epigenomic landscape of a tumor vs normal reference epigenomes can be used as marker to help determine pathways that are activated in disease progression. As a proof of principle, we sought to examine the functional significance of epigenomic changes in the human myeloid leukemia cell line HL60, by comparing it to reference epigenomes.

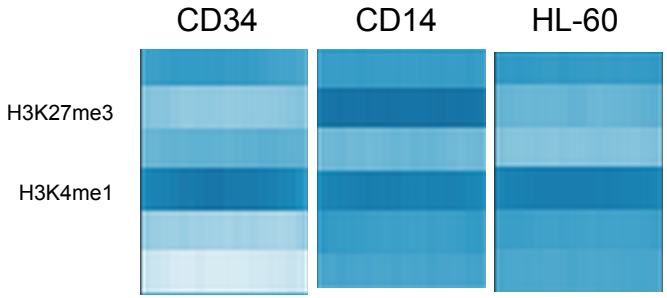
Results: Comparison of reference epigenomes with HL60 histone modifications correctly identify HL60 as being of myeloid origin, suggesting one may perform an initial epigenomic characterization of cell lines/tissues of unknown origin, or which have not been fully characterized, by virtual comparison to reference epigenomes. In addition, further computational approaches point to several pathways and transcriptional regulators previously shown to play a role in myeloid biology.

Summary of Results



Results: Comparison (heatmap) of reference epigenomes with HL60 histone modifications correctly identify HL60 as being of myeloid origin, suggesting one may perform an initial characterization of cell lines/tissues of unknown origin by virtual comparison to reference epigenomes. Spark tool was then used to identify 1025 enhancers that underwent epigenomic transition. GREAT tool was used to determine pathways cis-regulatory regions associated with these enhancers. C-MYB, previously known, transcriptional regulator was found to be significantly enriched.

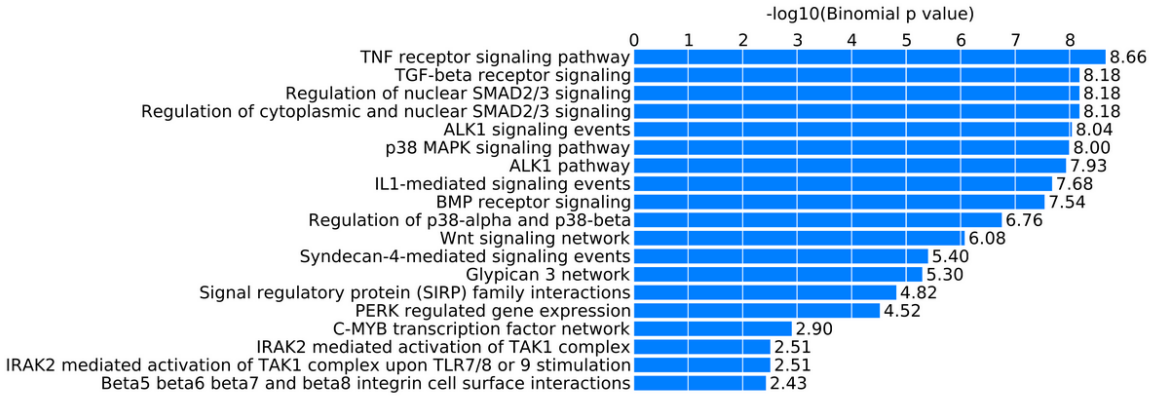
Spark



Epigenomic transitions of 1025 enhancers

GREAT

Pathway Commons



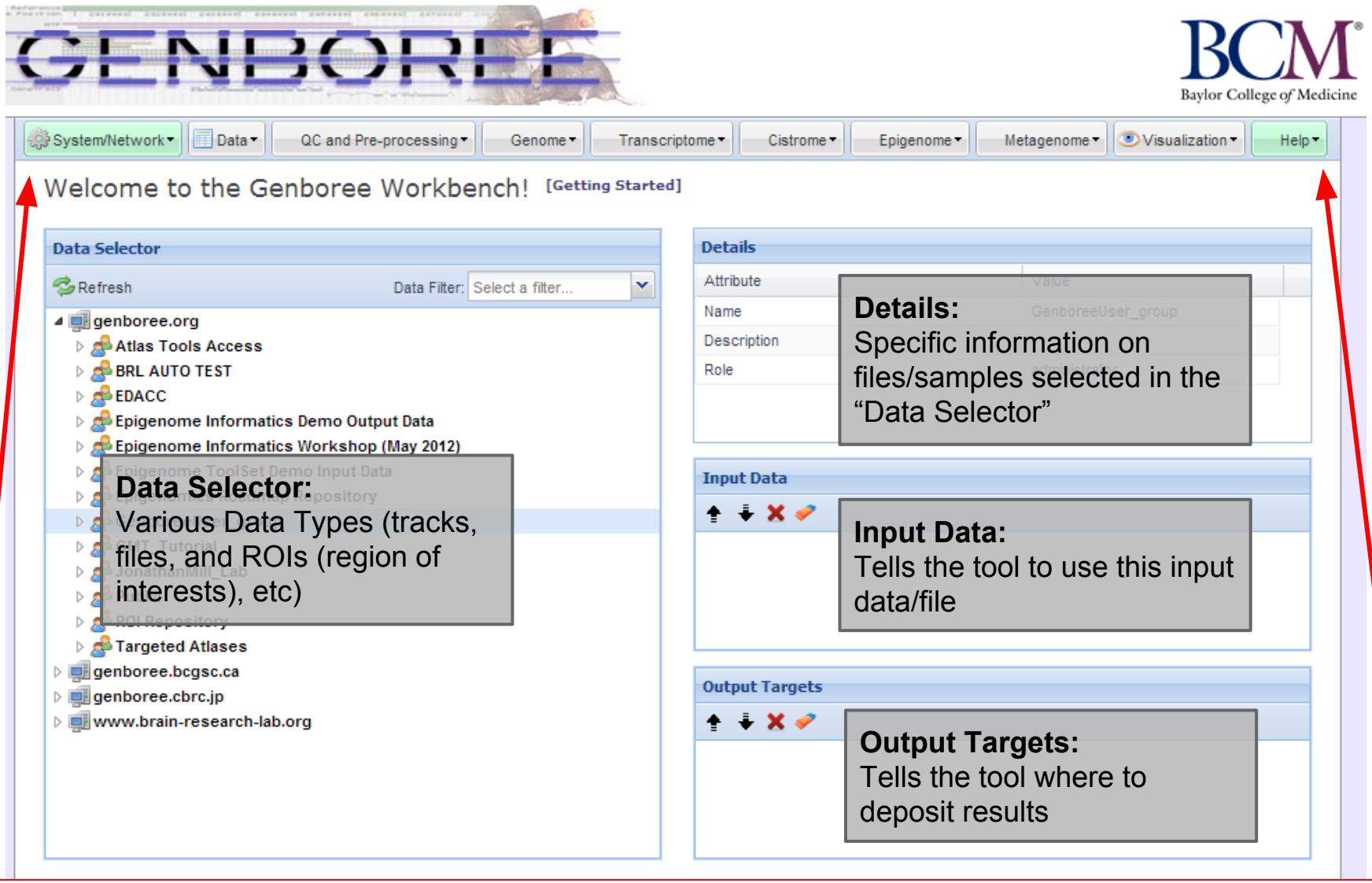
Use Case Overview

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

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

- Methodology
- Steps for reproducing the results

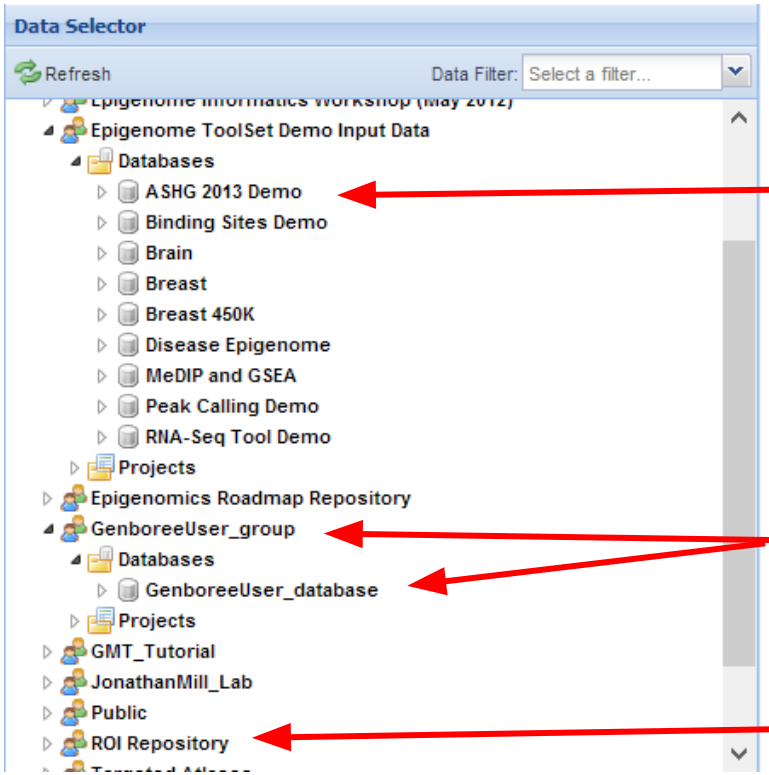
The Genboree Workbench: Web-based Data Management & Analysis



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

Preparation Prior to Starting the Use Case

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

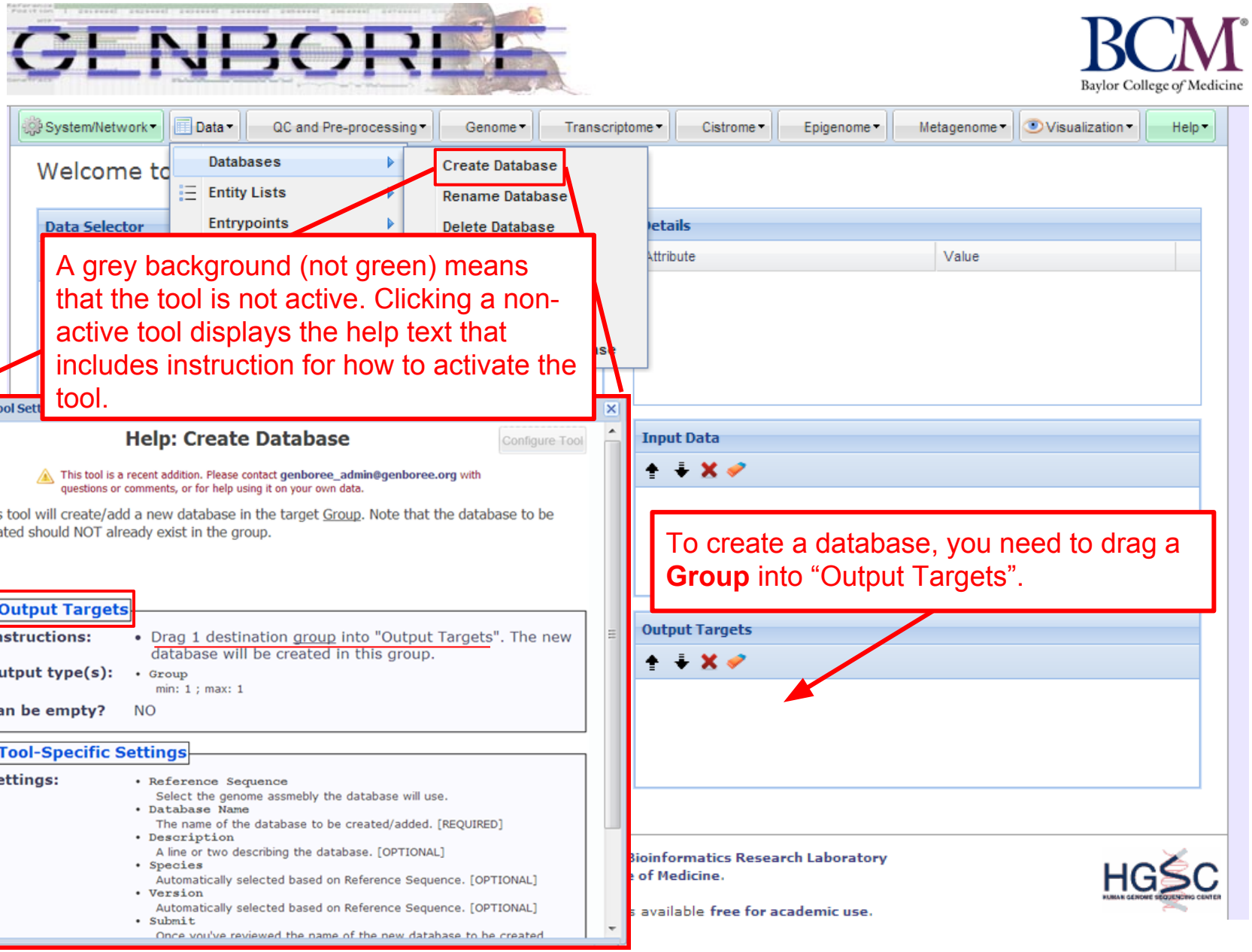


Under "Epigenome Toolset Demo Input Data" you will find "ASHG 2013 Demo" database, where we have provided you with sample data to try out the use cases

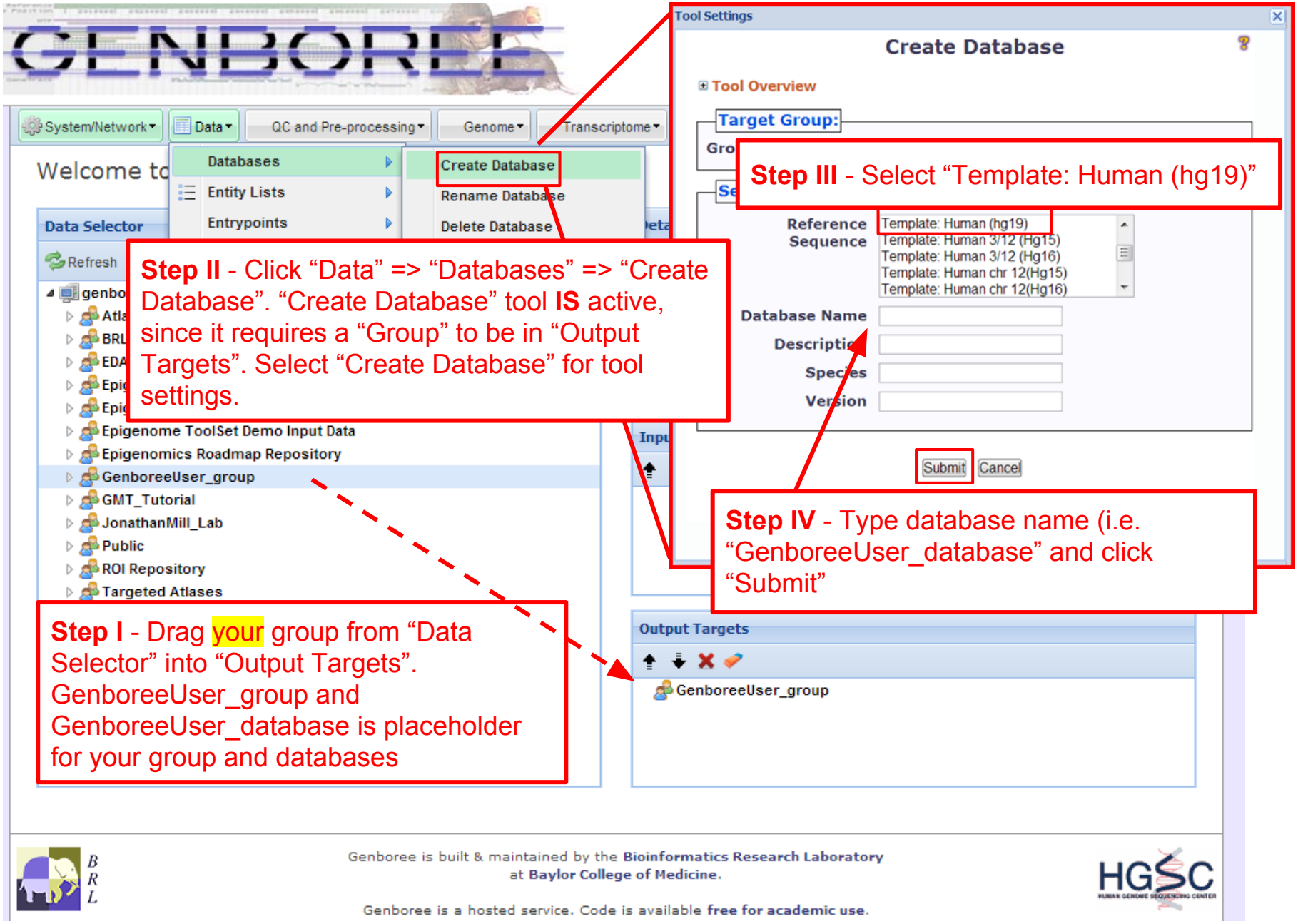
When making screenshots for providing instructions we have used "GenboreeUser_group" and "GenboreeUser_database" as output targets, however, you will use your own group and database. Following slides will show you how to create database and project

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

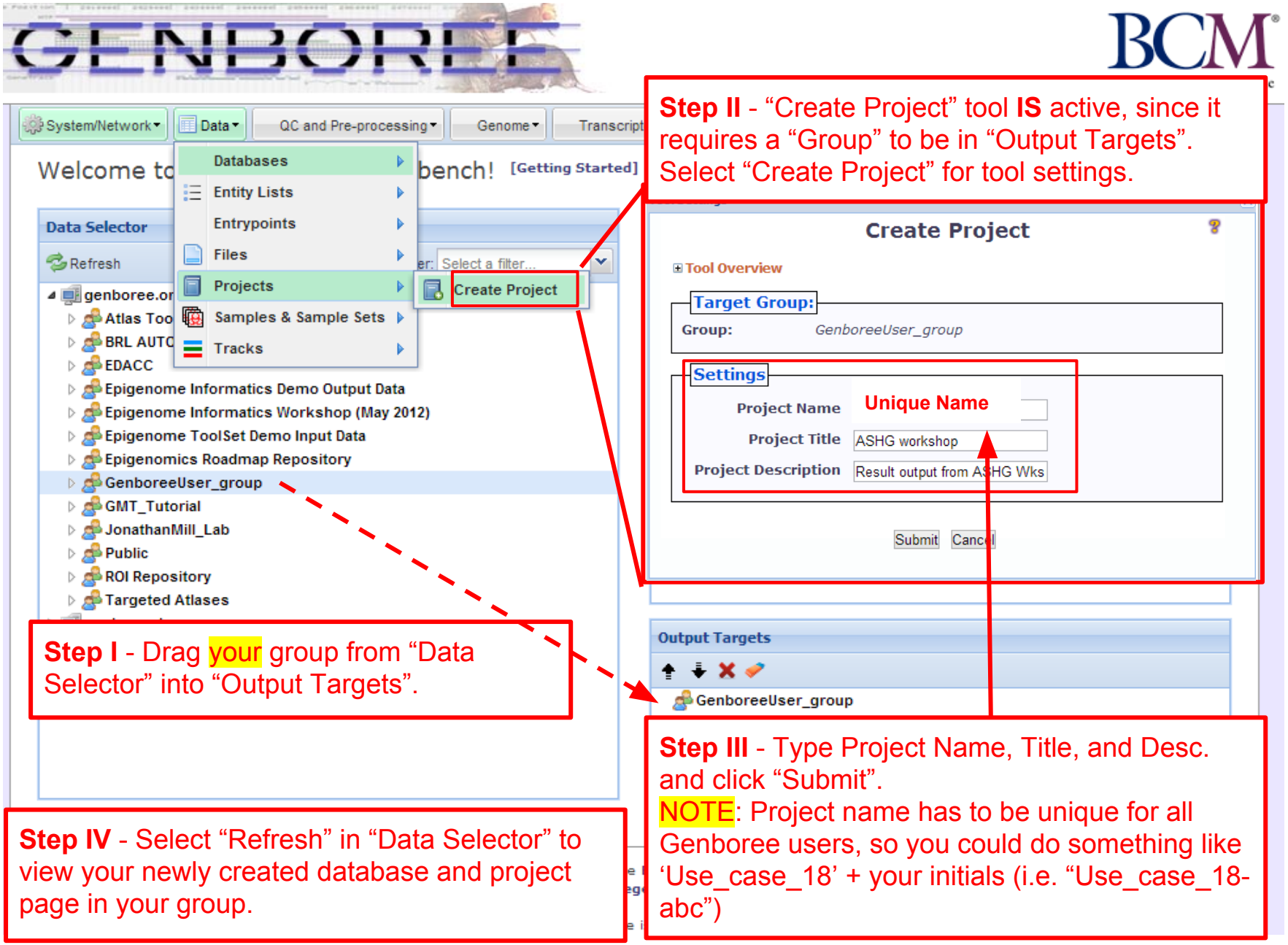
Display Tool Setting “Help” dialogue box in the Workbench



Steps for Creating a Database



Steps for Creating a Project page



Step I - Drag your group from "Data Selector" into "Output Targets".

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

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")

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

Description of Regions of Interest (ROIs) Tracks

Source of ROIs that will be used in this analysis:

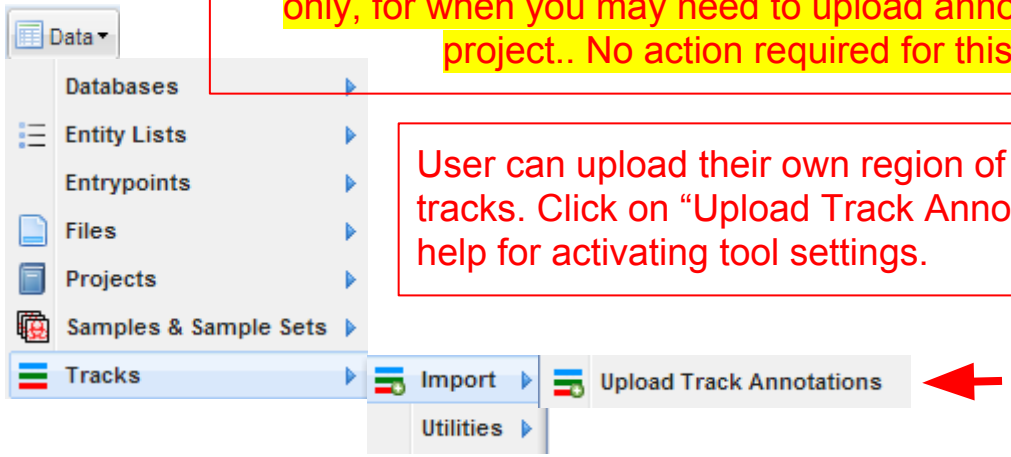
- **Track "GeneProteinCoding_promoter":** This track contains promoters of protein coding genes that were defined using Gencode V10 annotations (www.gencodegenes.org/), with transcription start sites (TSS) +/- 1500bp. The track contains 20,007 promoters from human genome build Hg19.

- **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 "release9_H3K4me3":** This track contains enhancers obtained from Manolis Kellis. Enhancers here are defined as H3K4me3 marks from the NIH Roadmap Consortium data (www.epigenomeatlas.org), wherein the coordinates were defined by ChromHMM.[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.

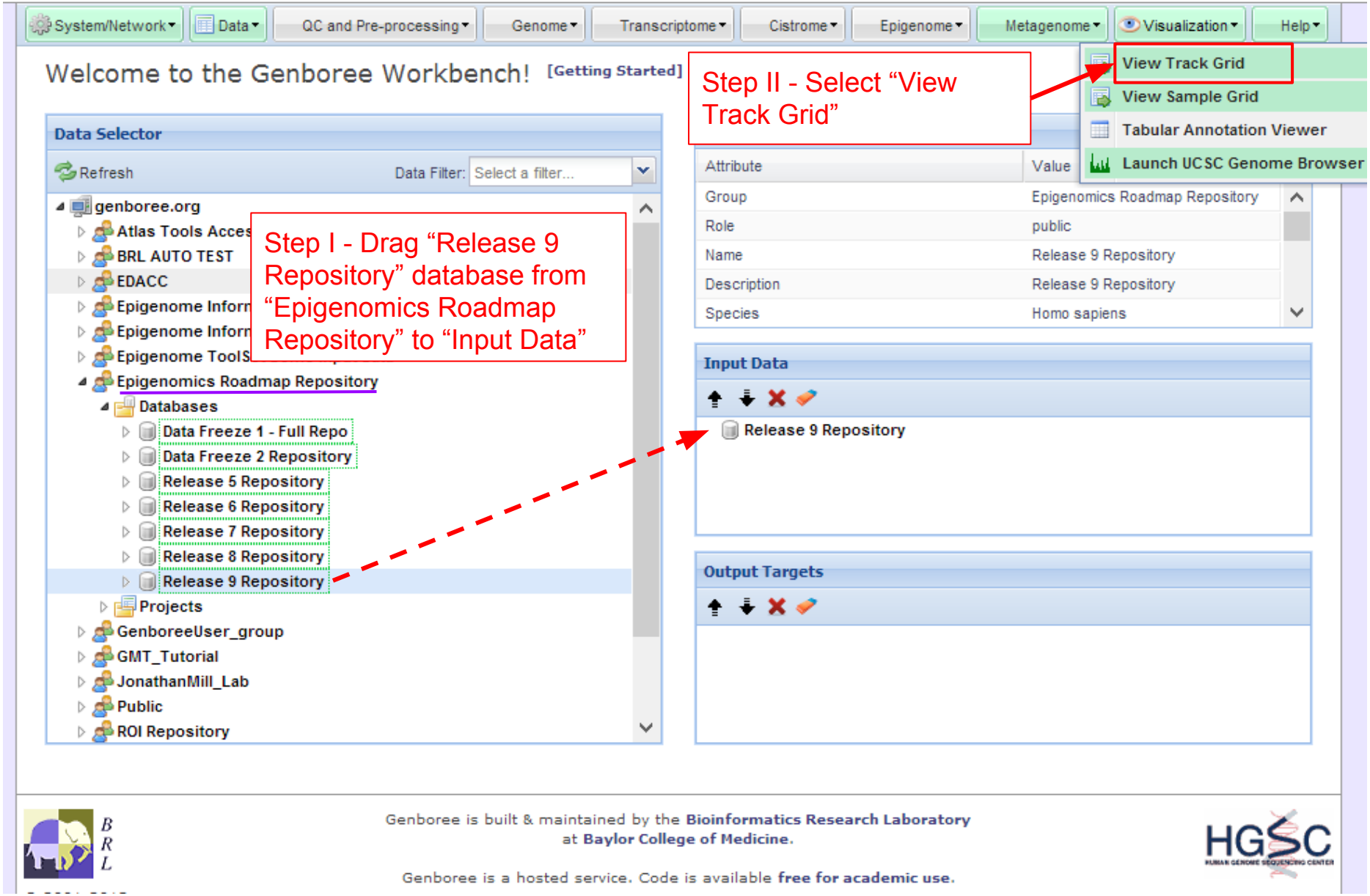
Upload your own Track Annotations:



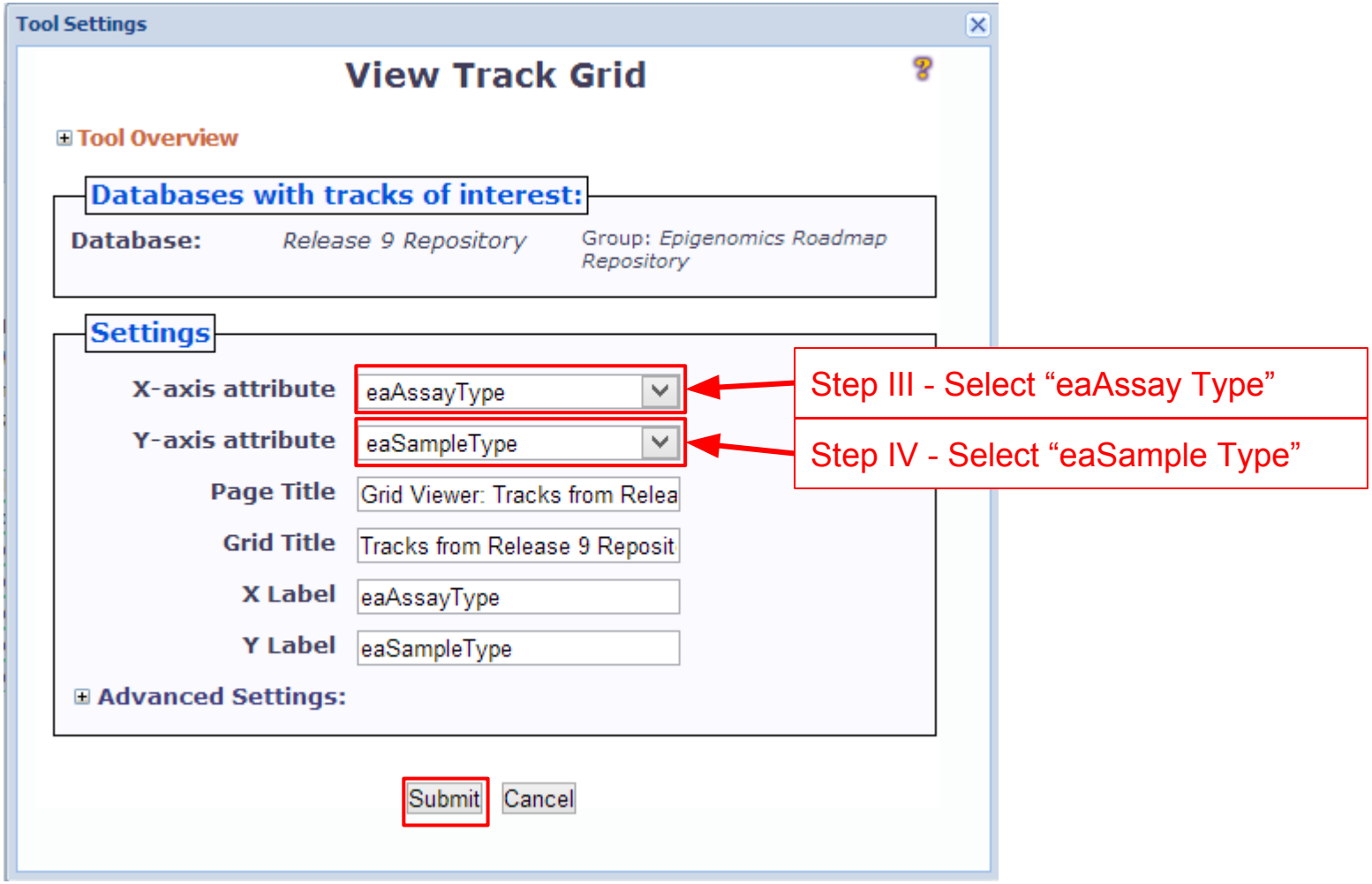
User can upload their own region of interests tracks. Click on "Upload Track Annotations" to see help for activating tool settings.

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

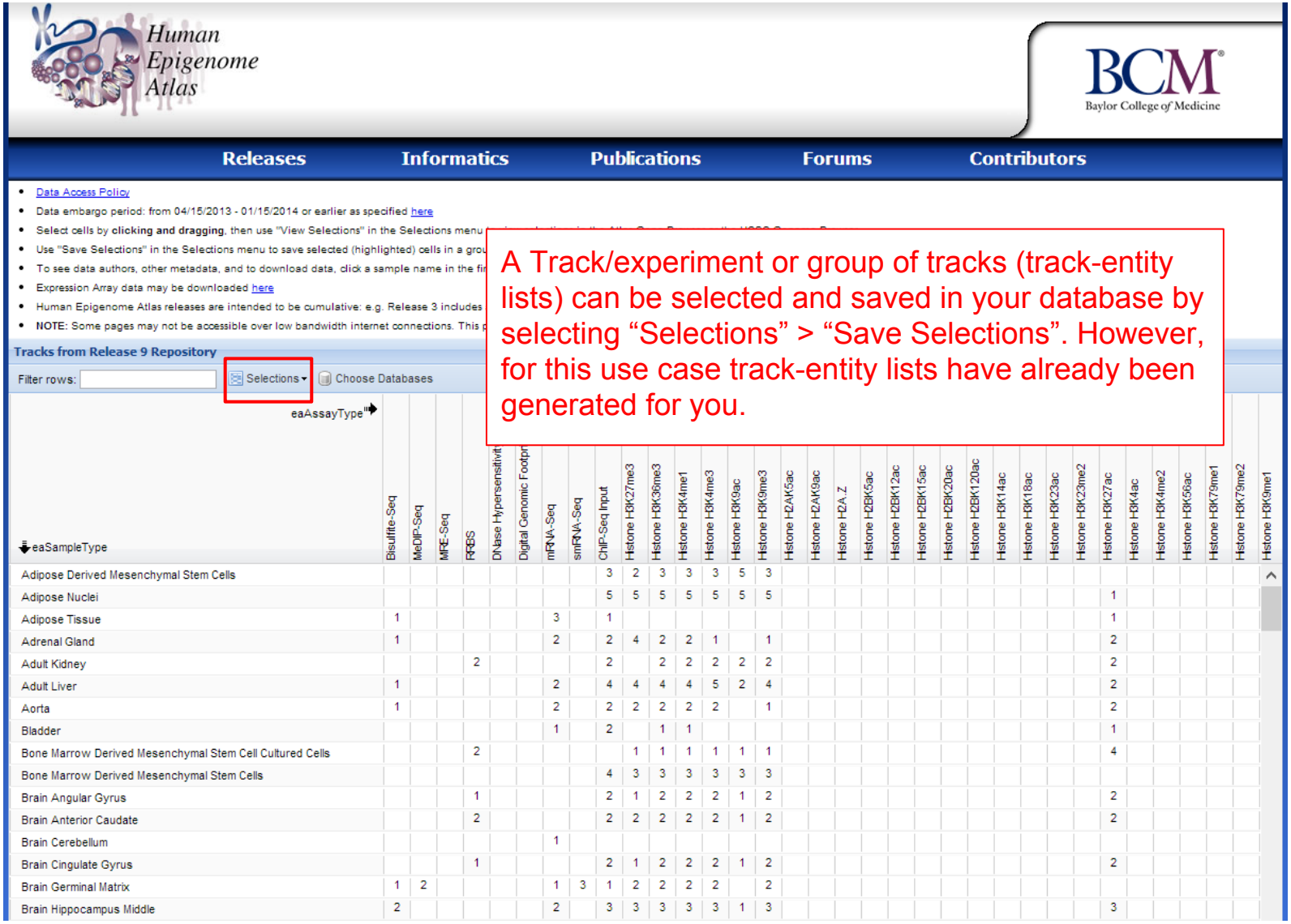
Follow these steps to view Track Grid of data from the Roadmap Epigenome Project



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



Track Grid view of the data from Roadmap Epigenome Project



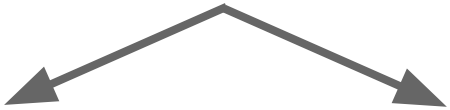
eaSampleType \ eaAssayType	Bisulfite-Seq	MeDIP-Seq	MRE-Seq	RRBS	DNase Hypersensitivit	Digital Genomic Footpr	mRNA-Seq	smRNA-Seq	ChIP-Seq Input	Histone H3K27me3	Histone H3K36me3	Histone H3K4me1	Histone H3K4me3	Histone H3K9ac	Histone H3K9me3	Histone H2AK5ac	Histone H2AK9ac	Histone H2A.Z	Histone H2BK5ac	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																			
Brain Hippocampus Middle	2						2		3	3	3	3	3	1	3													3						

Methodology Overview

Clustering/Heatmap: select your experimental tracks (provided here as HL60) and epigenomes for comparison from the Human Epigenome Atlas to find closest reference epigenome

LIMMA: to find regions with differentially modified histone signals between two groups of data tracks.

Spark: visualizes epigenomic profiles on a genome-wide scale by clustering or collapsing regions with similar “epigenomic footprint”



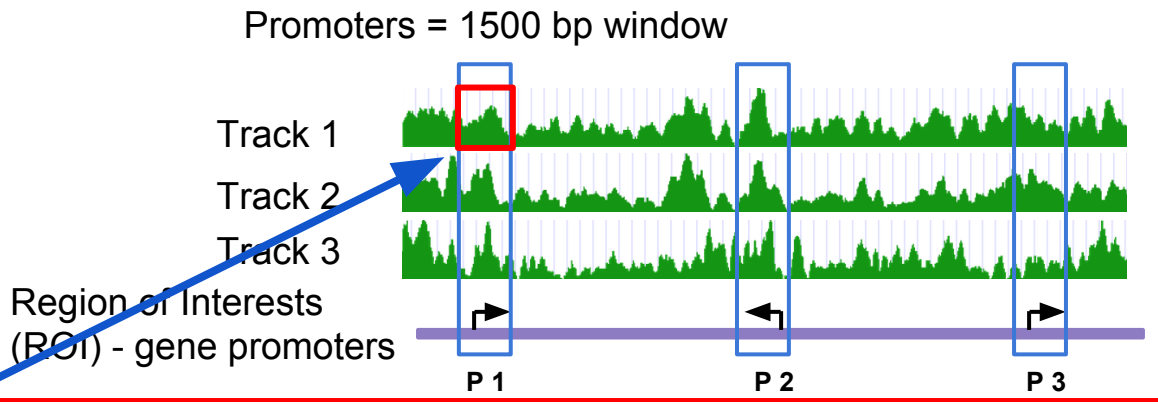
GREAT: assesses functional significance of cis-regulatory regions.

HOMER: de novo motif discovery.

Use Case 19
Supplemental Slides

Methodology: Clustering/Heatmap

I. Data Selection



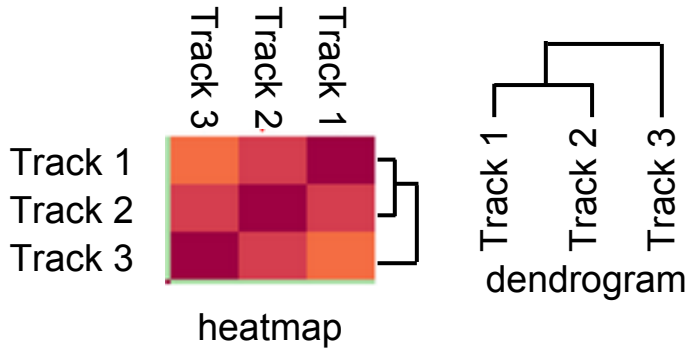
II. Signal processing- Heatmap Tool calculates average signal for each track and ROI (promoter) combination

	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

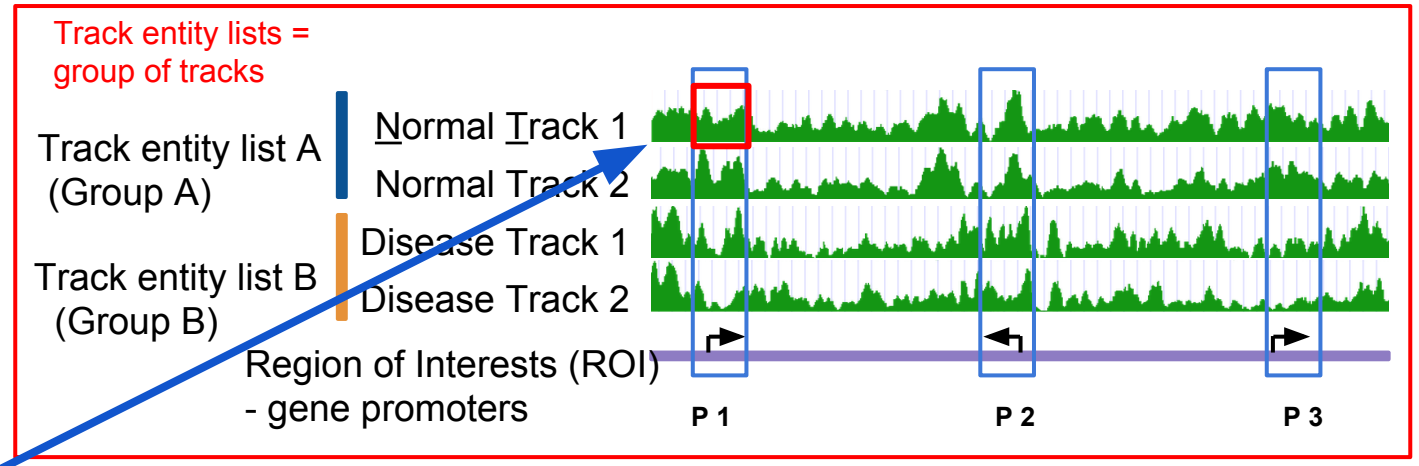
Correlation values are dynamically scaled and represented in heatmap.



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.

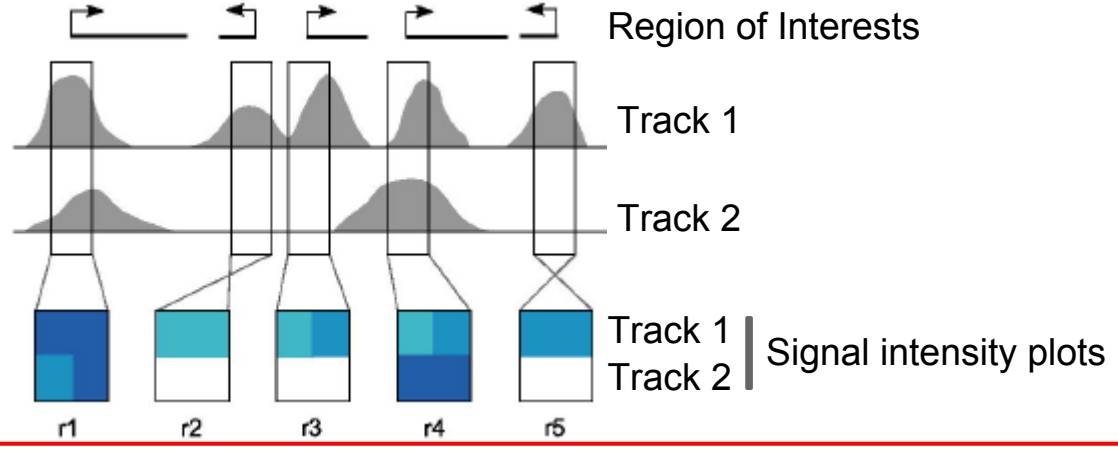
III. Results output

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

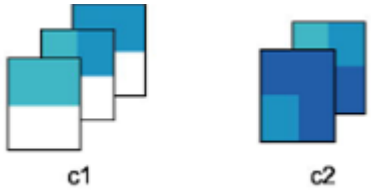
Methodology: Spark

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

I. Preprocessing
(ROIs = r1 - r5)

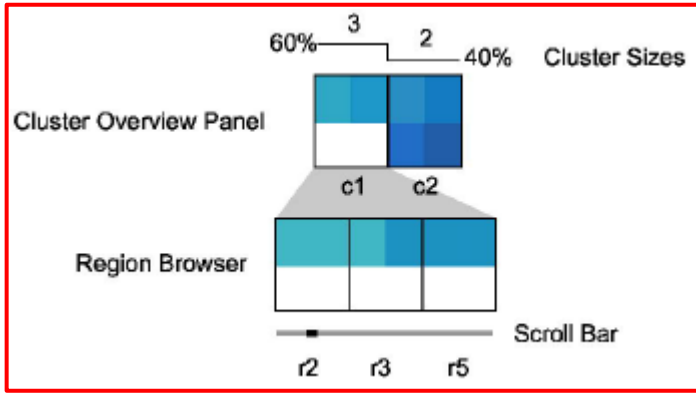


II. Clustering (C1, C2)



Regions with similar intensity plots are clustered. K-means clustering is used.

III. Interactive Visualization



Users can then interactively split clusters into subclusters to drill down and focus on specific biological patterns of interest: active vs poised promoters, for example

Nielson et. al., Gen Res 2012

Methodology: GREAT & HOMER

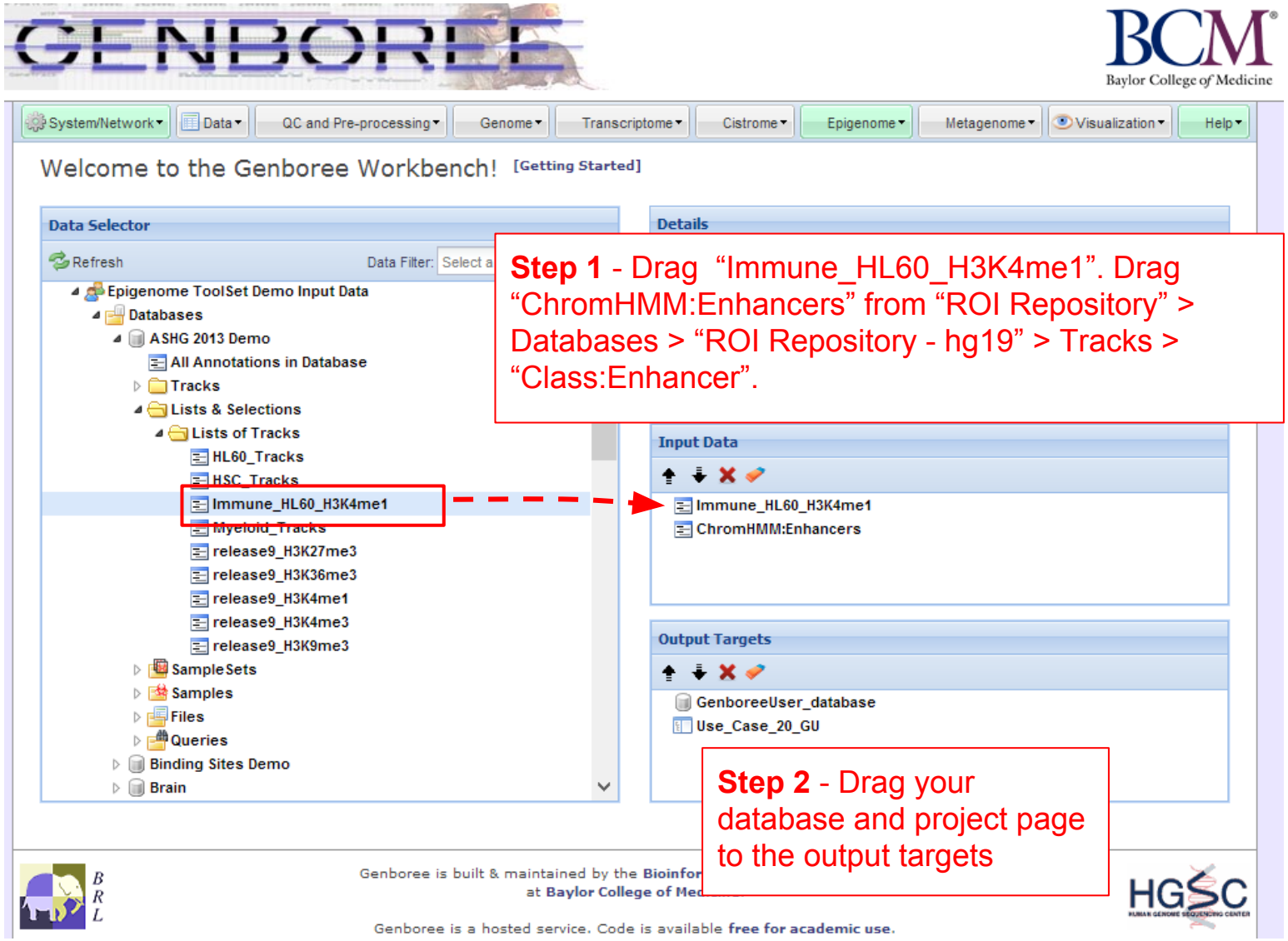
The patterns identified by Spark require further investigation to assess biological meaning

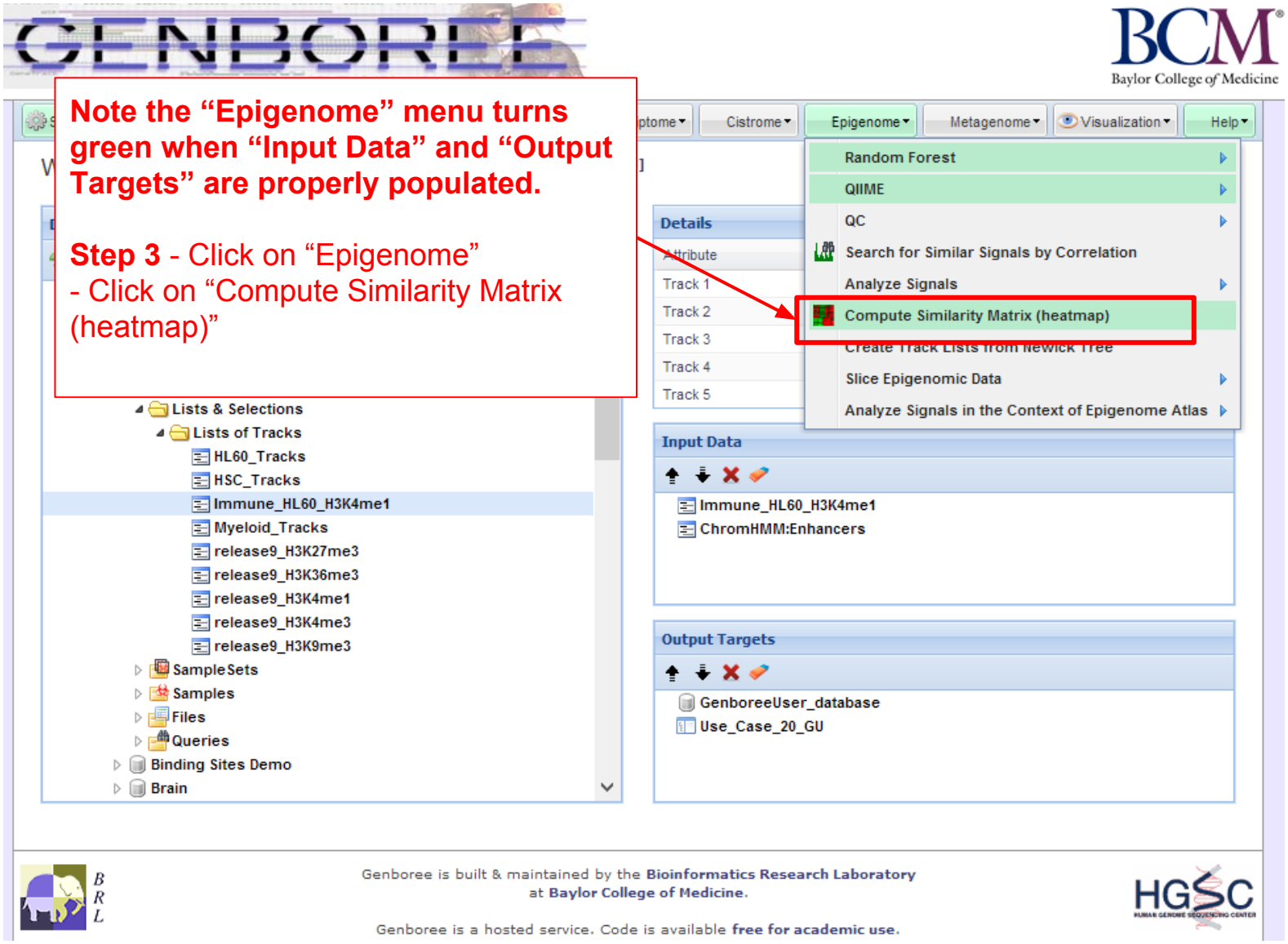
- GREAT[2] (Genomic Regions Enrichment of Annotations Tool) analysis
  - Assesses functional significance of cis-regulatory regions, here identified by Spark
  - Associates both proximal and distal binding events (Chip-Seq) with putative target genes
  - Uses gene annotation from several ontologies to 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

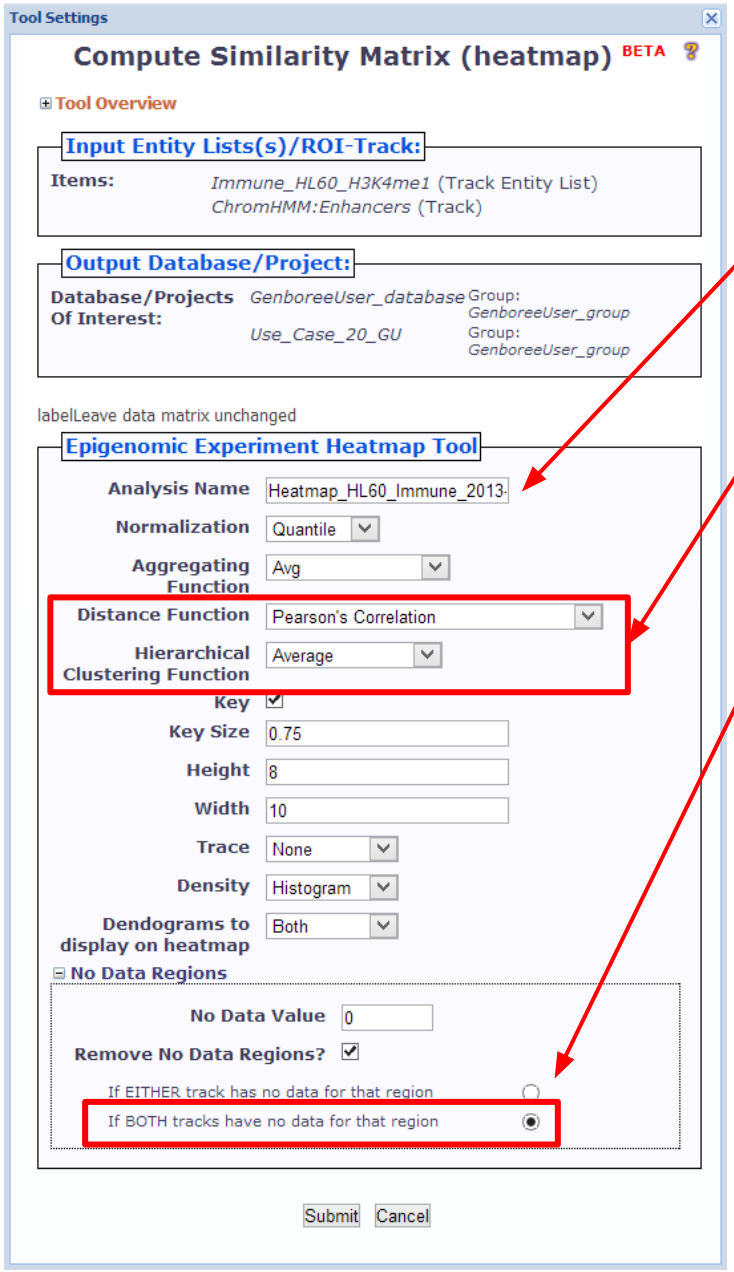
- HOMER[3] analysis:
  - Utilizes a de novo motif discovery algorithm
  - Scores motifs by looking for motifs with differential enrichment between two sets of sequences
  - Enrichment measured using the cummulative hypergeometric distribution (or cummulative binomial distribution for large data sets)
  - 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

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

[3]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).


GENBOREE
BCM
Baylor College of Medicine
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:
Epigenome ToolSet Demo Input Data
Databases
ASHG 2013 Demo
All Annotations in Database
Tracks
Lists & Selections
Lists of Tracks
HL60_Tracks
HSC_Tracks
Immune_HL60_H3K4me1
Myeloid_Tracks
release9_H3K27me3
release9_H3K36me3
release9_H3K4me1
release9_H3K4me3
release9_H3K9me3
SampleSets
Samples
Files
Queries
Binding Sites Demo
Brain
Details
Input Data
Immune_HL60_H3K4me1
ChromHMM:Enhancers
Output Targets
GenboreeUser_database
Use_Case_20_GU
Step 1 - Drag “Immune_HL60_H3K4me1”. Drag “ChromHMM:Enhancers” from “ROI Repository” > Databases > “ROI Repository - hg19” > Tracks > “Class:Enhancer”.
Step 2 - Drag your database and project page to the output targets
Genboree is built & maintained by the Bioinfor
at Baylor College of Med
Genboree is a hosted service. Code is available free for academic use.
HGSC


GENBOREE
BCM
Baylor College of Medicine
Note the "Epigenome" menu turns green when "Input Data" and "Output Targets" are properly populated.
Step 3 - Click on "Epigenome"
- Click on "Compute Similarity Matrix (heatmap)"
Cistrome
Epigenome
Metagenome
Visualization
Help
Random Forest
QIIME
QC
Search for Similar Signals by Correlation
Analyze Signals
Compute Similarity Matrix (heatmap)
Slice Epigenomic Data
Analyze Signals in the Context of Epigenome Atlas
Details
Attribute
Track 1
Track 2
Track 3
Track 4
Track 5
Lists & Selections
Lists of Tracks
HL60_Tracks
HSC_Tracks
Immune_HL60_H3K4me1
Myeloid_Tracks
release9_H3K27me3
release9_H3K36me3
release9_H3K4me1
release9_H3K4me3
release9_H3K9me3
SampleSets
Samples
Files
Queries
Binding Sites Demo
Brain
Input Data
Immune_HL60_H3K4me1
ChromHMM:Enhancers
Output Targets
GenboreeUser_database
Use_Case_20_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.
BRL
HGSC

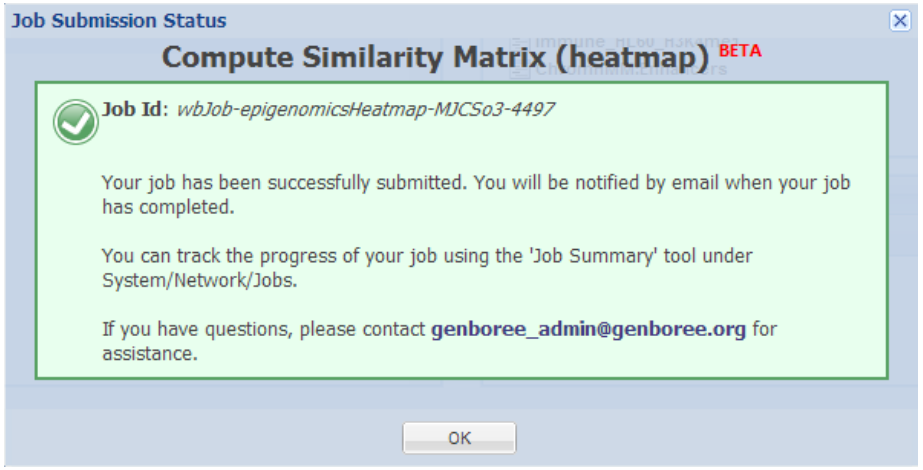


Step 4 - A default "Analysis Name" is generated by Genboree. It is recommended that all text and the time stamp be kept, and that you append some unique text to the beginning to help you distinguish different jobs run from the same tool.

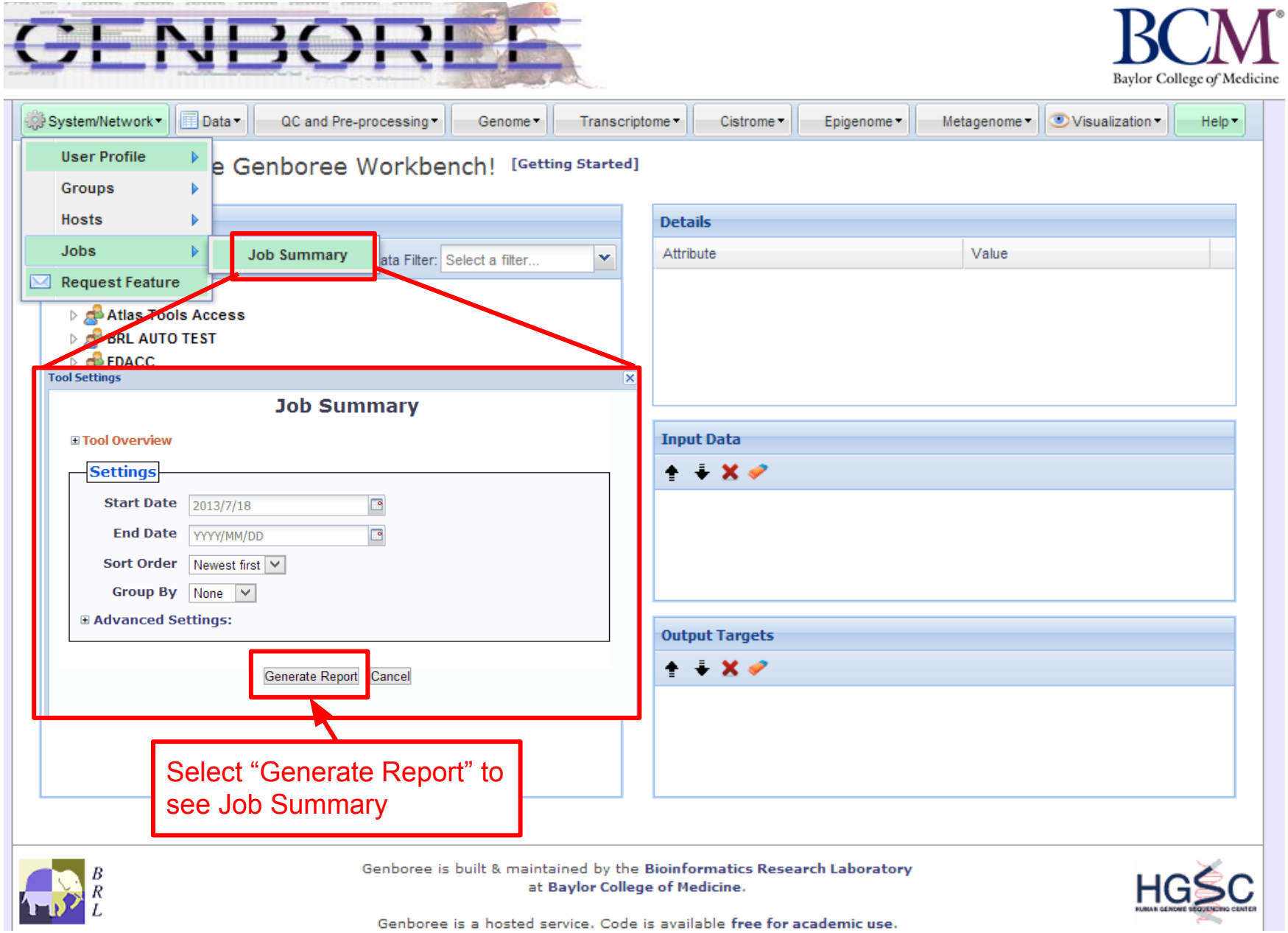
Step 5 - Select Pearson's Correlation as distance function and Average as hierarchical clustering function

Step 6 - Select to remove data if both tracks have no data for that region

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



Status of the jobs submitted can be obtained through Job Summary



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

Hello

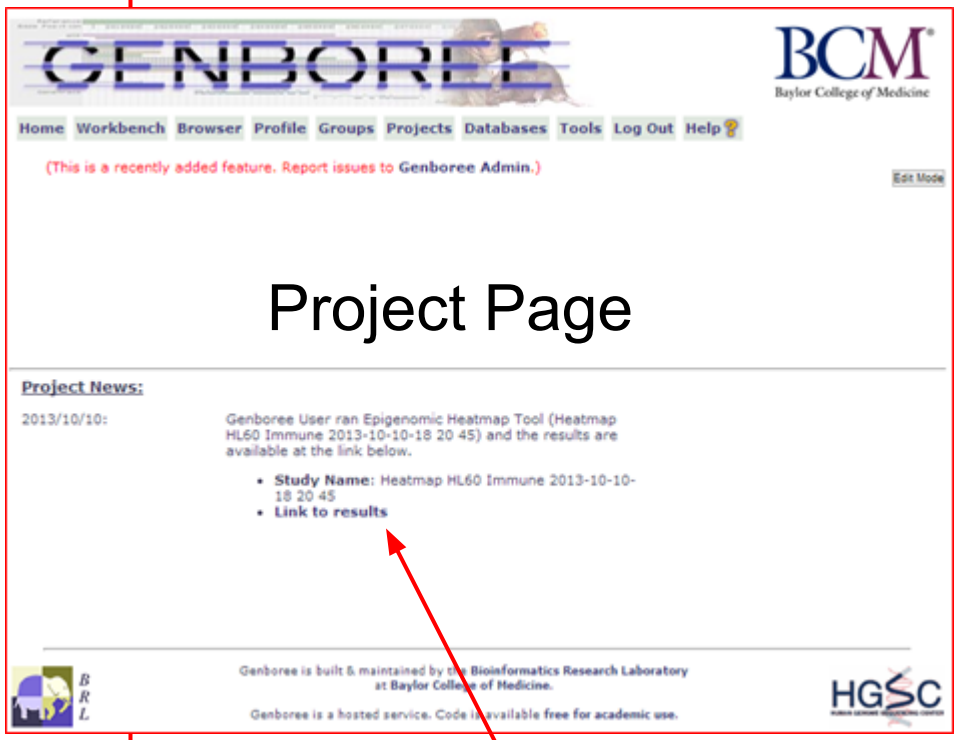
Your Compute Similarity Matrix (heatmap) job completed successfully.

Job Summary:
JobID - wbJob-epigenomicsHeatmap-e3CshO-5730
Analysis Name - EpigenomeExpHeatmap2013-10-09-00:42:01
Inputs:
1. Entitylist -
2. Trk -
3. Entitylist -
Outputs:
1. Db -
2. Prj -
Settings:
analysisName - EpigenomeExpHeatmap2013-10-09-00:42:01
dendograms - both
density - histogram
distfun - cor
hclustfun - average
height - 8
key - TRUE
keySize - 0.75
naG
nor
rem
rep
spa
trac
wid

- The Genboree Team

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

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



Click on the “Link to results” in your Project page

Table of Content: Epigenomic HeatMap

Study Name: Heatmap HL60 Immune 2013-10-10-18 20 45
User: Genboree User
Date: 2013/10/10 19:13 CDT

Epigenomic HeatMap Plots

Heatmap
Correlation plot

Click on the heatmap to see which reference epigenome profiled in the NIH Roadmap Epigenome Project does HL60 cluster with

Newick Tree Visualizations

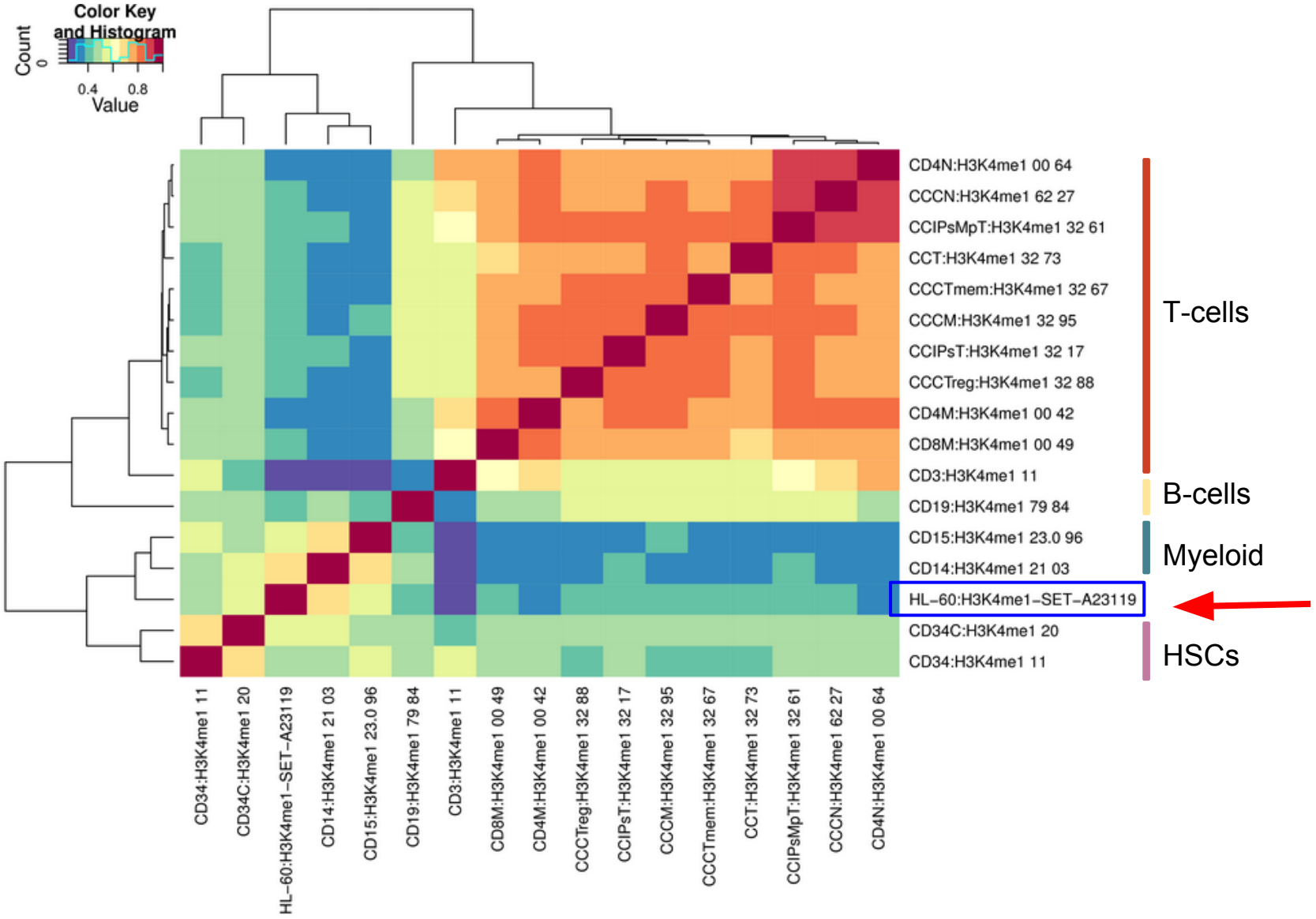
Equal Branch Lengths
Rows [PNG] [SVG]
Columns [PNG] [SVG]

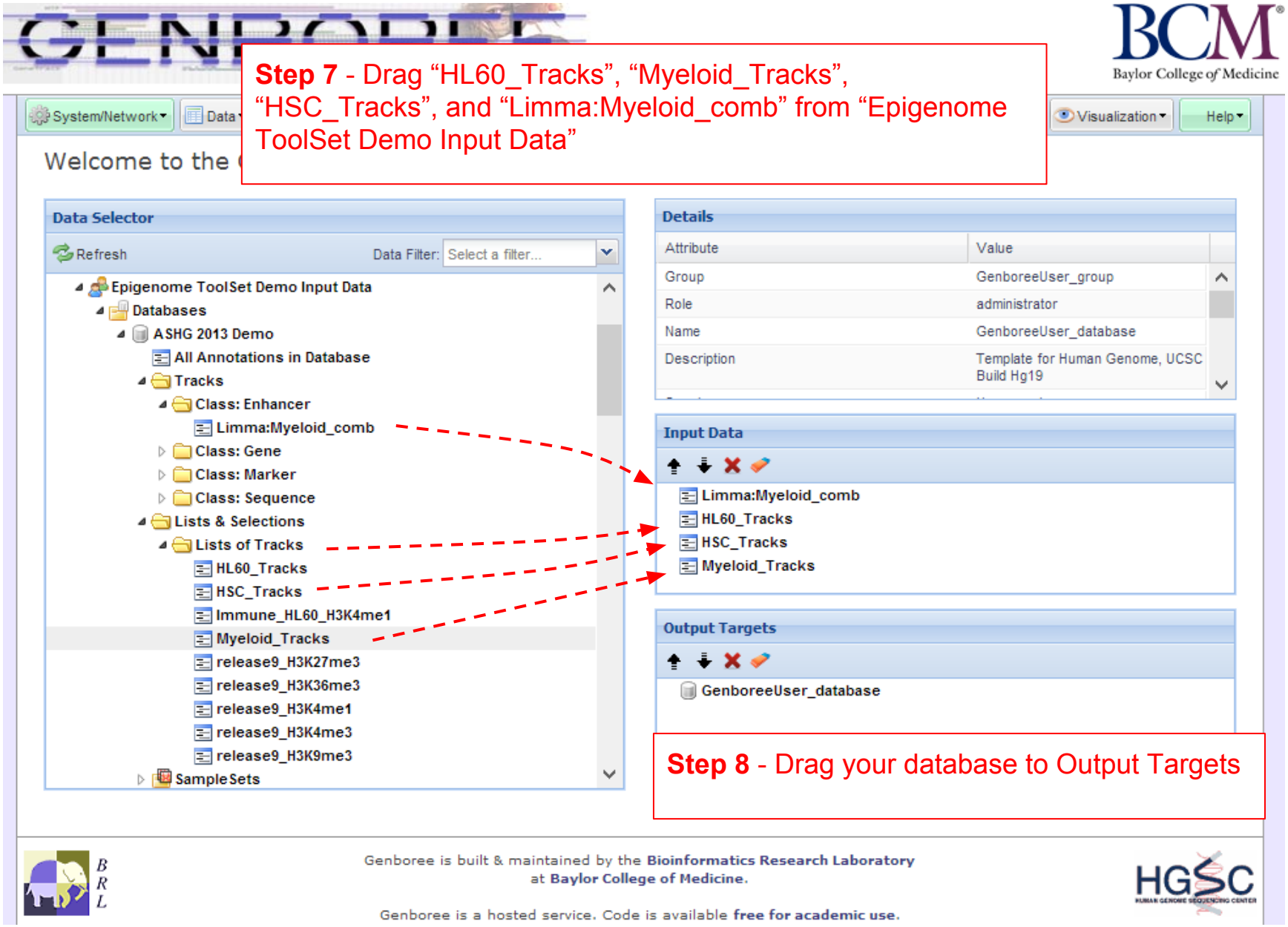
Scaled Branch Lengths
Rows [PNG] [SVG]
Columns PNG] SVG]

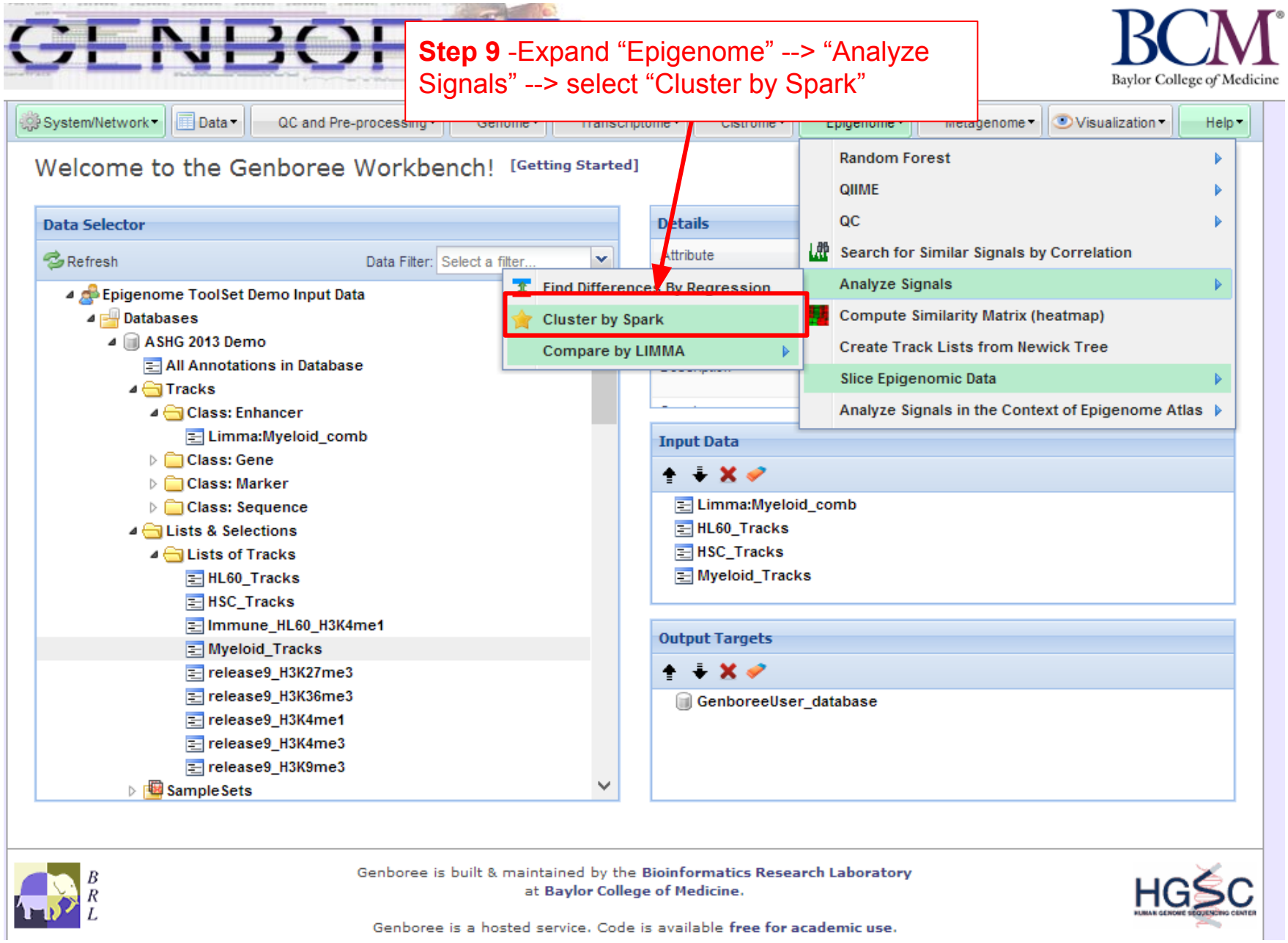
Natural Log Scaled Branch Lengths
Rows [PNG] [SVG]
Columns [PNG] [SVG]

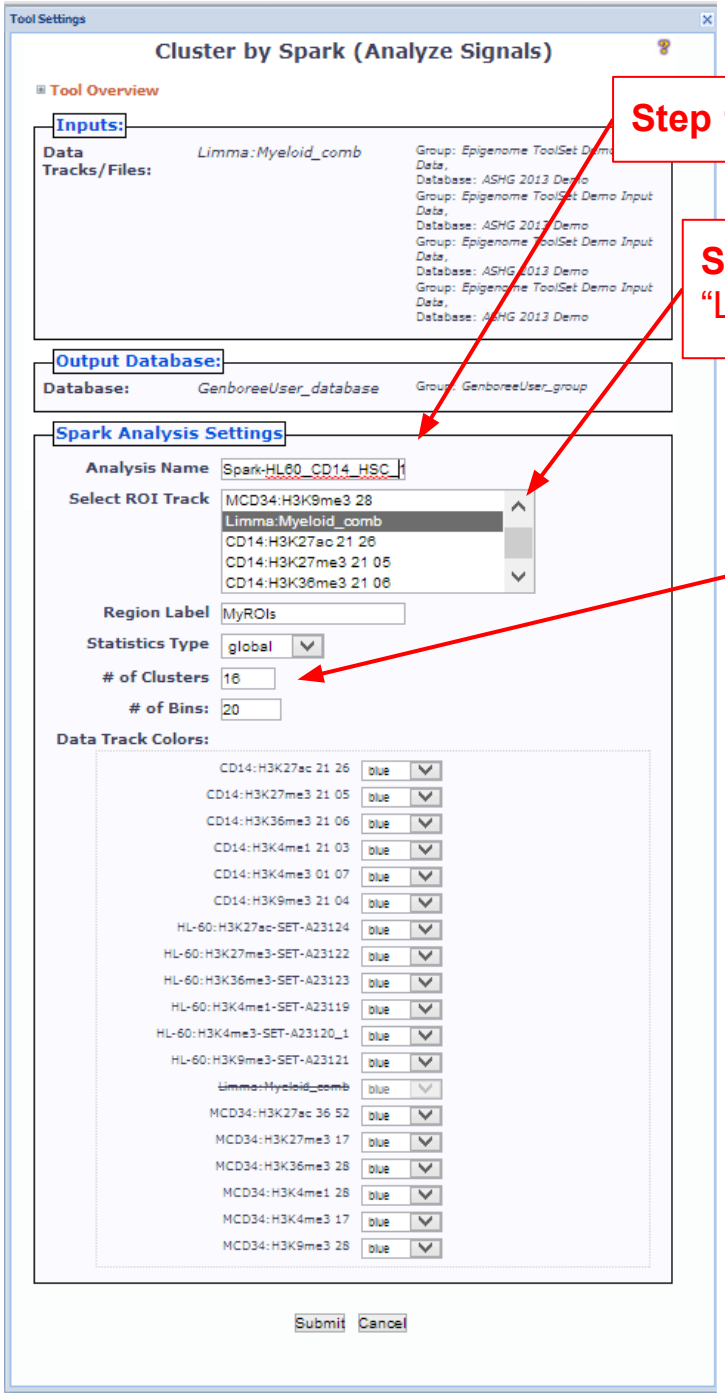
Log10 Scaled Branch Lengths
Rows [PNG] [SVG]
Columns [PNG] [SVG]

Heatmap results indicate that closest reference epigenomes for HL-60 is CD14 and CD15 of Myeloid lineage




Step 7 - Drag “HL60_Tracks”, “Myeloid_Tracks”, “HSC_Tracks”, and “Limma:Myeloid_comb” from “Epigenome ToolSet Demo Input Data”
BCM
Baylor College of Medicine
System/Network
Visualization
Help
Welcome to the
Data Selector
Refresh
Data Filter:
Select a filter...
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
SampleSets
Details
Attribute
Value
Group
GenboreeUser_group
Role
administrator
Name
GenboreeUser_database
Description
Template for Human Genome, UCSC Build Hg19
Input Data
Limma:Myeloid_comb
HL60_Tracks
HSC_Tracks
Myeloid_Tracks
Output Targets
GenboreeUser_database
Step 8 - Drag your database to 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


Step 9 -Expand “Epigenome” --> “Analyze Signals” --> select “Cluster by Spark”
BCM
Baylor College of Medicine
System/Network
Data
QC and Pre-processing
Visualization
Help
Welcome to the Genboree Workbench!
[Getting Started]
Data Selector
Refresh
Data Filter:
Select a filter...
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
SampleSets
Find Differences By Regression
Cluster by Spark
Compare by LIMMA
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
Details
Attribute
Input Data
Limma:Myeloid_comb
HL60_Tracks
HSC_Tracks
Myeloid_Tracks
Output Targets
GenboreeUser_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.
HGSC

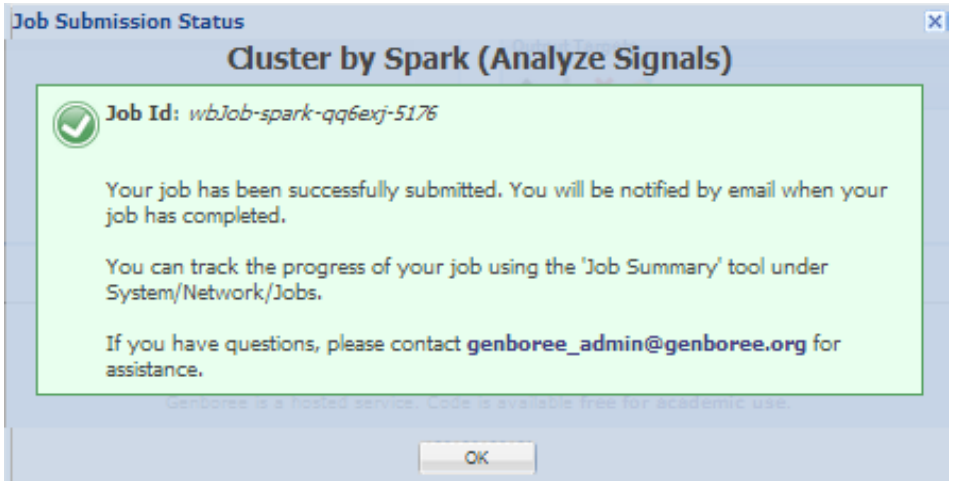


Step 10 - Add identifier text at beginning of analysis name

Step 11 - Select ROI (Region of Interest) Track "Limma:Myeloid_comb"

Step 12 - Enter the # of Clusters "16". During visualization, User can change number of clusters based on their expert knowledge.

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



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

Your Spark job completed successfully.

Job Summary:
JobID - wbJob-spark-AsqKlJ-9045
Analysis Name -
Inputs:
# of Data Tracks -
ROI Track -
Outputs:
Output DB -
Output Host - genboree.org
Settings:
k -
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.jnlp

Step 13 - Download Spark GUI

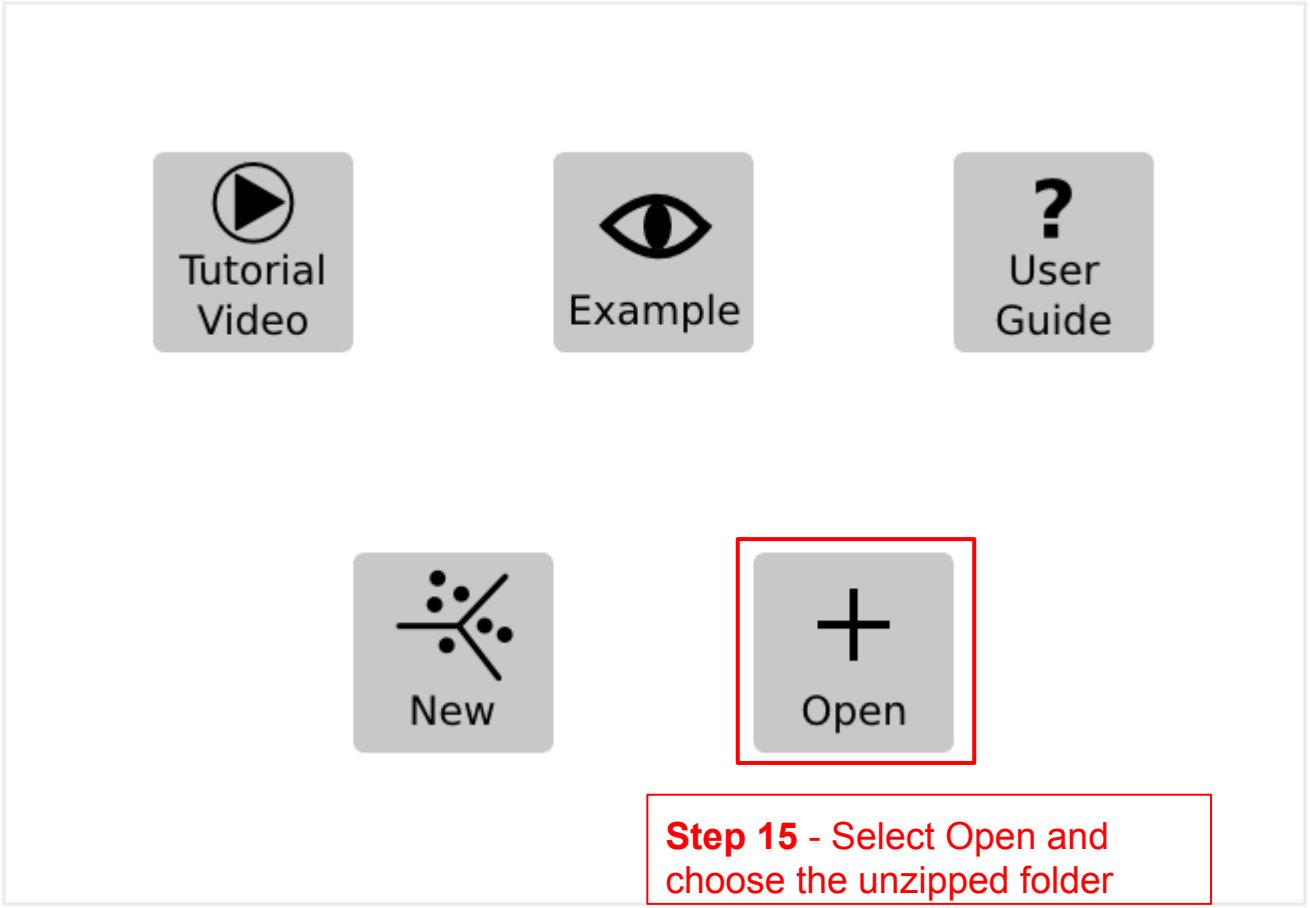
- The Genboree Team

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:
*
*
*

Step 14 - Download Spark Results and UNZIP the Folder

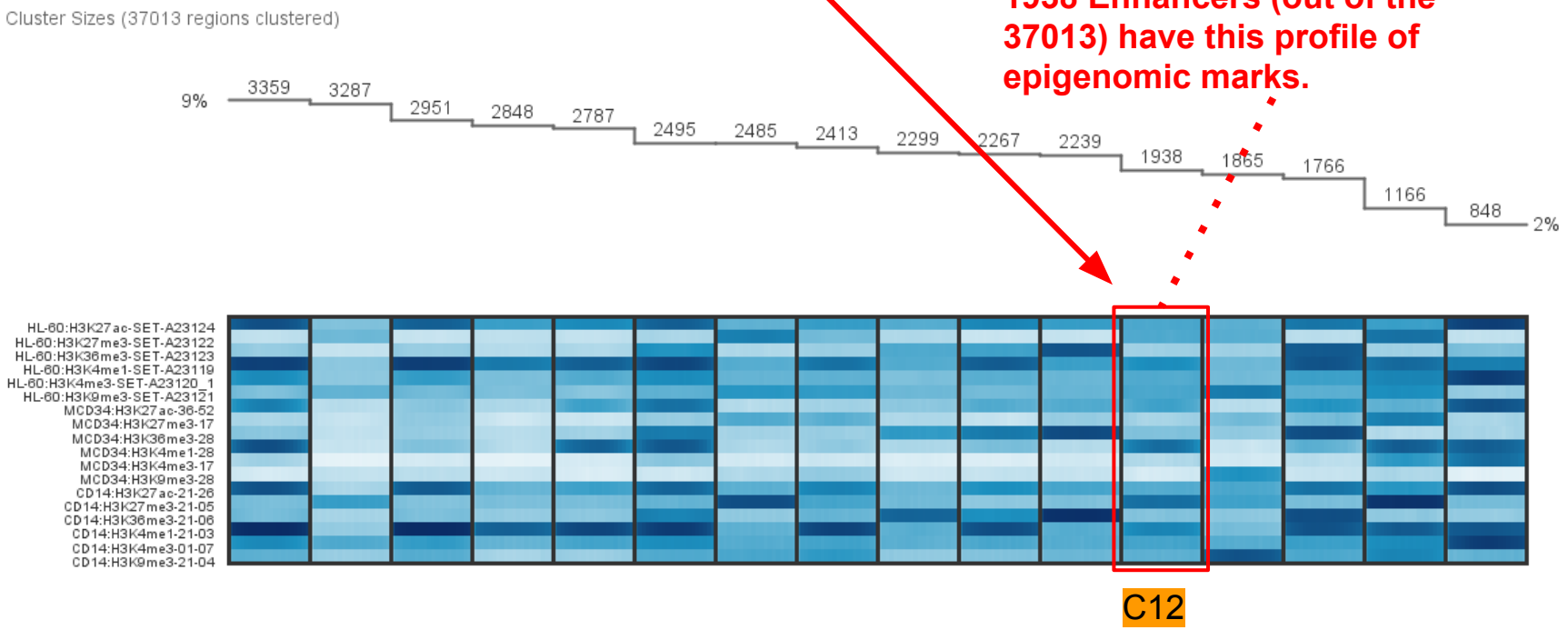
Result File URLs (click or paste in browser to access file):
FILE:
URL:
http://genboree.org/java-bin/apiCaller.jsp?rsrcPath=http%3A%2F%2Fgenboree.org%2FREST%2Fv1%2Fgrp%2Fvamin_group%2Fdb%2FUseCase%2Ffile%2FSpark%2520-%2520Results%2FSpark_Myeloid_HSC%2FSpark_Myeloid_HSC.zip%2Fdata%3F&fileDownload=true&promptForLogin=true&errorFormal=html

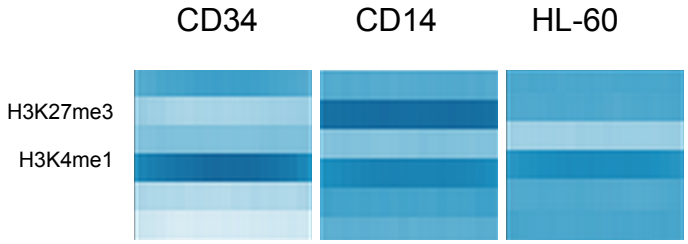
SPARK GUI


Tutorial Video
Example
User Guide
New
Open
Step 15 - Select Open and choose the unzipped folder

Step 16: Split the cluster to find enhancers that have active enhancer profile in HL60

1938 Enhancers (out of the 37013) have this profile of epigenomic marks.

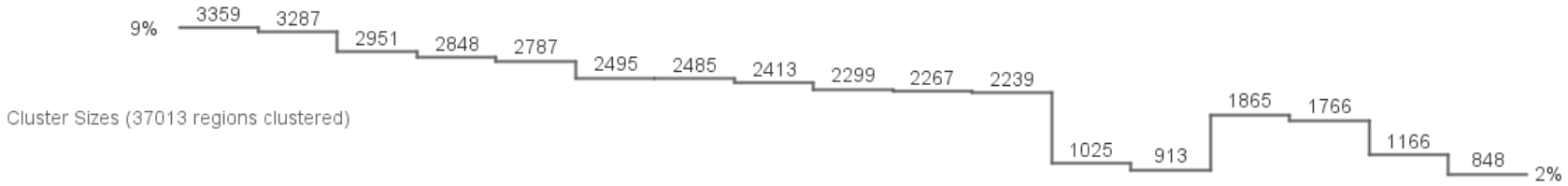


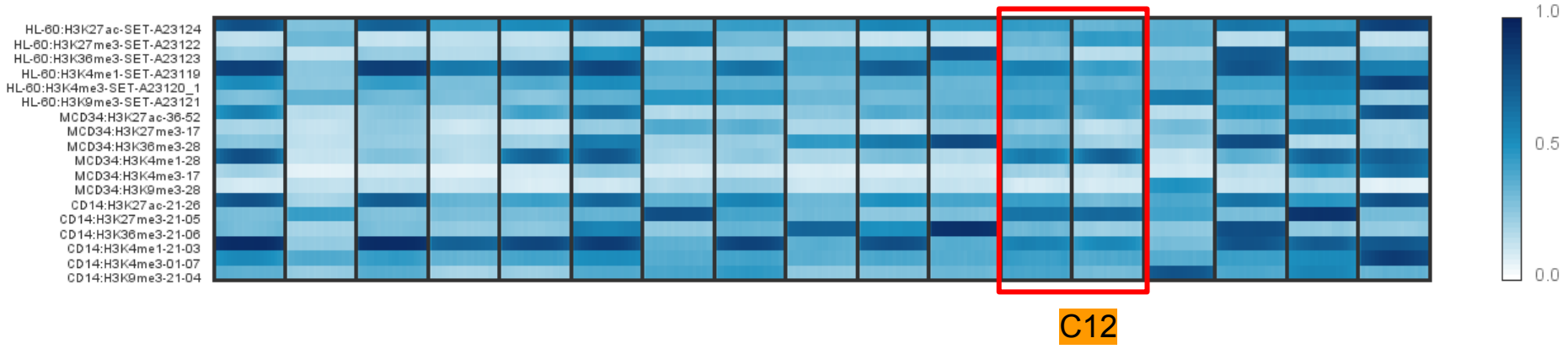


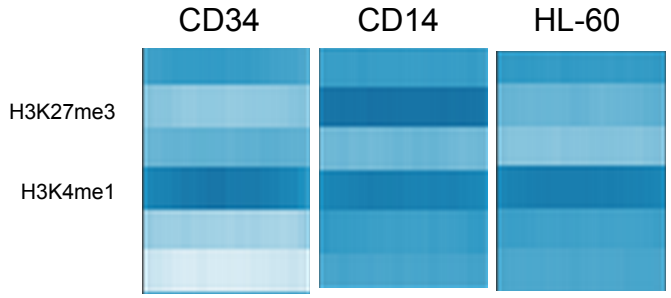
Epigenomic transitions of 1938 enhancers

Clustering (C12) reveals a shift from active epigenomic marks in human immune stem cells (MD34) to poised epigenomic marks in myeloid cells (CD14) and then active in HL-60

Upon splitting cluster C12 into two clusters, notice that the cluster with 1025 enhancer regions show a clear shift of HL60 in the active state (H3K4me1 and H3K27ac), while the cluster with 913 shows a poised state (H3K4me1 and H3K27ac).

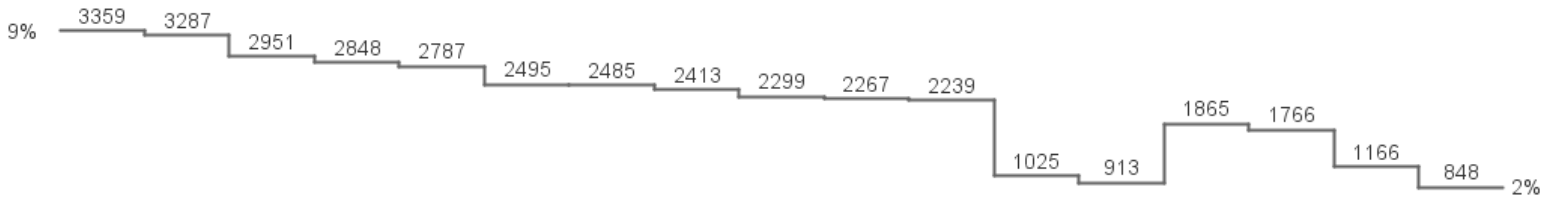


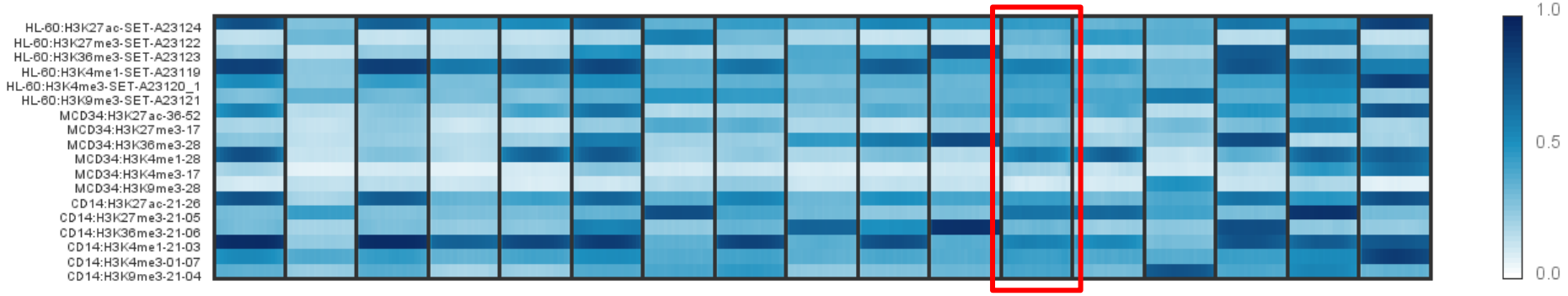




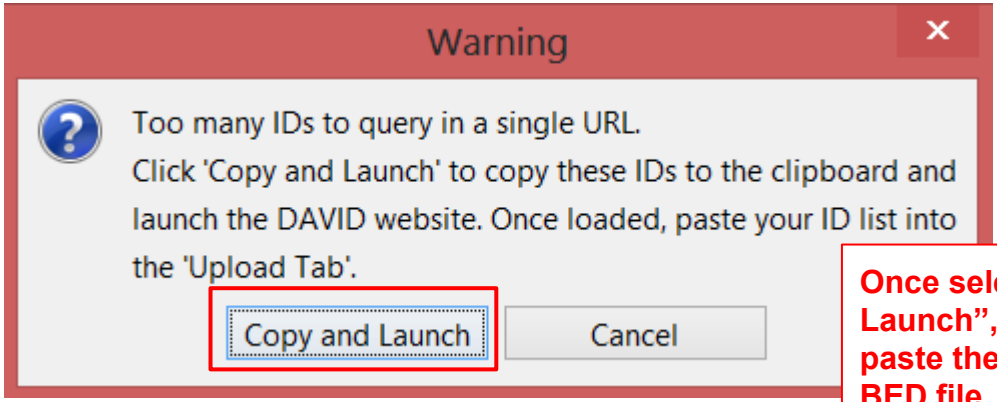
Epigenomic transitions of 1025 enhancers

The next few slides will walk you through some data manipulations in Excel which are required to generate BED files. BED files will supply the annotations for subequent steps

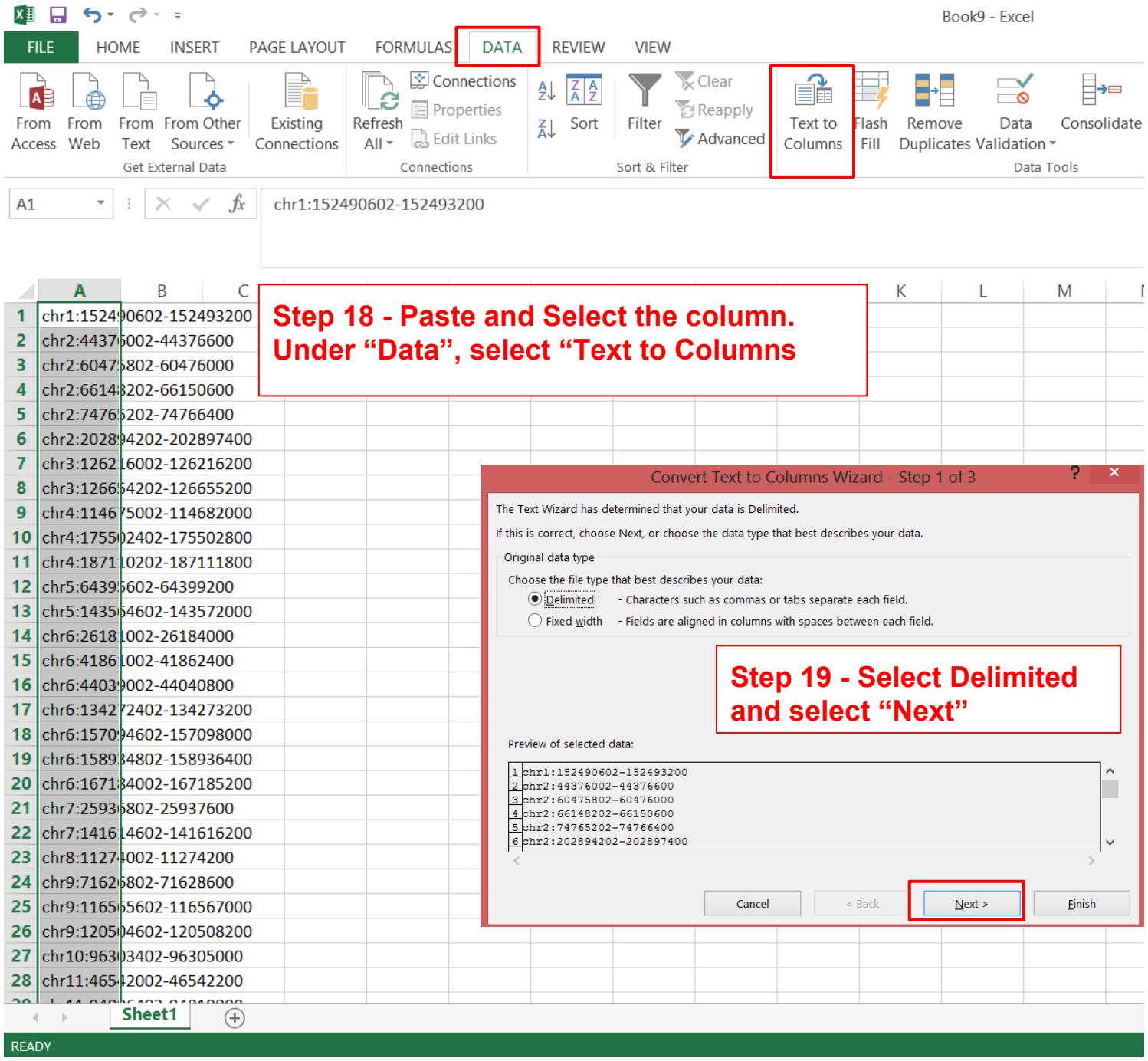


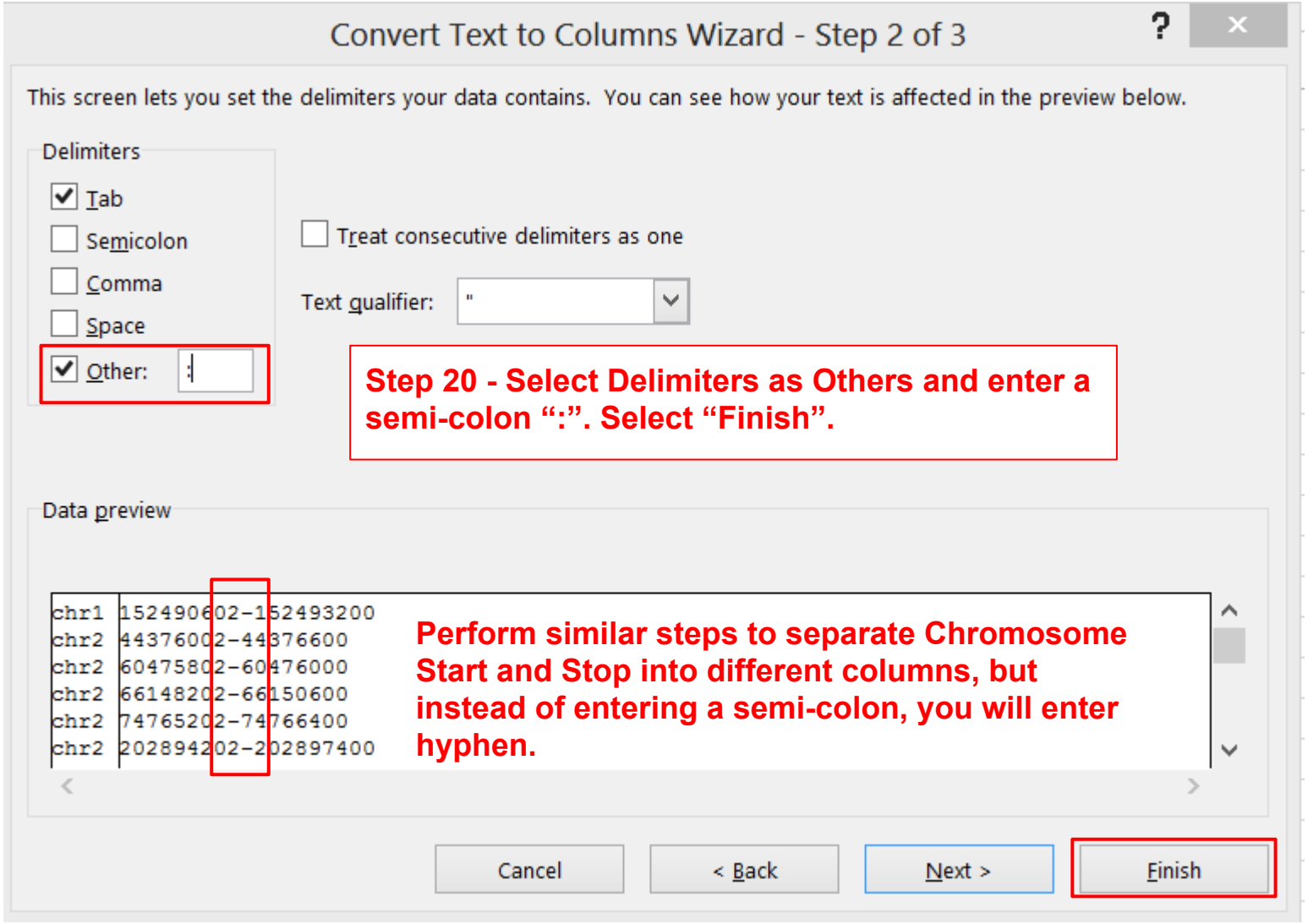


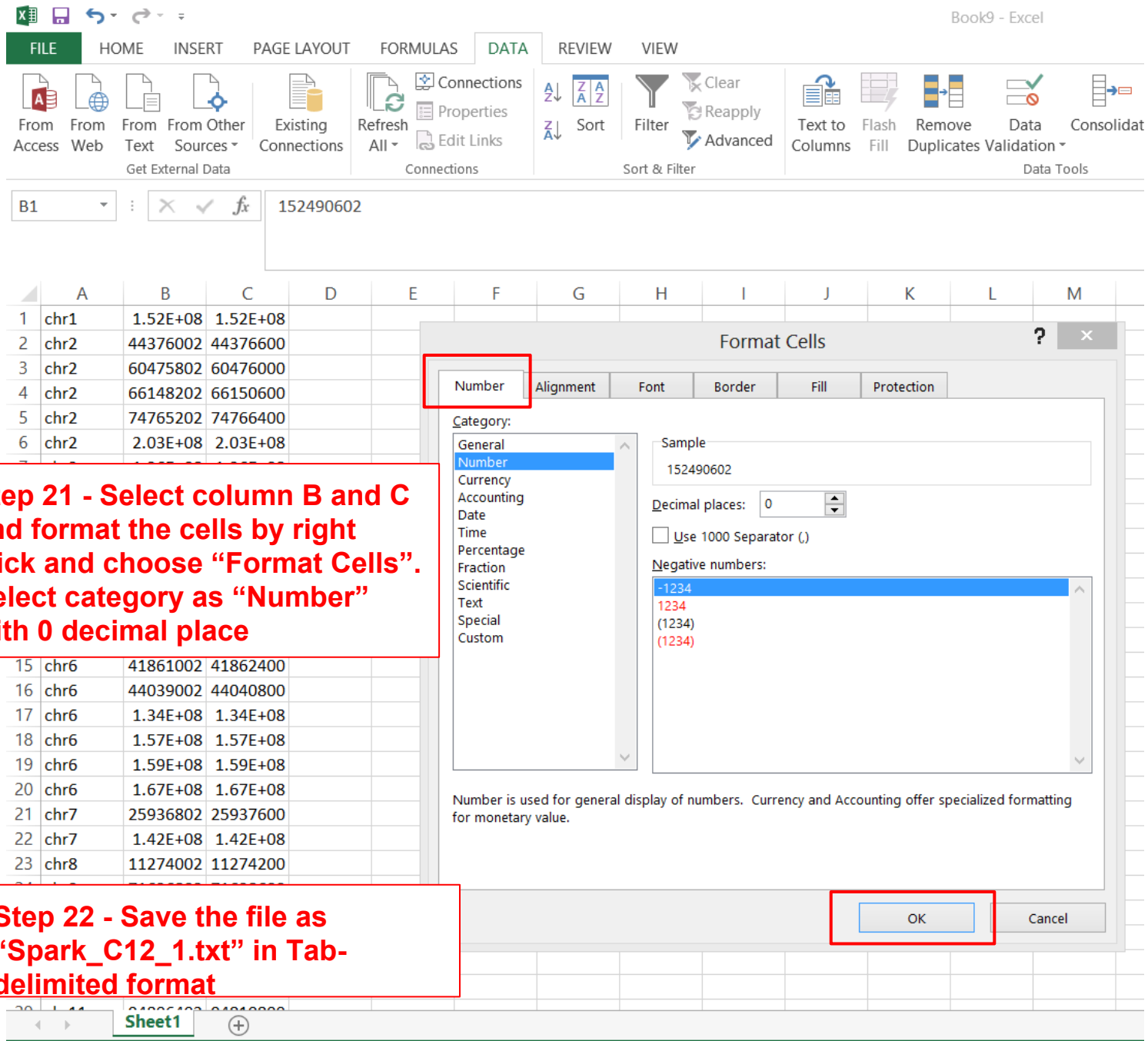
Step 17 - Select Copy and Launch. Web-browser will be launched. We will do Motif and GREAT analysis via Genboree. We will not use DAVID as regions are enhancers and not genes.



Once selected "Copy and Launch", open excel and paste the region IDs to make BED file


Book9 - Excel
FILE
HOME
INSERT
PAGE LAYOUT
FORMULAS
DATA
REVIEW
VIEW
From Access
From Web
From Text
From Other Sources
Existing Connections
Get External Data
Refresh All
Connections
Properties
Edit Links
Connections
Sort
Filter
Clear
Reapply
Advanced
Sort & Filter
Text to Columns
Flash Fill
Remove Duplicates
Data Validation
Consolidate
Data Tools
A1
chr1:152490602-152493200
Step 18 - Paste and Select the column.
Under "Data", select "Text to Columns
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.
Step 19 - Select Delimited and select "Next"
Preview of selected data:
1 chr1:152490602-152493200
2 chr2:44376002-44376600
3 chr2:60475802-60476000
4 chr2:66148202-66150600
5 chr2:74765202-74766400
6 chr2:202894202-202897400
Cancel
< Back
Next >
Finish
Sheet1
READY

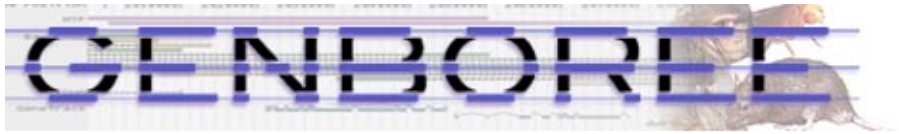

Convert Text to Columns Wizard - Step 2 of 3
This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.
Delimiters
Tab
Semicolon
Comma
Space
Other: :
Treat consecutive delimiters as one
Text qualifier: "
Step 20 - Select Delimiters as Others and enter a semi-colon ":". Select "Finish".
Data preview
chr1 152490602-152493200
chr2 44376002-44376600
chr2 60475802-60476000
chr2 66148202-66150600
chr2 74765202-74766400
chr2 202894202-202897400
Perform similar steps to separate Chromosome Start and Stop into different columns, but instead of entering a semi-colon, you will enter hyphen.
Cancel
< Back
Next >
Finish



	A	B	C
1	chr1	1.52E+08	1.52E+08
2	chr2	44376002	44376600
3	chr2	60475802	60476000
4	chr2	66148202	66150600
5	chr2	74765202	74766400
6	chr2	2.03E+08	2.03E+08
15	chr6	41861002	41862400
16	chr6	44039002	44040800
17	chr6	1.34E+08	1.34E+08
18	chr6	1.57E+08	1.57E+08
19	chr6	1.59E+08	1.59E+08
20	chr6	1.67E+08	1.67E+08
21	chr7	25936802	25937600
22	chr7	1.42E+08	1.42E+08
23	chr8	11274002	11274200

Step 21 - Select column B and C and format the cells by right click and choose “Format Cells”. Select category as “Number” with 0 decimal place

Step 22 - Save the file as “Spark_C12_1.txt” in Tab-delimited format



BCM
Baylor College of Medicine

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

Welcome to the Genboree Workbench! [Getting Started]

Step 24 -Expand "Data" > "Tracks" > "Import" > "Upload Track Annotations"

Data Selector

Refresh

- genboree.org
  - Atlas Tools Access
  - BRL AUTO TEST
  - 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
  - vamin_group
- genboree.bcgsc.ca
- genboree.cbrc.jp
- www.brain-research-lab.org

ails

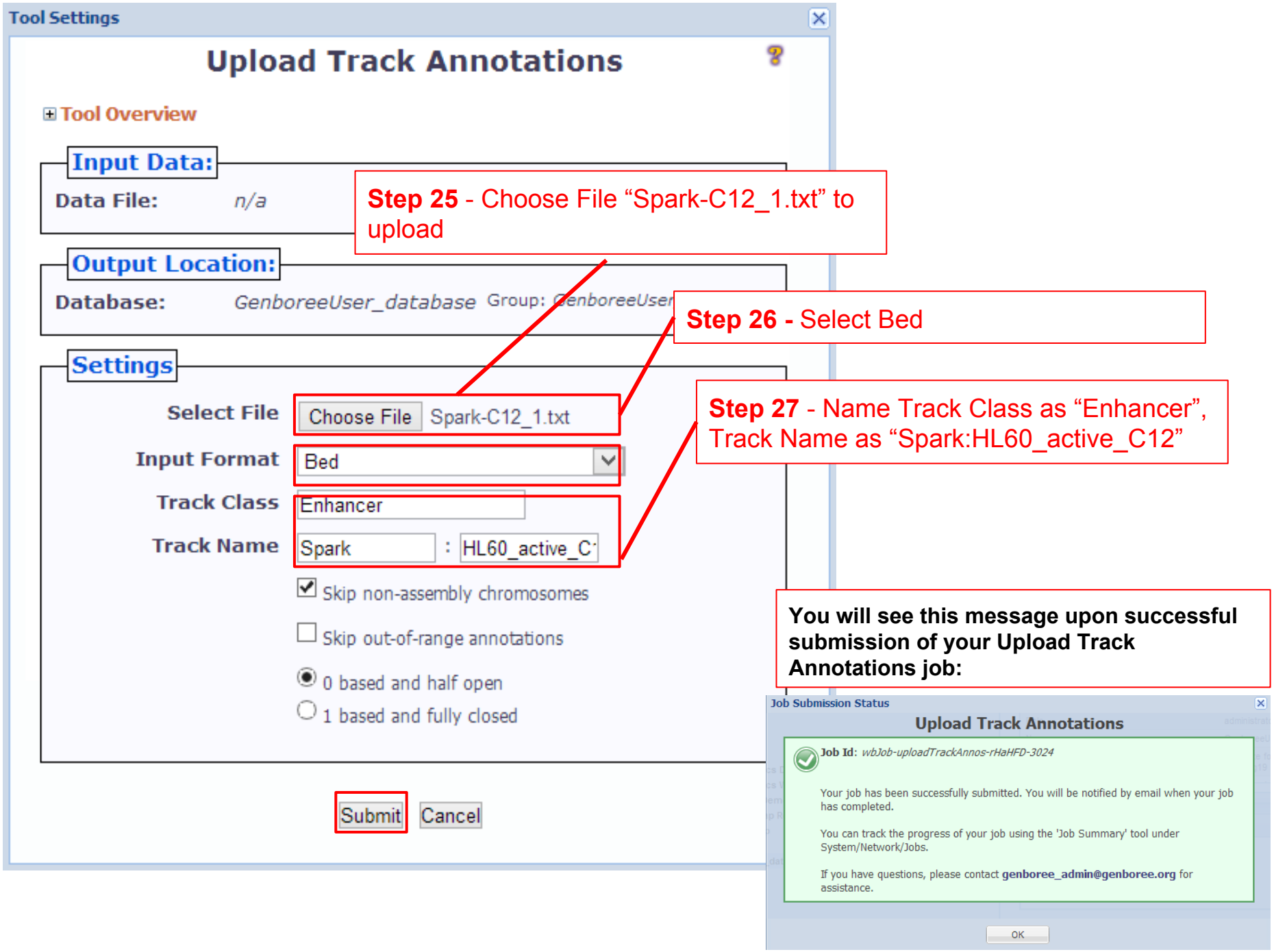
ibute	Value
Group	GenboreeUser_group
Role	administrator
Name	GenboreeUser_database
Description	Template for Human Genome, UCSC Build Hg19

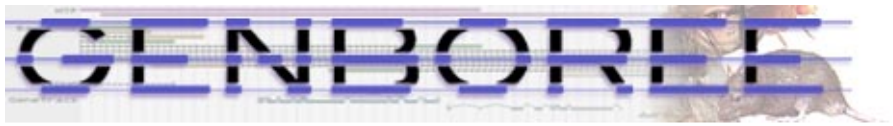
Input Data

Output Targets

GenboreeUser_database

Step 23 -To upload track annotations, drag your database


Tool Settings
Upload Track Annotations
Tool Overview
Input Data:
Data File: n/a
Output Location:
Database: GenboreeUser_database Group: GenboreeUser
Settings
Select File
Choose File
Spark-C12_1.txt
Input Format
Bed
Track Class
Enhancer
Track Name
Spark
:
HL60_active_C
Skip non-assembly chromosomes
Skip out-of-range annotations
0 based and half open
1 based and fully closed
Submit
Cancel
Step 25 - Choose File "Spark-C12_1.txt" to upload
Step 26 - Select Bed
Step 27 - Name Track Class as "Enhancer", Track Name as "Spark:HL60_active_C12"
You will see this message upon successful submission of your Upload Track Annotations job:
Job Submission Status
Upload Track Annotations
Job Id: wbJob-uploadTrackAnnos-rHaHFD-3024
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.
OK



Baylor College of Medicine

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

Welcome to the Genboree Workbench! [Getting Started]

Data Selector

Refresh

Select “Refresh” to view the uploaded track

- GenboreeUser_group
  - Databases
    - GenboreeUser_database
      - All Annotations in Database
      - Tracks
        - Class: BranchSpecificROI
        - Class: Class
        - Class: Enhancer
          - Spark:HL60_active_C12
        - Class: Gene
        - Class: High Density Score Data
        - Class: MACS
        - Class: Marker
        - Class: Sequence
        - Class: SPARK
      - Lists & Selections
      - SampleSets
      - Samples
      - Files
      - Queries
  - Projects
    - GenboreeUser_project

Details

Attribute	Value
View Link	Link to Project
Group	GenboreeUser_group
Name	Use_Case_20_GU

Input Data

Spark:HL60_active_C12

Output Targets

GenboreeUser_database
Use_Case_20_GU

Step 28 - Drag Spark: HL60_active_C12 from your database in Input Data. Drag Your Database and Project Page in Output Targets

BCM
Baylor College of Medicine

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

Analyze Structural Variants
Find Motifs
SNPs
GREAT
HOMER

Step 29 - Expand "Genome" > "Find Motifs" > "HOMER"

Welcome to the Genboree Workbe

Data Selector

Refresh | Data Filter:

- GenboreeUser_group
  - Databases
    - GenboreeUser_database
      - All Annotations in Database
      - Tracks
        - Class: BranchSpecificROI
        - Class: Class
        - Class: Enhancer
          - Spark:HL60_active_C12
        - Class: Gene
        - Class: High Density Score Data
        - Class: MACS
        - Class: Marker
        - Class: Sequence
        - Class: SPARK
      - Lists & Selections
      - SampleSets
      - Samples
      - Files
      - Queries
  - Projects
    - GenboreeUser_project

	Value
View Link	Link to Project
Group	GenboreeUser_group
Name	Use_Case_20_GU

Input Data

Spark:HL60_active_C12

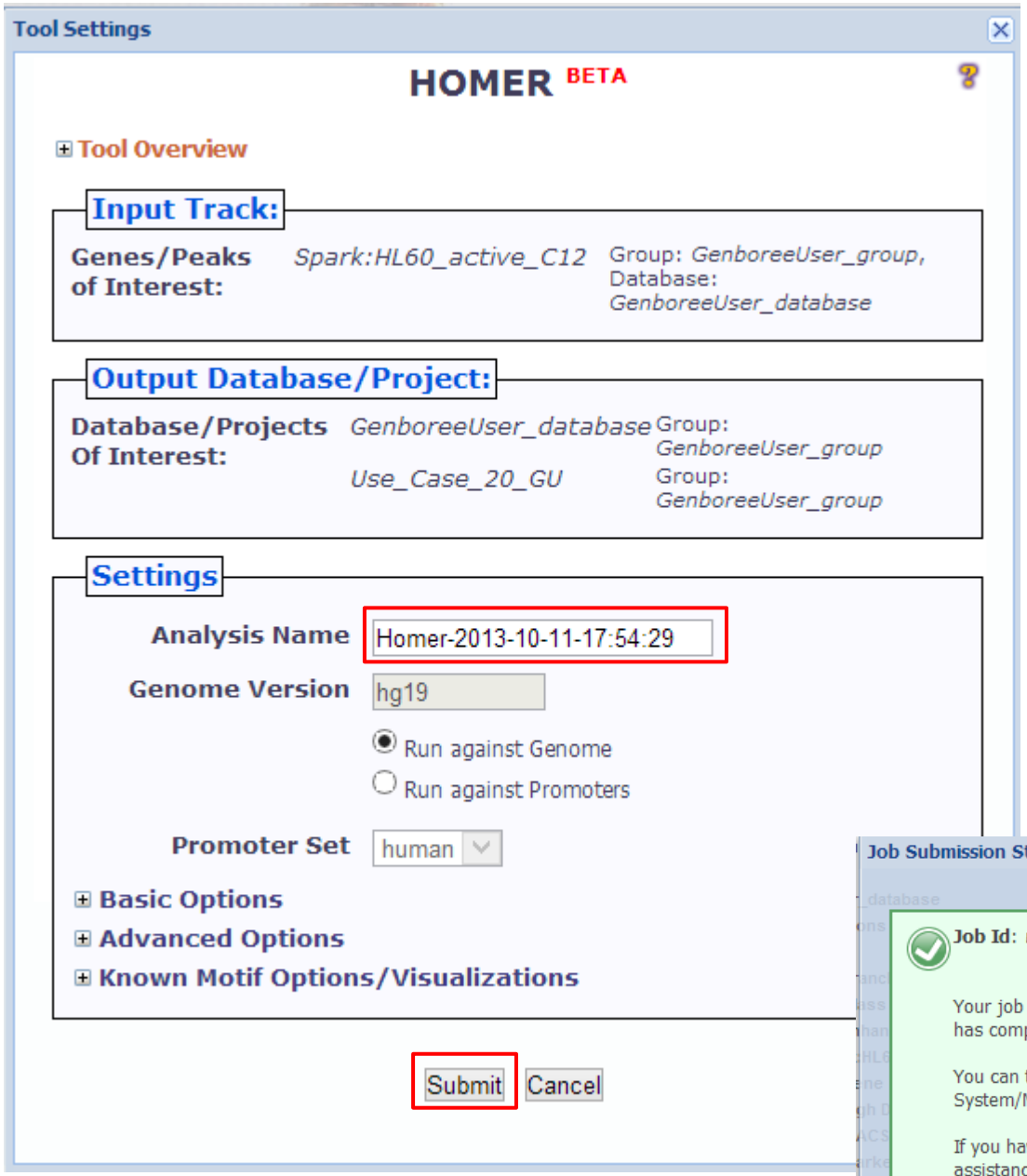
Output Targets

GenboreeUser_database
Use_Case_20_GU

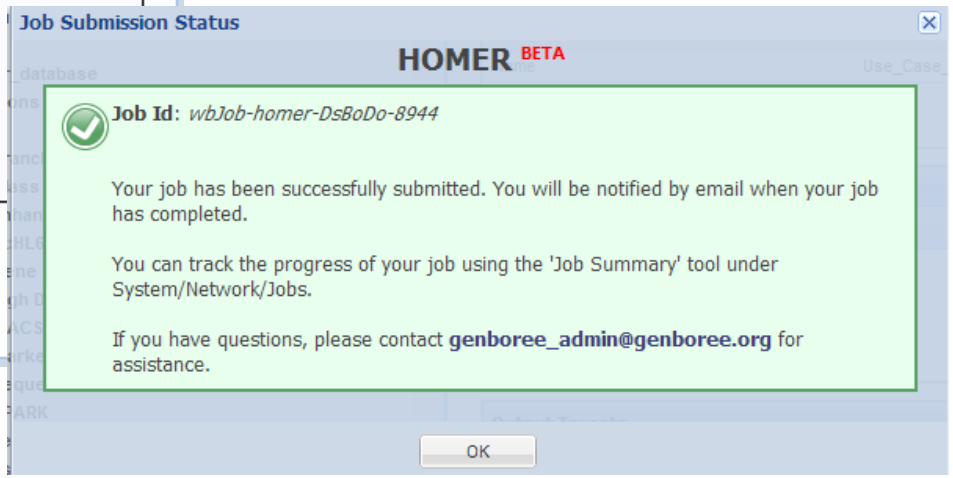
Genboree is built & maintained by the **Bioinformatics Research Laboratory** at **Baylor College of Medicine.**

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

HGSC
HUMAN GENOME SEQUENCING CENTER



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



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

Hello

Your HOMER job completed successfully.

Job Summary:
JobID - wbJob-homer-hsthEq-0654

Additional Info:
Target Group:
Target Database:

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

from HOMER un

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

Project Page

Project News:

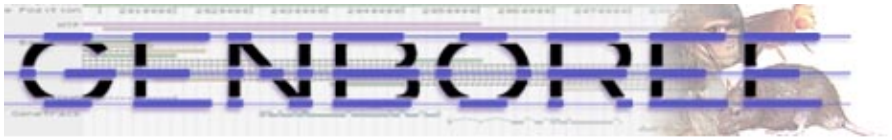
2013/10/11: Genboree User ran a GREAT (Genomic Regions Enrichment of Annotations Tool) job (wbJob-great-DqocDC-0173). Click the link below to perform live analysis with GREAT:

- **Peform GREAT anlaysis**

2013/10/11: Genboree User ran a HOMER job (wbJob-homer-DsBoDo-8944) and the results are available at the link below.

- **Link to Homer Results**

Links to HOMER results of motifs that are enriched for the regions chosen in Spark cluster.



BCM
Baylor College of Medicine

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

Analyze Structural Variants
Find Motifs
SNPs
GREAT

Welcome to the Genboree Workbe

Step 31 - Expand "Genome" > select "GREAT"

Data Selector

Refresh Data Filter:

- GenboreeUser_group
  - Databases
    - GenboreeUser_database
      - All Annotations in Database
      - Tracks
        - Class: BranchSpecificROI
        - Class: Class
        - Class: Enhancer
          - Spark:HL60_active_C12
        - Class: Gene
        - Class: High Density Score Data
        - Class: MACS
        - Class: Marker
        - Class: Sequence
        - Class: SPARK
      - Lists & Selections
      - SampleSets
      - Samples
      - Files
      - Queries
  - Projects
    - GenboreeUser_project

View Link

Group	GenboreeUser_group
Name	Use_Case_20_GU

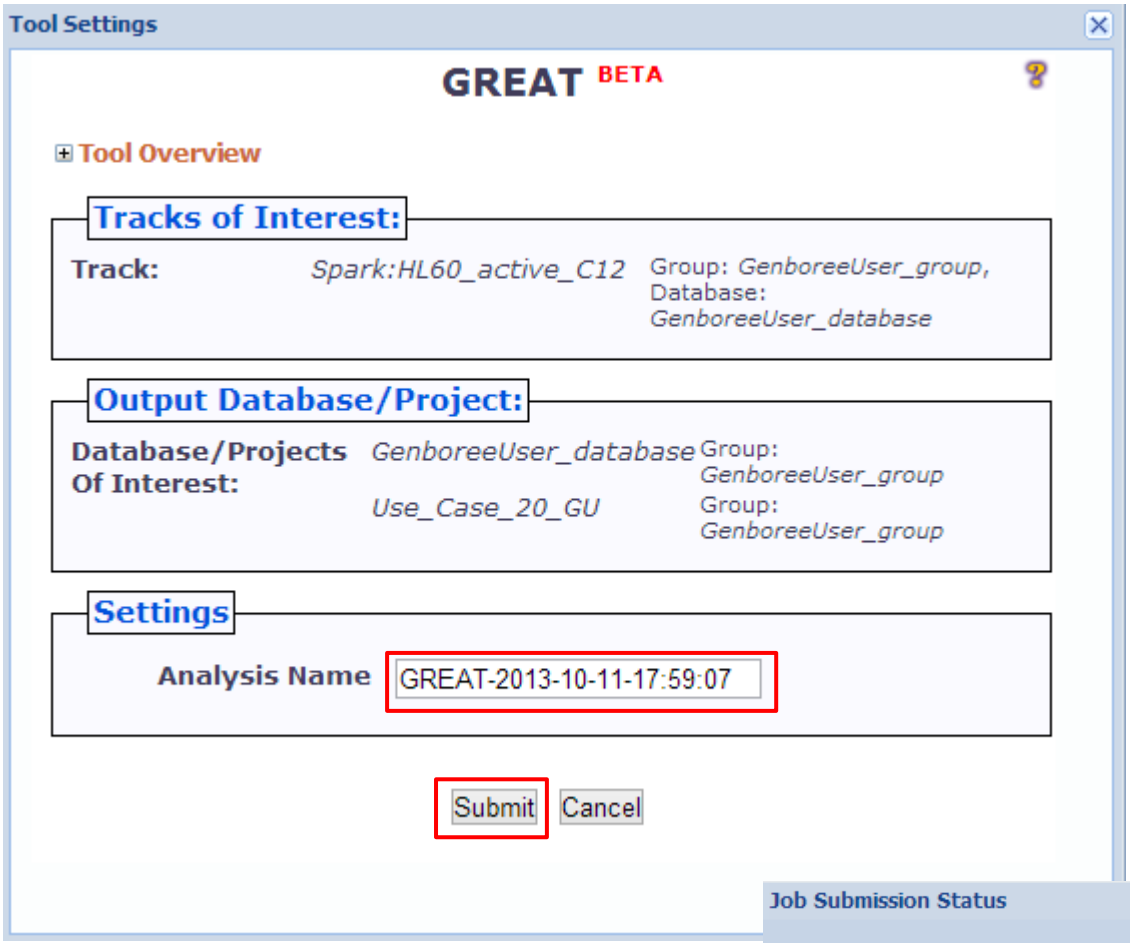
Input Data

Spark:HL60_active_C12

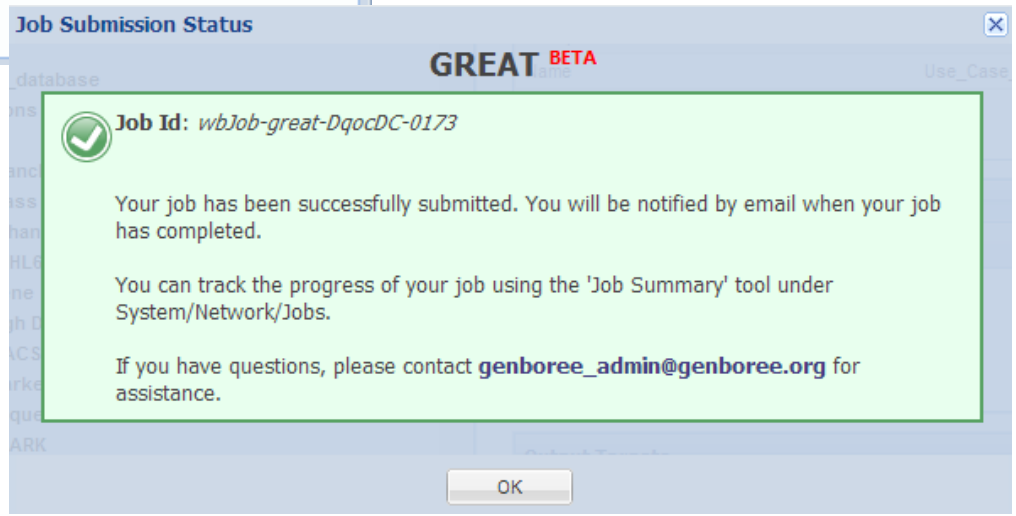
Step 30 - Drag Spark: HL60_active_C12 from your database in Input Data. Drag Your Database and Project Page in Output Targets

Output Targets

GenboreeUser_database
Use_Case_20_GU



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



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:

2013/10/11: Genboree User ran a GREAT (Genomic Regions Enrichment of Annotations Tool) job (wbJob-great-DqocDC-0173). Click the link below to perform live analysis with GREAT:

- **Peform GREAT anlaysis**

2013/10/10: Genboree User ran Epigenomic Heatmap Tool (Heatmap HL60 Immune 2013-10-10-18 20 45) and the results are available at the link below.

- **Study Name:** Heatmap HL60 Immune 2013-10-10-18 20 45
- **Link to results**

GREAT - Genomic Regions Enrichment of Annotations Tool

Host	Group	Database	Track	Link for live analysis
genboree.org	GenboreeUser_group	GenboreeUser_database	Spark:HL60_active_C12	Click here

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

Job Description

Job ID: 20131011-public-2.0.2-BkWIlW

Display name: external data

Test set: external data (1,025 genomic regions)
Show in UCSC genome browser. *How do I look at my regions in the genome?*

Background: Whole genome background

Assembly: Human: GRCh37 (UCSC hg19, Feb 2009) *What gene set does GREAT use?*

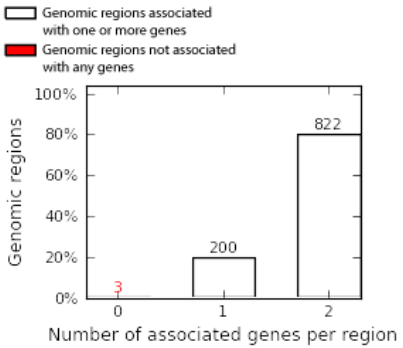
Associated genomic regions: Basal+extension (constitutive 5.0 kb upstream and 1.0 kb downstream, up to 1000.0 kb max extension). Curated regulatory domains are included.
3 of all 1,025 genomic regions (0.3%) are not associated with any genes.
View all genomic region-gene associations. *Which genes are my regions associated with?*
Revise the region-gene association rule. *How are my regions associated with genes?*

Region-Gene Association Graphs

What do these graphs illustrate?

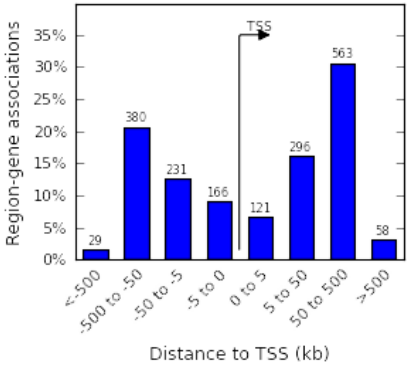
Number of associated genes per region

Download as PDF.



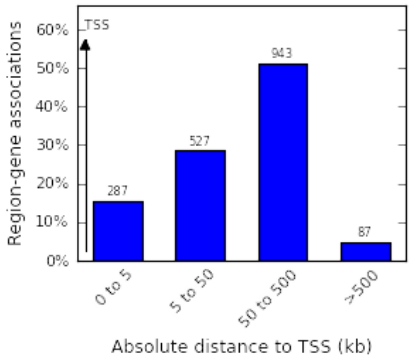
Binned by orientation and distance to TSS

Download as PDF.



Binned by absolute distance to TSS

Download as PDF.



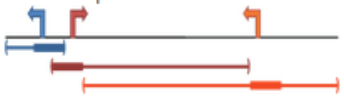
When should I use a background set?
What should my background regions file contain?

Association rule settings Hide settings »

Associating genomic regions with genes

GREAT calculates statistics by associating genomic regions with nearby genes and applying the gene annotations to the regions. Association is a two step process. First, every gene is assigned a regulatory domain. Then, each genomic region is associated with all genes whose regulatory domain it overlaps.

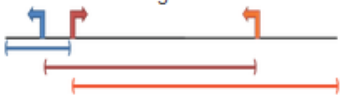
◉ Basal plus extension



Proximal: 5.0 kb upstream, 1.0 kb downstream, plus Distal up to 500.0 kb

Gene regulatory domain definition: Each gene is assigned a basal regulatory domain of a minimum distance upstream and downstream of the TSS (regardless of other nearby genes). The gene regulatory domain is extended in both directions to the nearest gene's basal domain but no more than the maximum extension in one direction.

○ Two nearest genes



within 1000.0 kb

Gene regulatory domain definition: Each gene is assigned a regulatory domain that extends in both directions to the nearest gene's TSS but no more than the maximum extension in one direction.

○ Single nearest gene

within 1000.0 kb

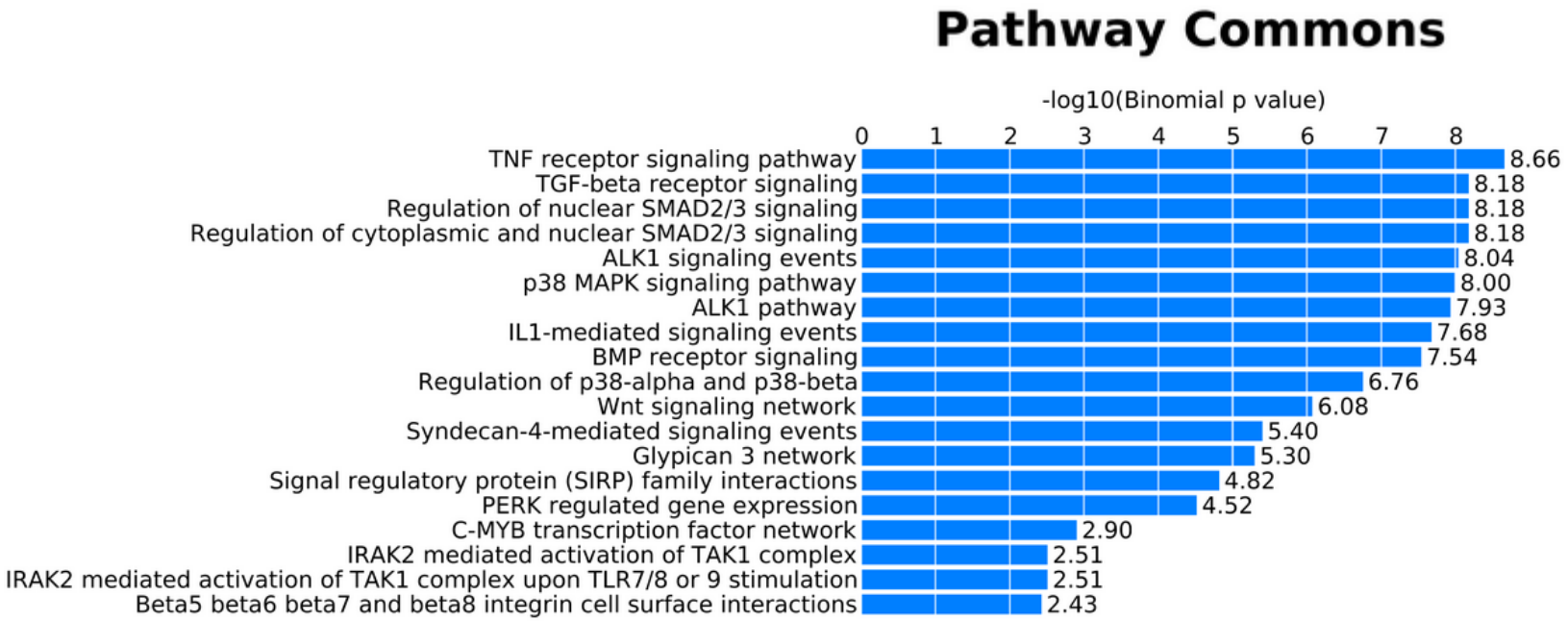
Gene regulatory domain definition: Each gene is assigned a regulatory domain that extends in both directions to the midpoint between the gene's TSS and the nearest gene's TSS but no more than the maximum extension in one direction.

Gene Transcription Start Site (TSS)

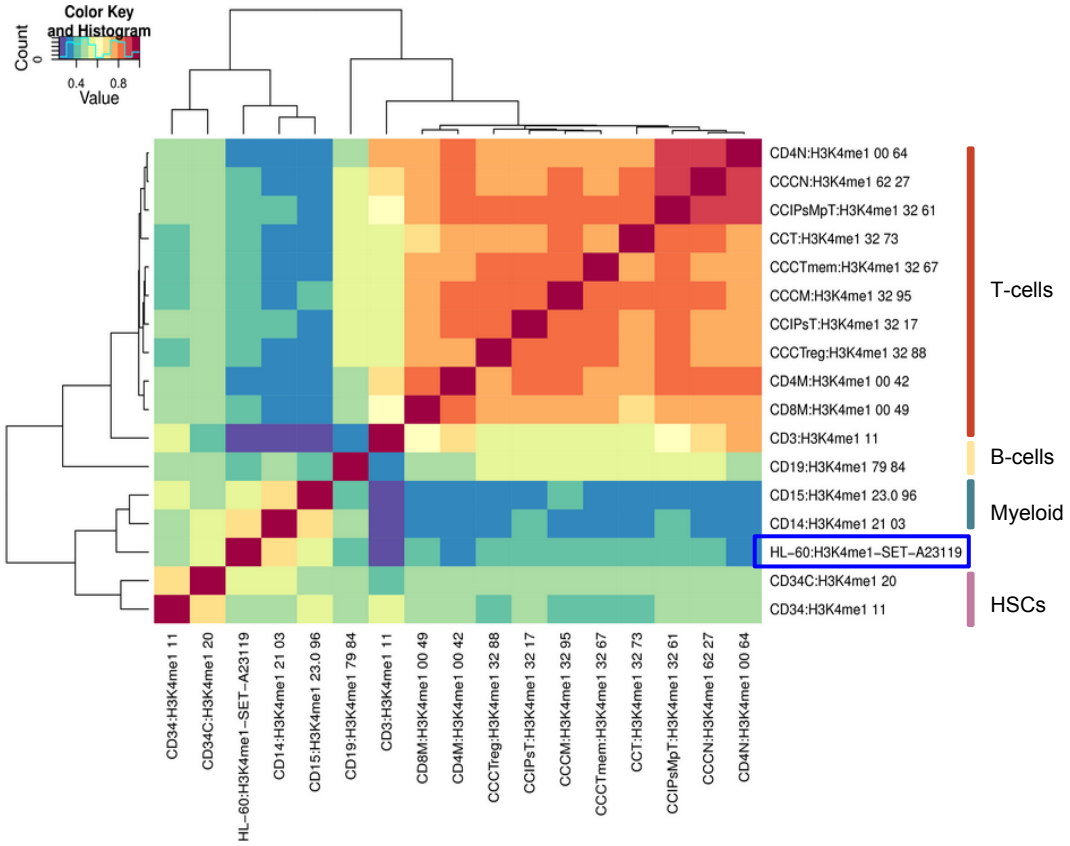
☑ Include curated regulatory domains *What are curated regulatory domains?*

Submit Reset

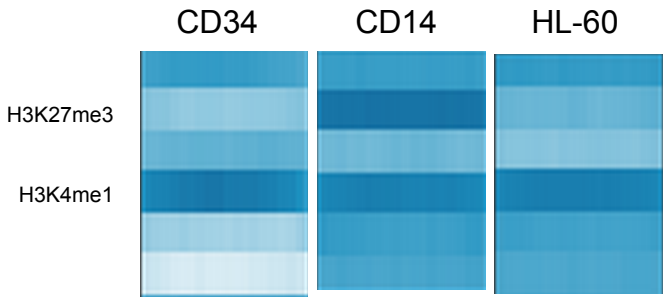
Pathways commons that are enriched using GREAT tool



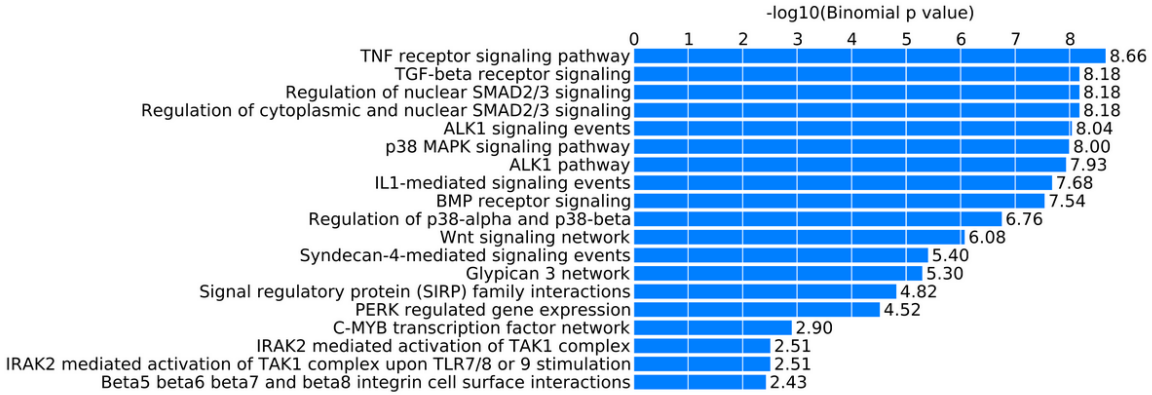
Summary and Interpretation of Results



Results: Comparison of reference epigenomes with HL60 histone modifications correctly identify HL60 as being of myeloid origin, suggesting one may perform an initial characterization of cell lines/tissues of unknown origin by virtual comparison to reference epigenomes. Further identifying enhancer regions that are undergoing epigenomic changes, Spark tool determined 1025 enhancers that underwent epigenomic transition. GREAT tool was used to determine pathways that these cis-regularoty regions are associated with. C-MYB, previously known, transcriptional regulator was found to be significantly enriched.



Epigenomic transitions of 1025 enhancers



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