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

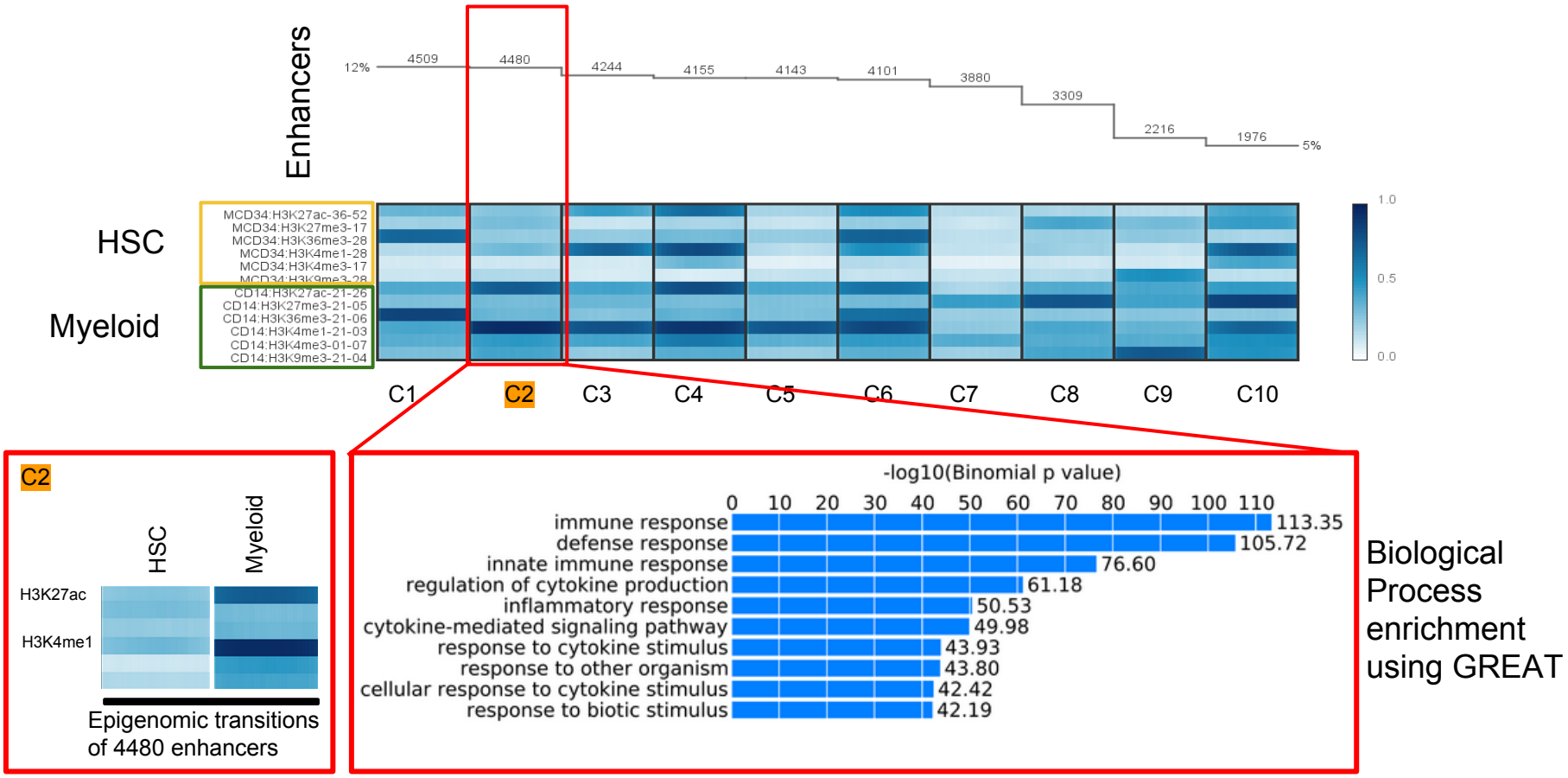
Baylor College of Medicine

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.

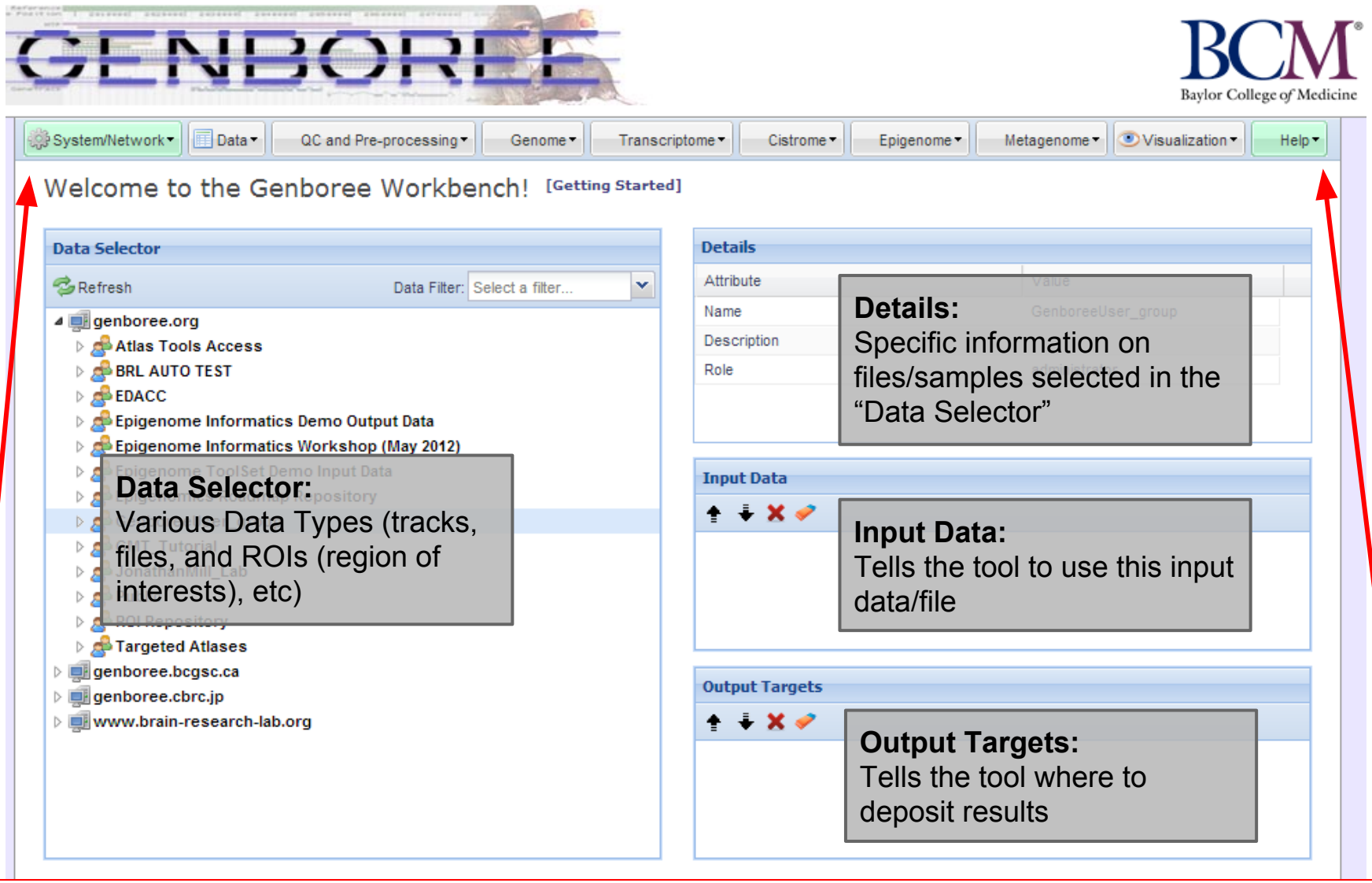
Use Case Overview

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

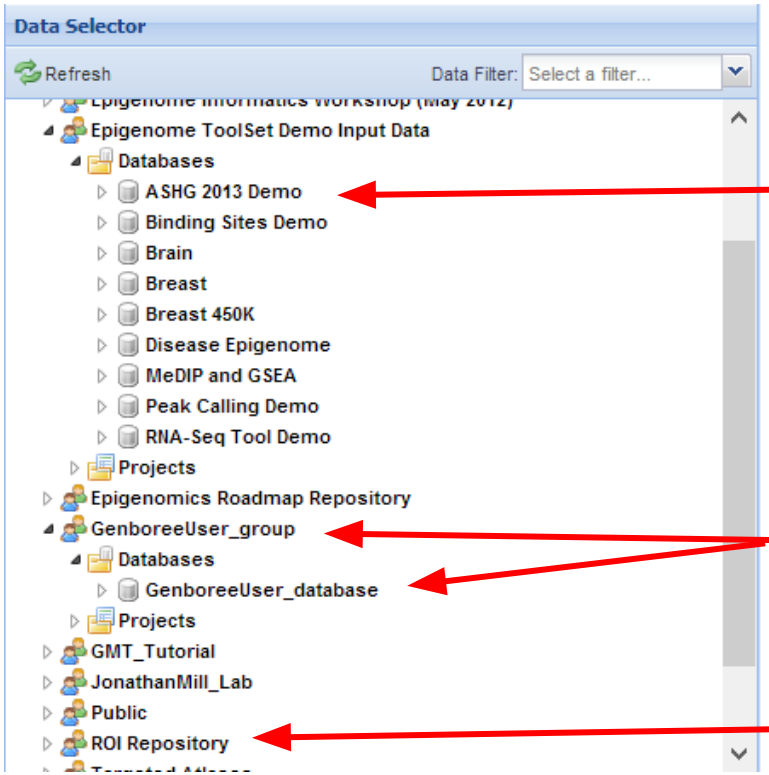
The Genboree Workbench: Web-based Data Management & Analysis



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.

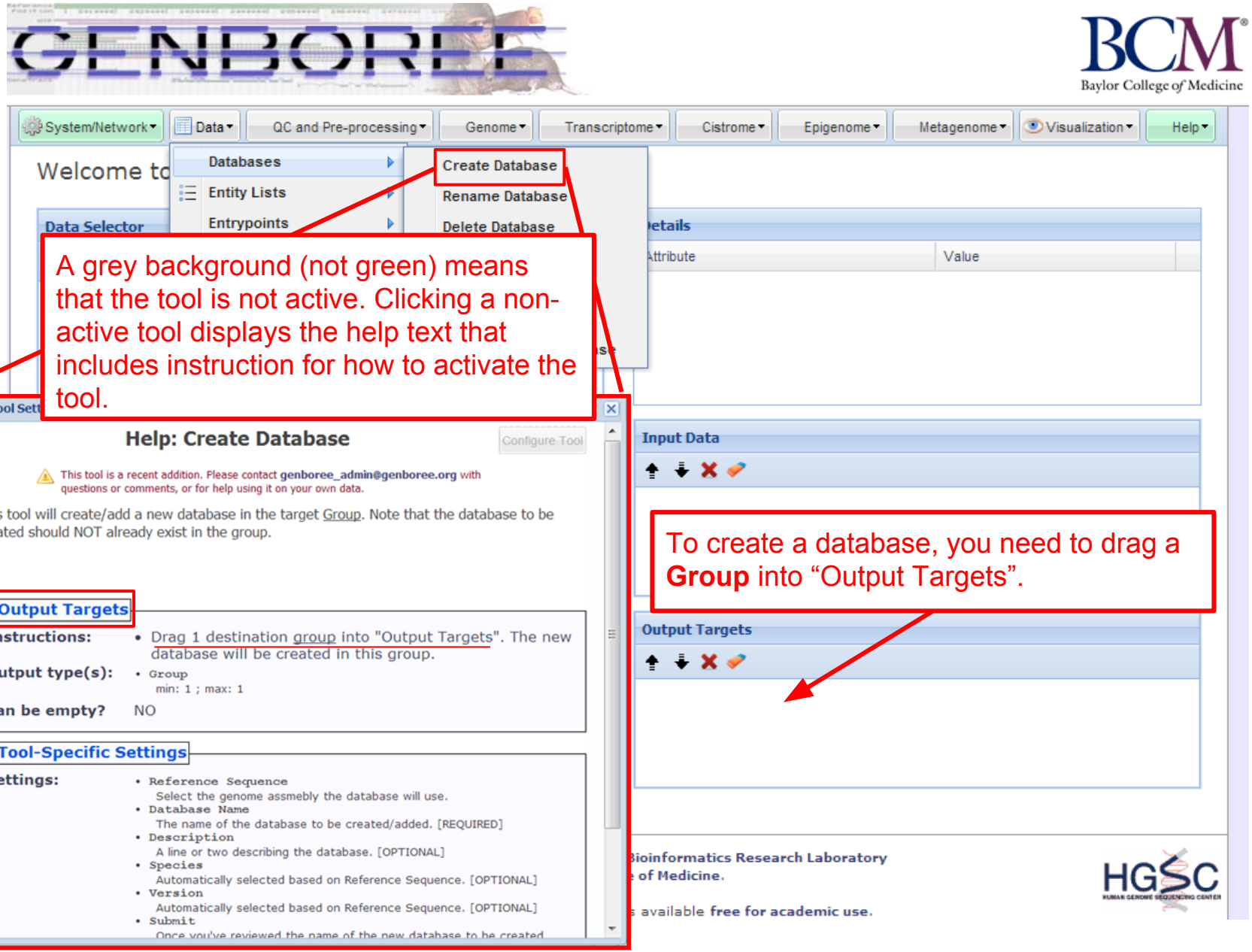


Under "Epigenome Toolset Demo Input Data" you will find "ASHG 2013 Demo" database, where we have provided you with sample data to try out the use cases

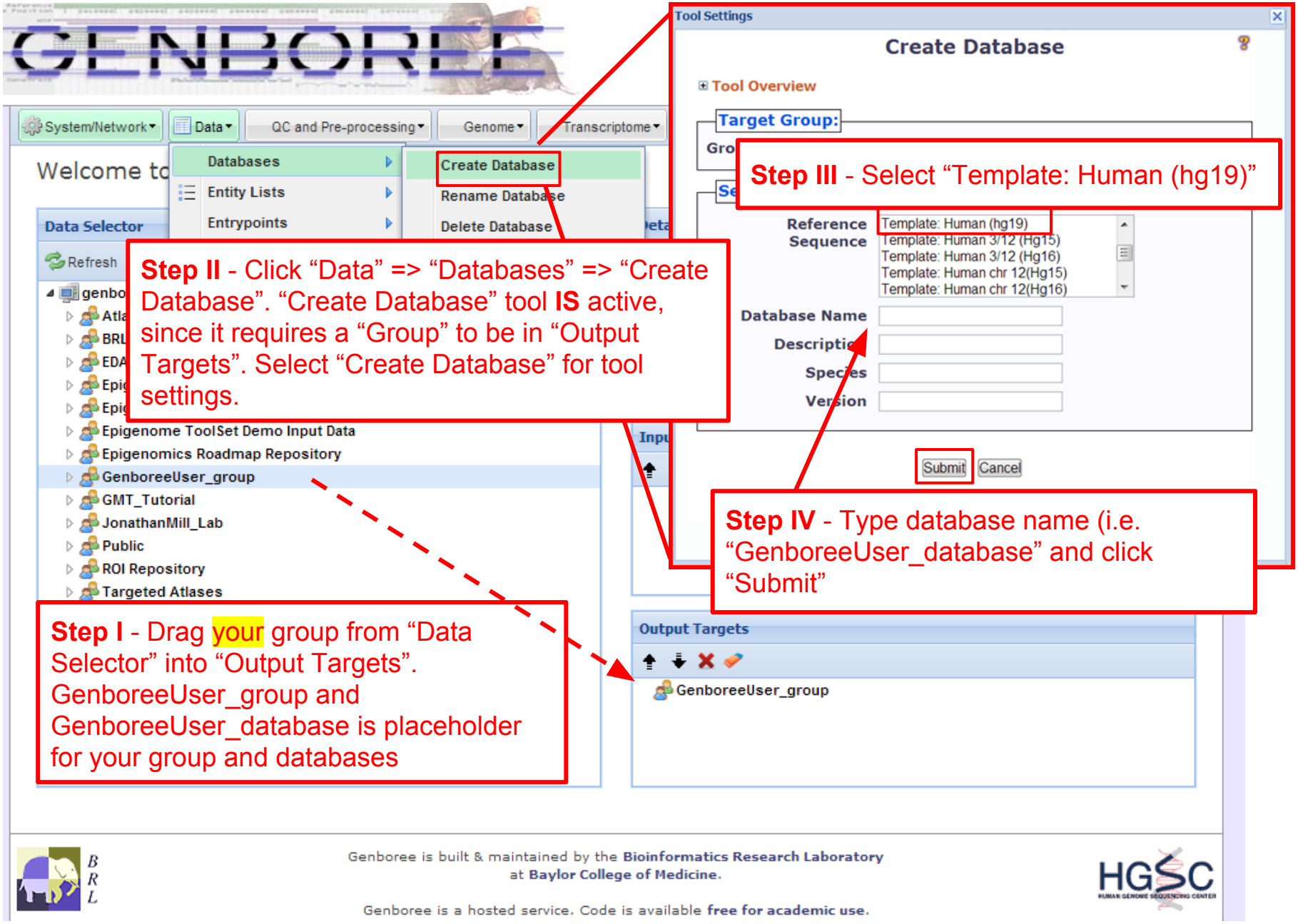
When making screenshots for providing instructions we have used "GenboreeUser_group" and "GenboreeUser_database" as output targets, however, you will use your own group and database. Following slides will show you how to create database and project

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

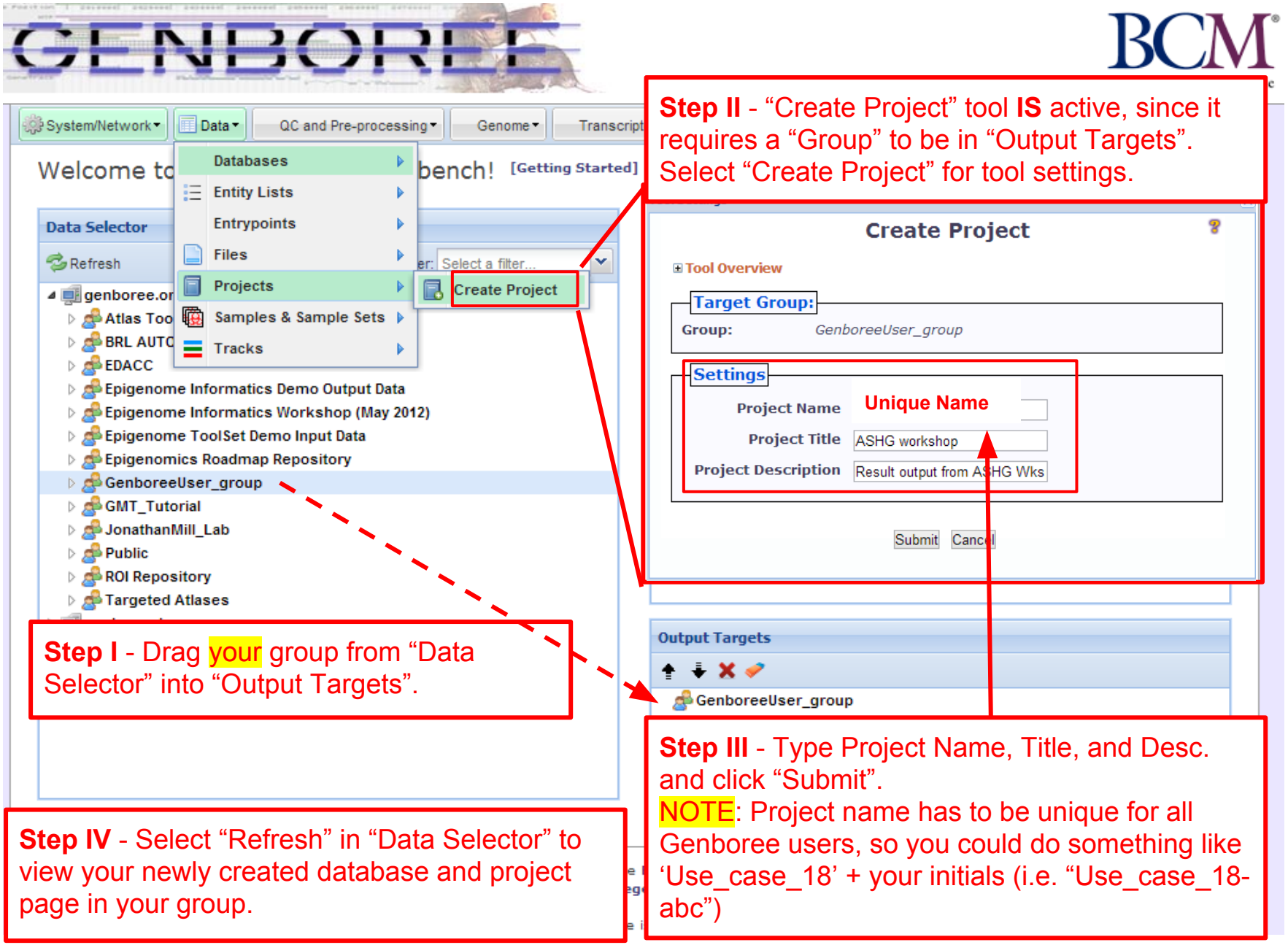
Display Tool Setting “Help” dialogue box in the Workbench



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:

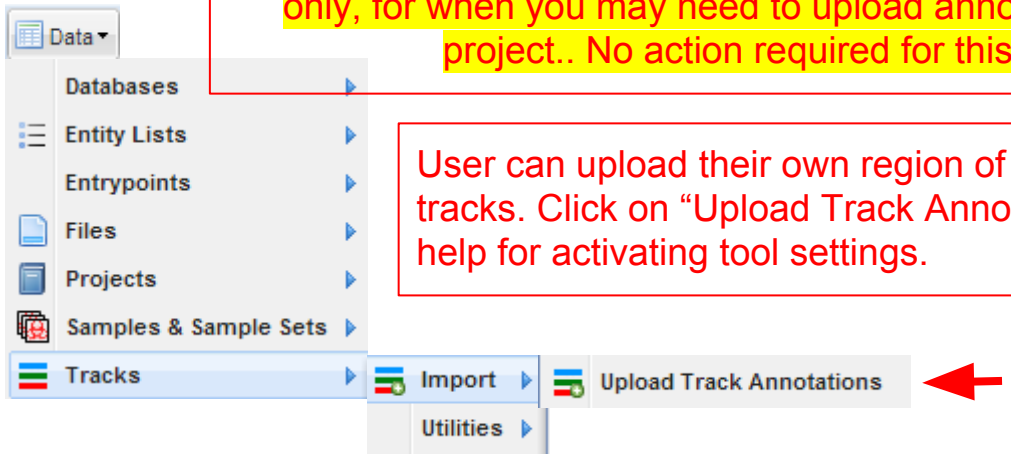
- **Track "GeneProteinCoding_promoter":** This track contains promoters of protein coding genes that were defined using Gencode V10 annotations (www.gencodegenes.org/), with transcription start sites (TSS) +/- 1500bp. The track contains 20,007 promoters from human genome build Hg19.

- **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 "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.

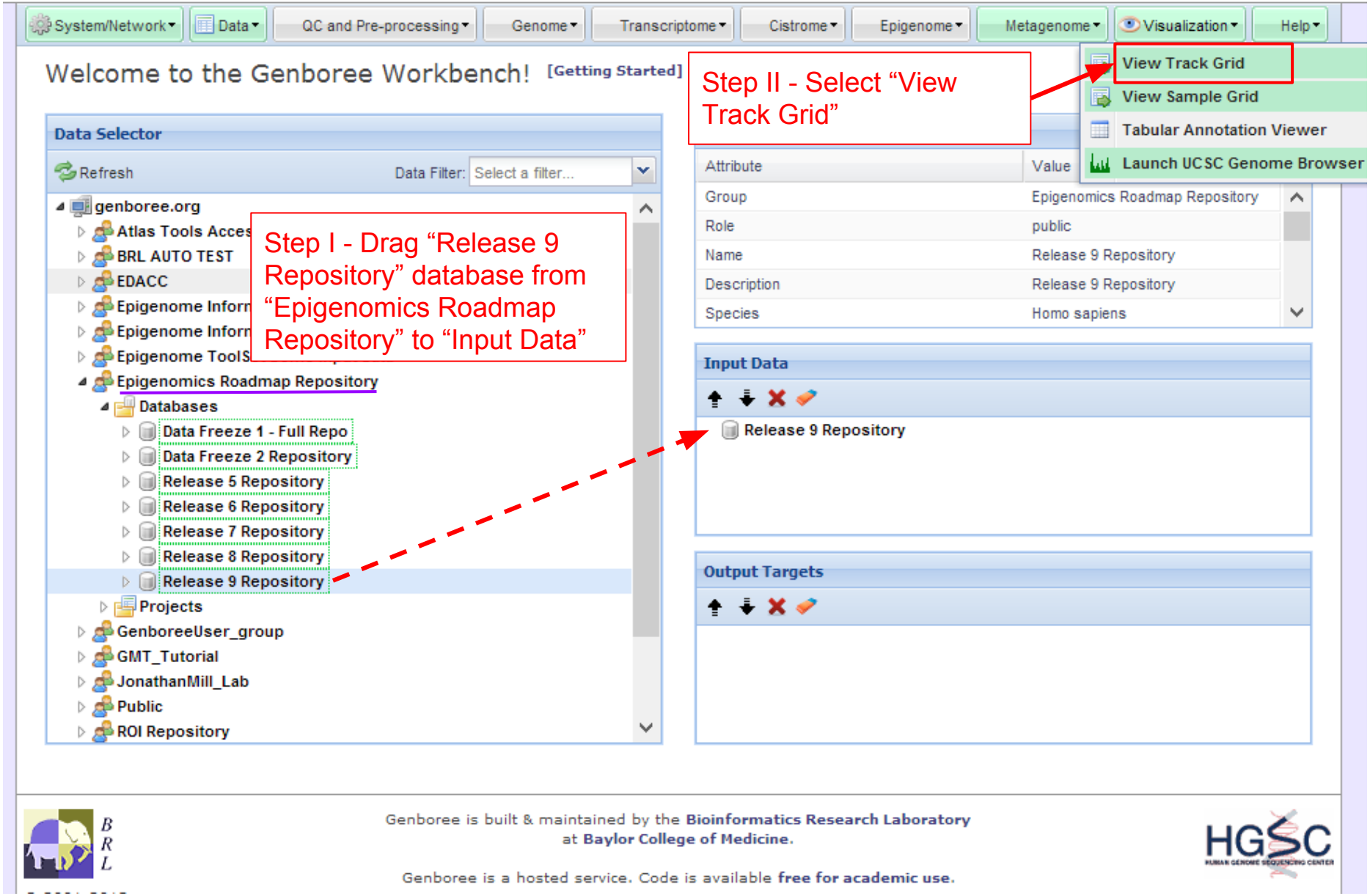
Upload your own Track Annotations:



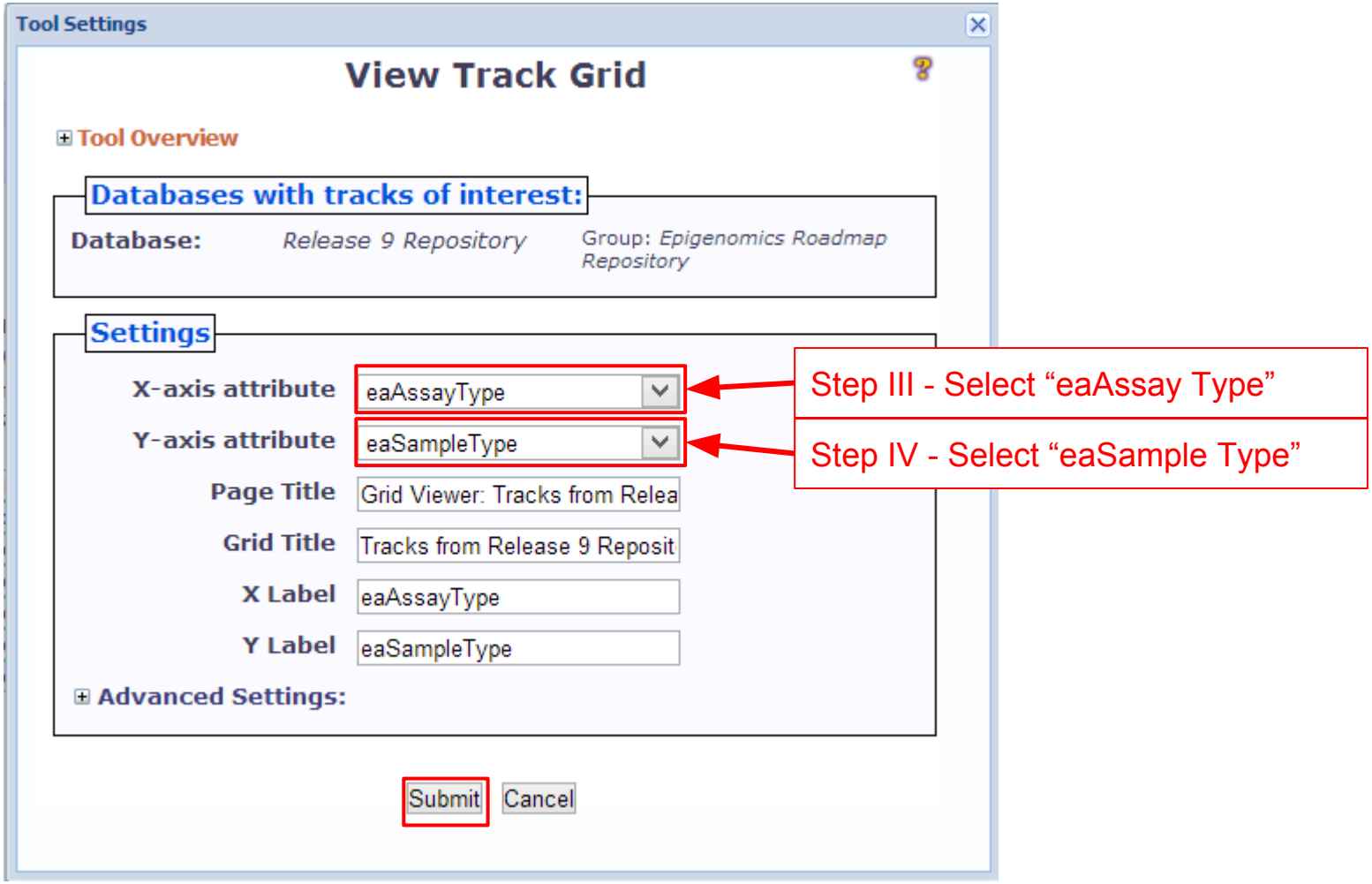
User can upload their own region of interests tracks. Click on "Upload Track Annotations" to see help for activating tool settings.

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

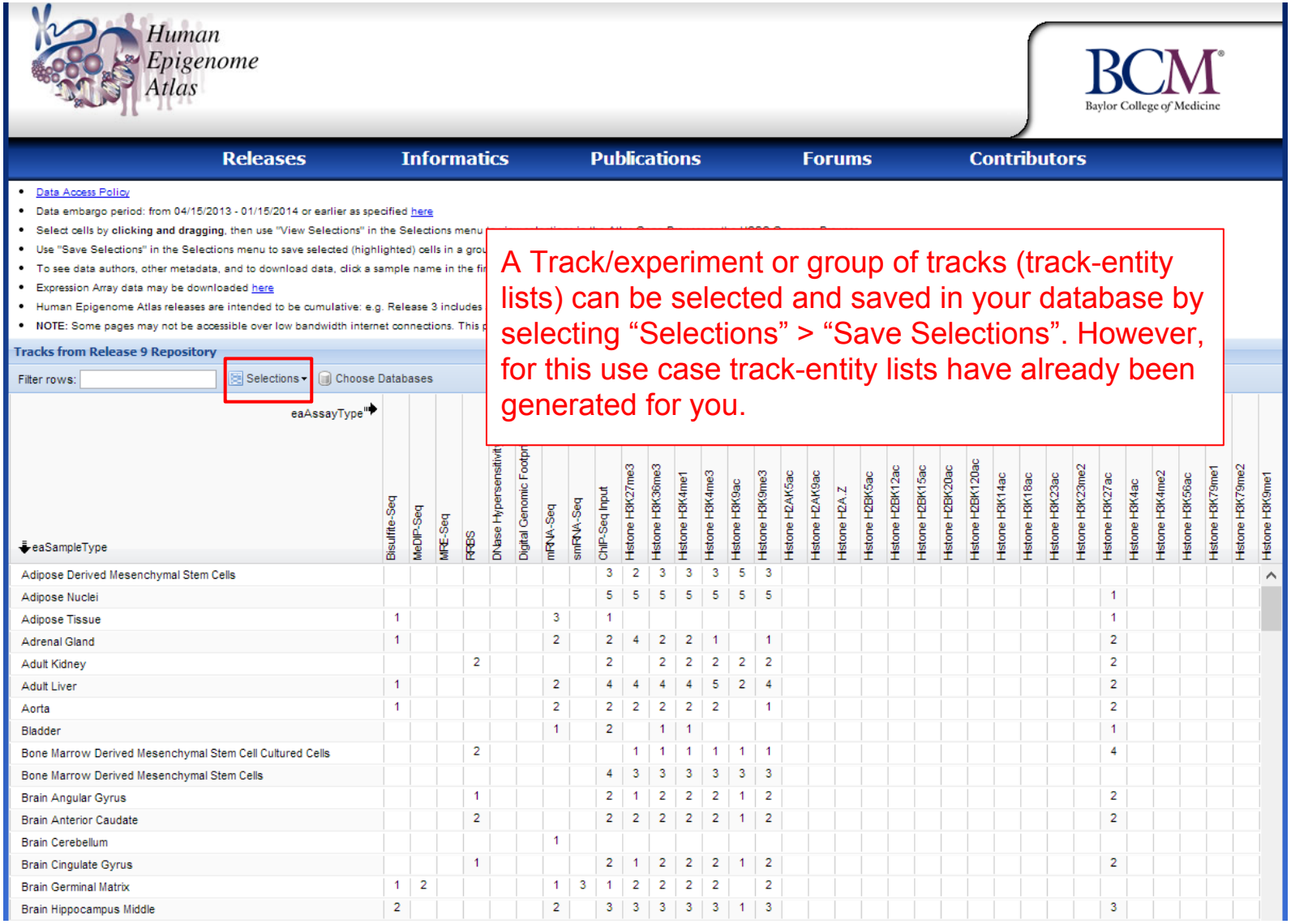
Follow these steps to view Track Grid of data from the Roadmap Epigenome Project



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						

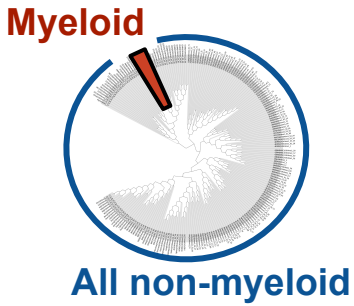
Methodology Overview

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



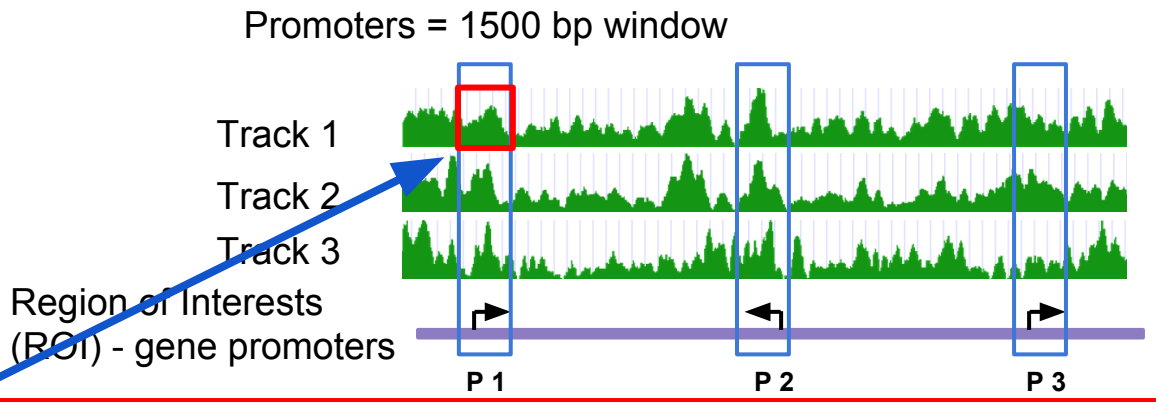
Spark: visualize epigenomic changes happening at LIMMA defined enhancer regions from hematopoietic stem cells to myeloid cell type by clustering regions with similar “epigenomic footprint” transitions

GREAT: assesses functional significance of cis-regulatory regions.

HOMER: de novo motif discovery.

Methodology: Clustering/Heatmap

I. Data Selection



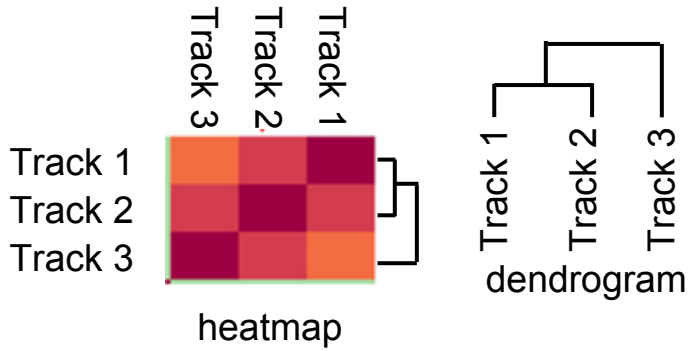
II. Signal processing- Heatmap Tool calculates average signal for each track and ROI (promoter) combination

	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

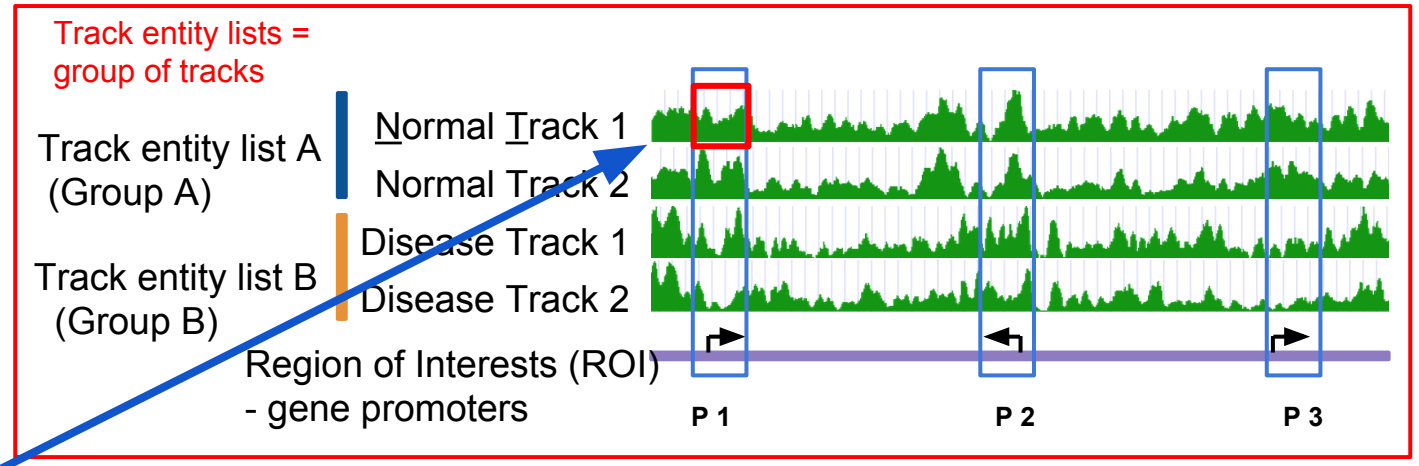
Correlation values are dynamically scaled and represented in heatmap.



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.

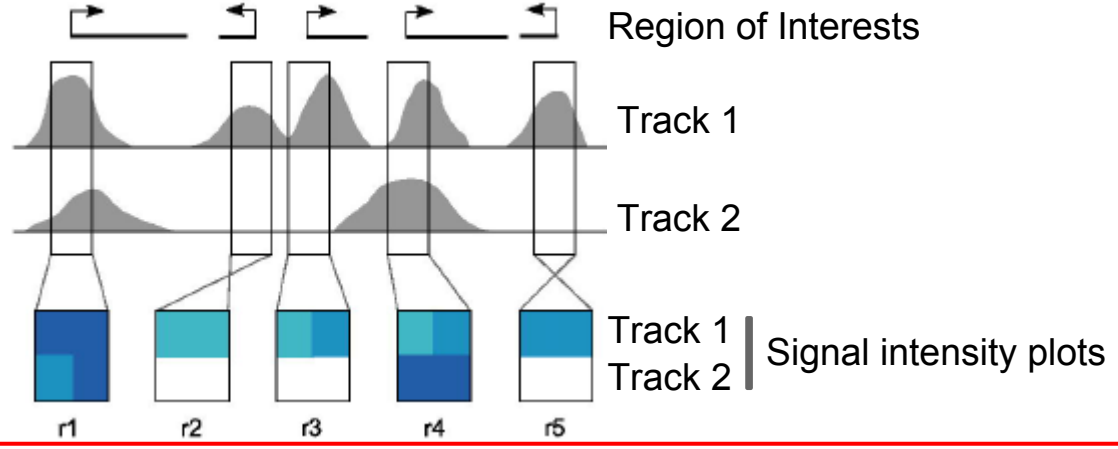
III. Results output

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

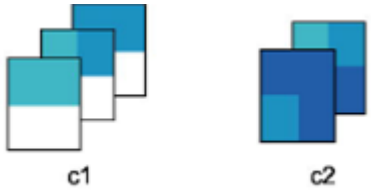
Methodology: Spark

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

I. Preprocessing
(ROIs = r1 - r5)

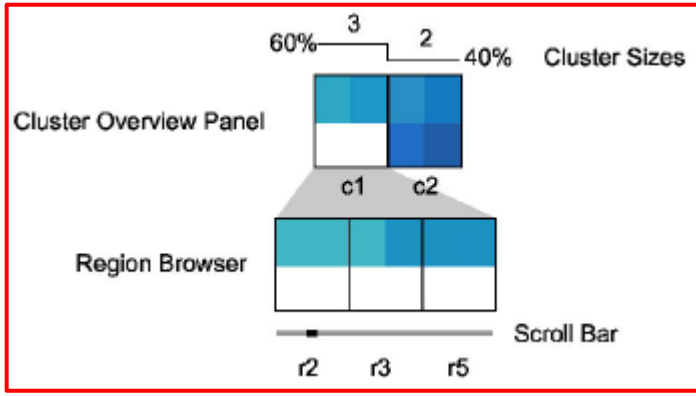


II. Clustering (C1, C2)



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

Watch Video

http://sparkinsight.org/

Nielson et. al., Gen Res 2012

Methodology: GREAT & HOMER

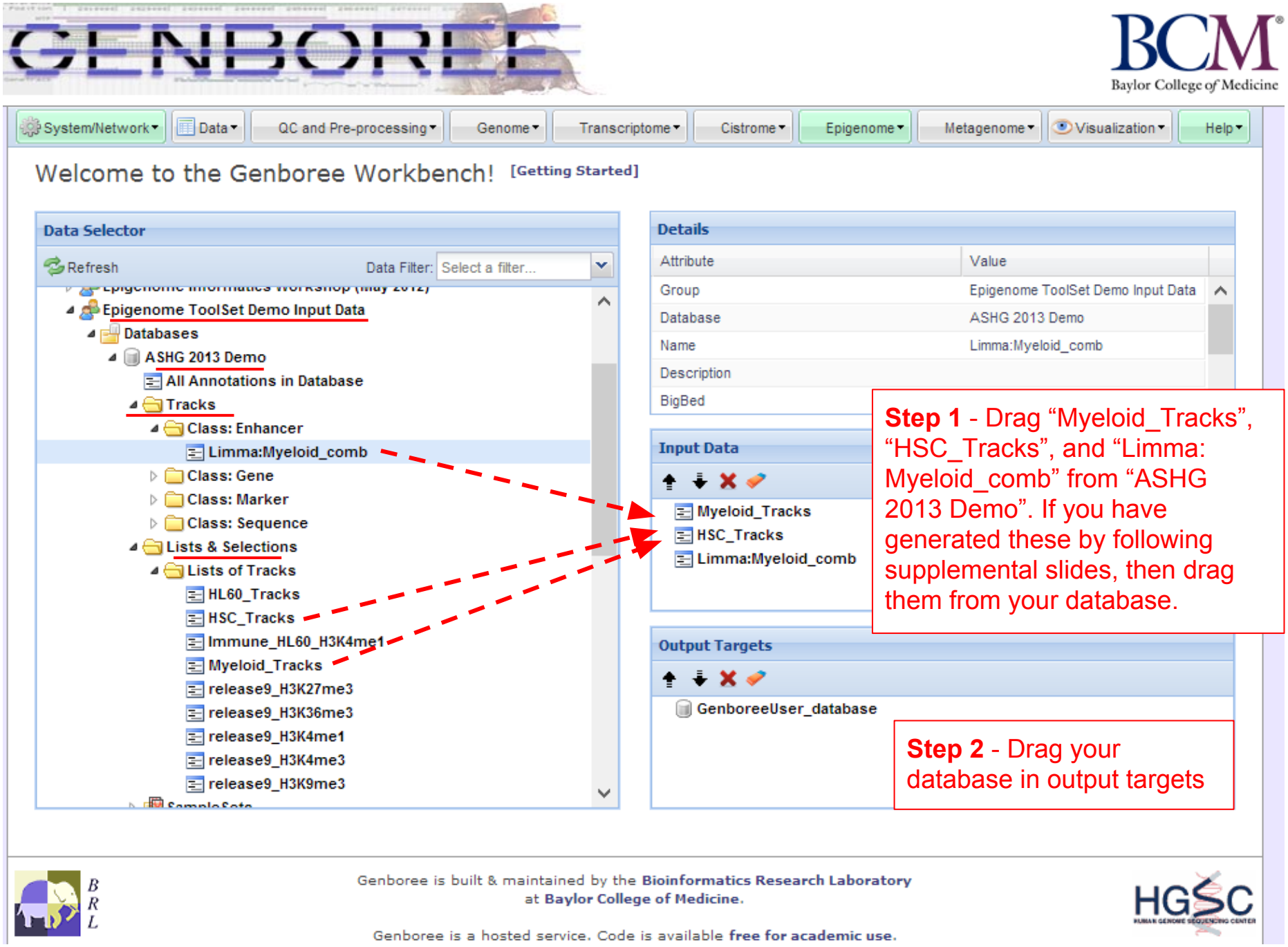
The patterns identified by Spark require further investigation to assess biological meaning

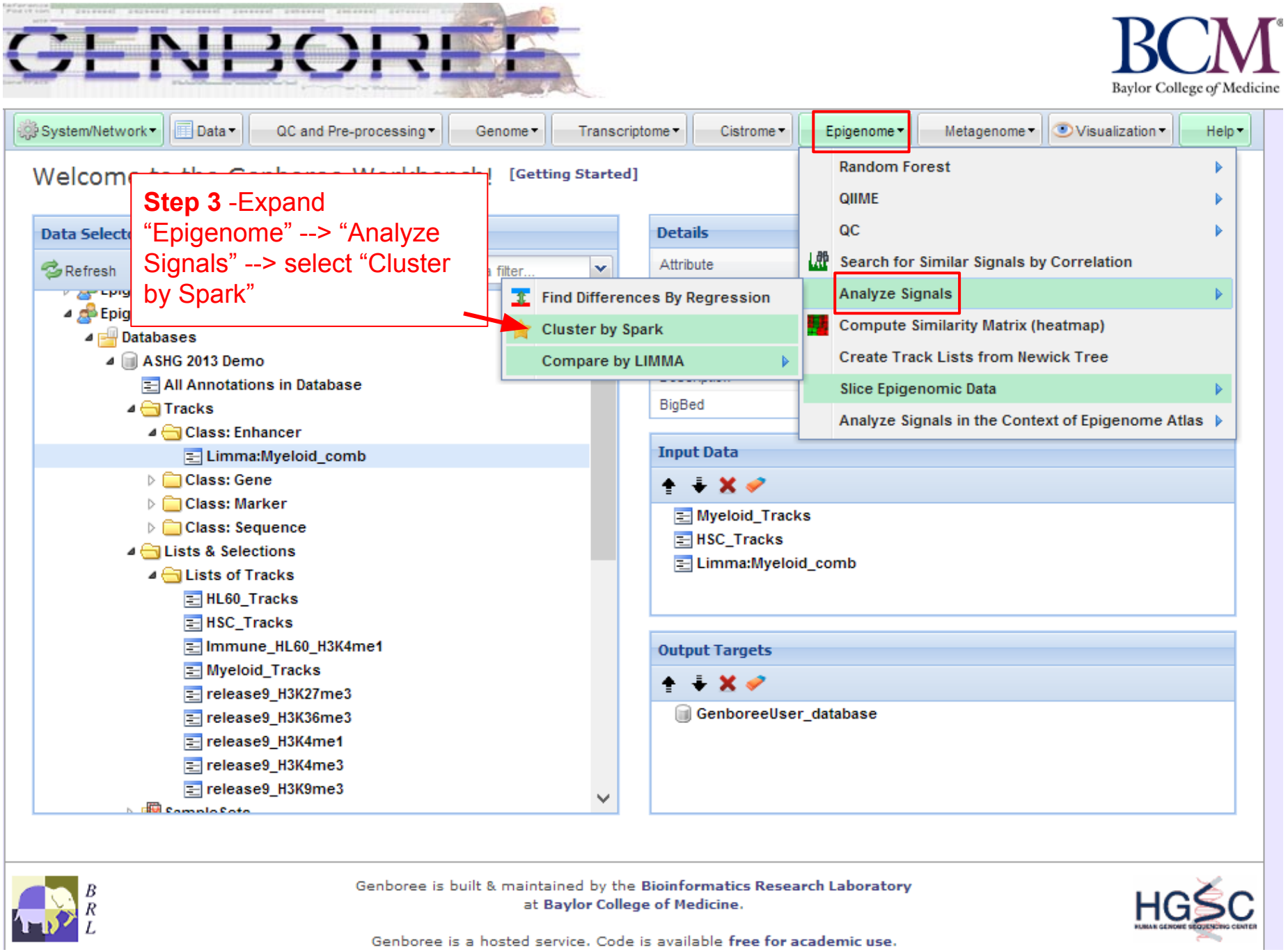
- GREAT[2] (Genomic Regions Enrichment of Annotations Tool) analysis
  - Assesses functional significance of cis-regulatory regions, here identified by Spark
  - Associates both proximal and distal binding events (Chip-Seq) with putative target genes
  - Uses gene annotation from several ontologies to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

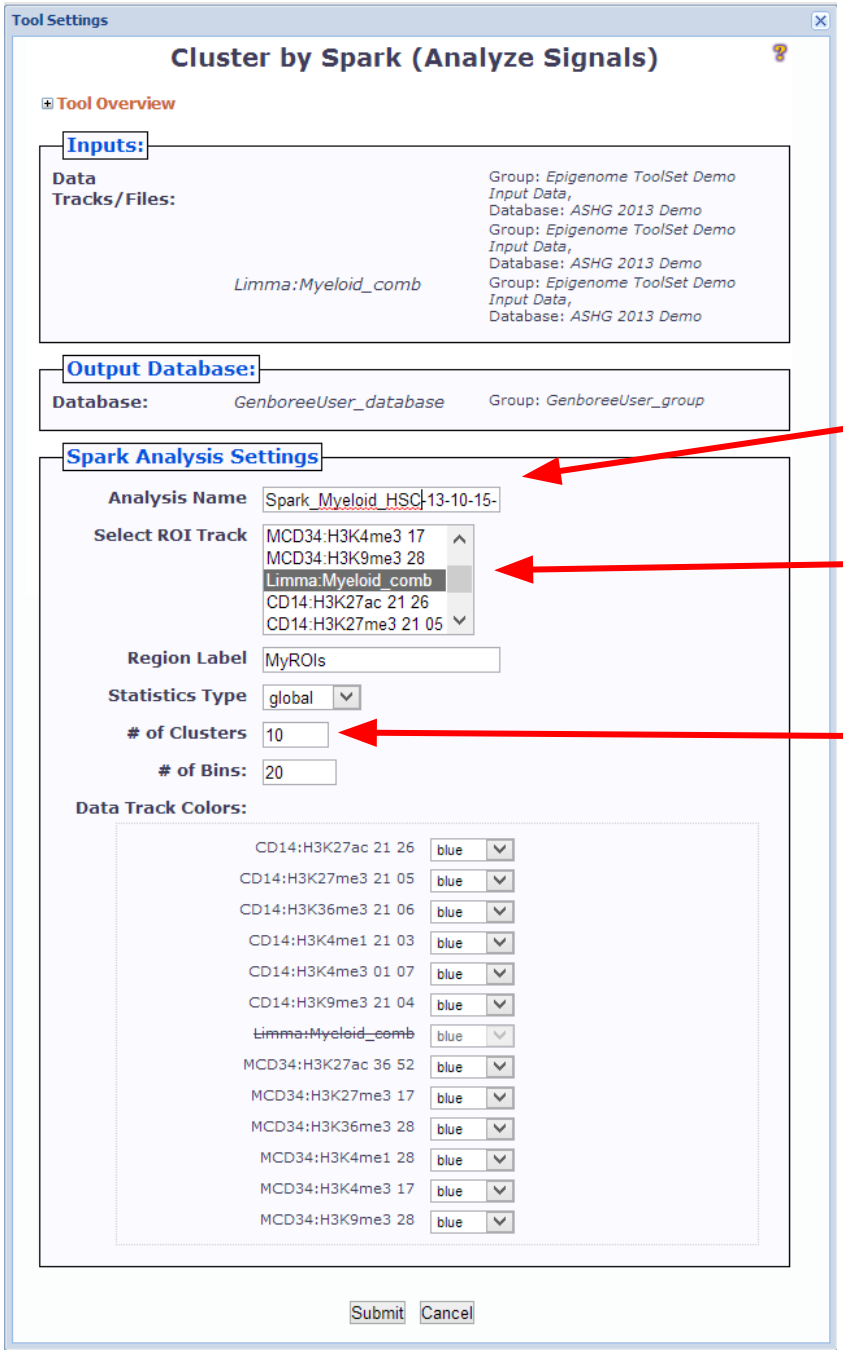
- HOMER[3] analysis:
  - Utilizes a de novo motif discovery algorithm
  - Scores motifs by looking for motifs with differential enrichment between two sets of sequences
  - Enrichment measured using the cummulative hypergeometric distribution (or cummulative binomial distribution for large data sets)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


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]
Step 3 -Expand "Epigenome" --> "Analyze Signals" --> select "Cluster by Spark"
Data Selecto
Refresh
Epig
Databases
ASHG 2013 Demo
All Annotations in Database
Tracks
Class: Enhancer
Limma:Myeloid_comb
Class: Gene
Class: Marker
Class: Sequence
Lists & Selections
Lists of Tracks
HL60_Tracks
HSC_Tracks
Immune_HL60_H3K4me1
Myeloid_Tracks
release9_H3K27me3
release9_H3K36me3
release9_H3K4me1
release9_H3K4me3
release9_H3K9me3
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
BigBed
Input Data
Myeloid_Tracks
HSC_Tracks
Limma:Myeloid_comb
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.
BRL
HGSC

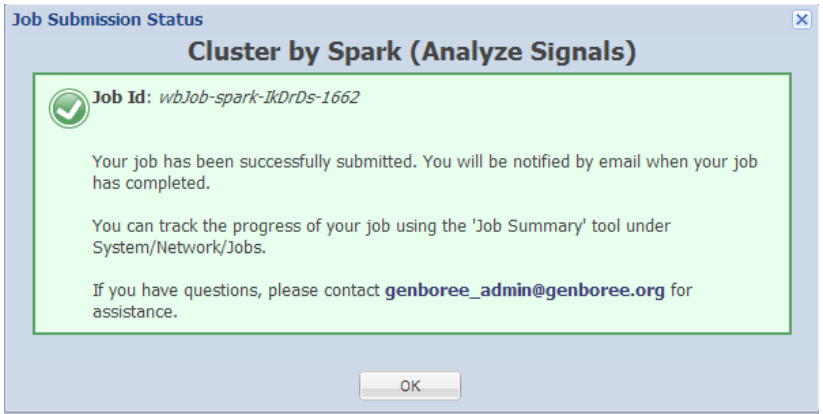


Step 4 - Analysis Name "Spark_Myeloid_HSC"

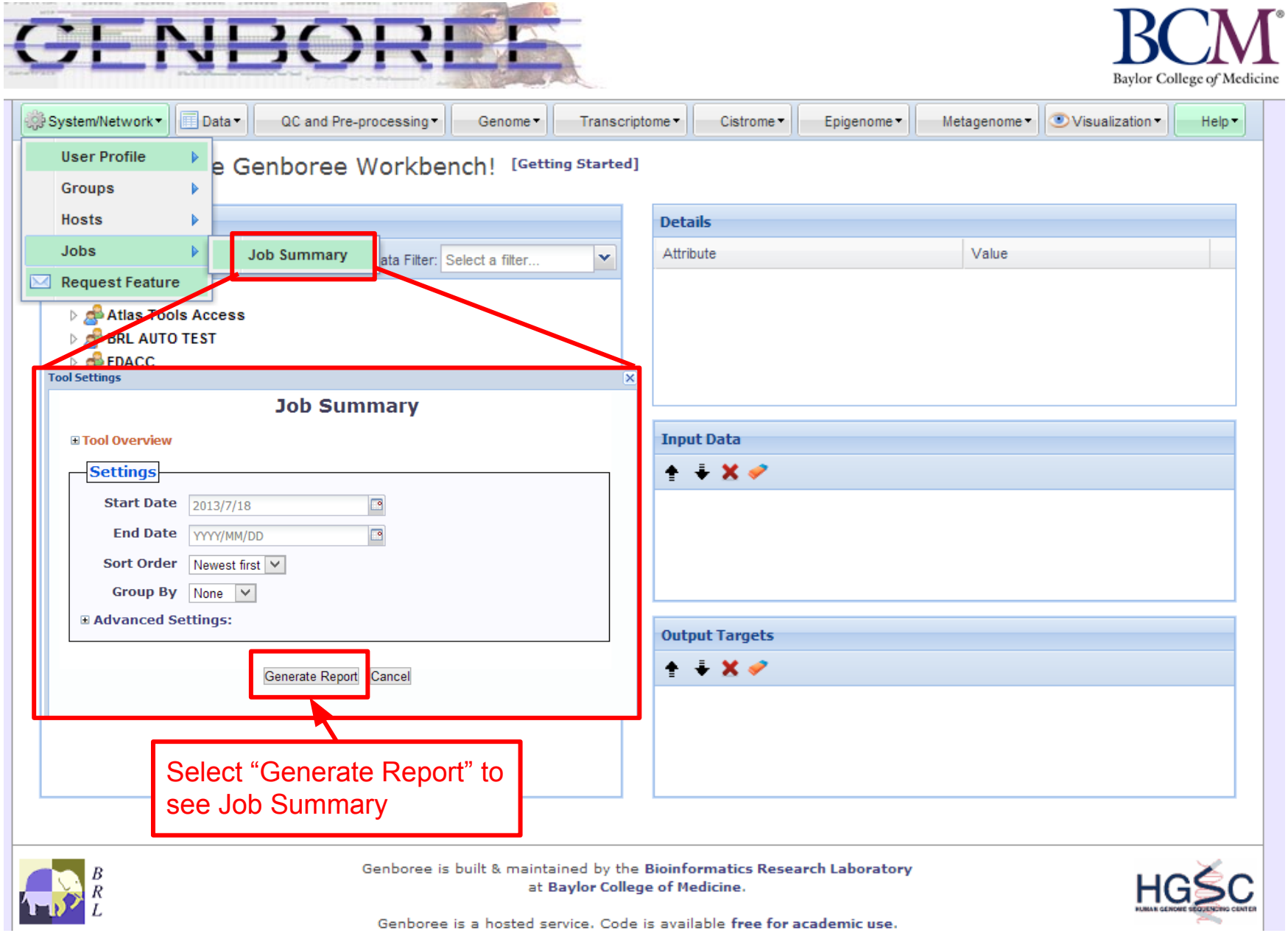
Step 5 - Select ROI (Region of Interest) Track "Limma_Myeloid_comb

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:



Status of the jobs submitted can be obtained through Job Summary



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

Your Spark job completed successfully.

Job Summary:
JobID - wbJob-spark-AsqKIJ-9045
Analysis Name -
Inputs:
# of Data Tracks - 12
ROI Track - Limma:Myeloid_comb
Outputs:
Output DB - UseCase
Output Host - genboree.org
Settings:
k - 10
normType - exp
numBins - 20
regionLabel -
statsType - global

Additional Info:
To view your results in the Spark GUI:
(a) download and unzip the results archive and then
(b) launch Spark via Java Web Start and open the analysis folder.
Spark Java Web Start Link:
http://www.bcgsc.ca/downloads/spark/current/start.jnlp

- The Genboree Team

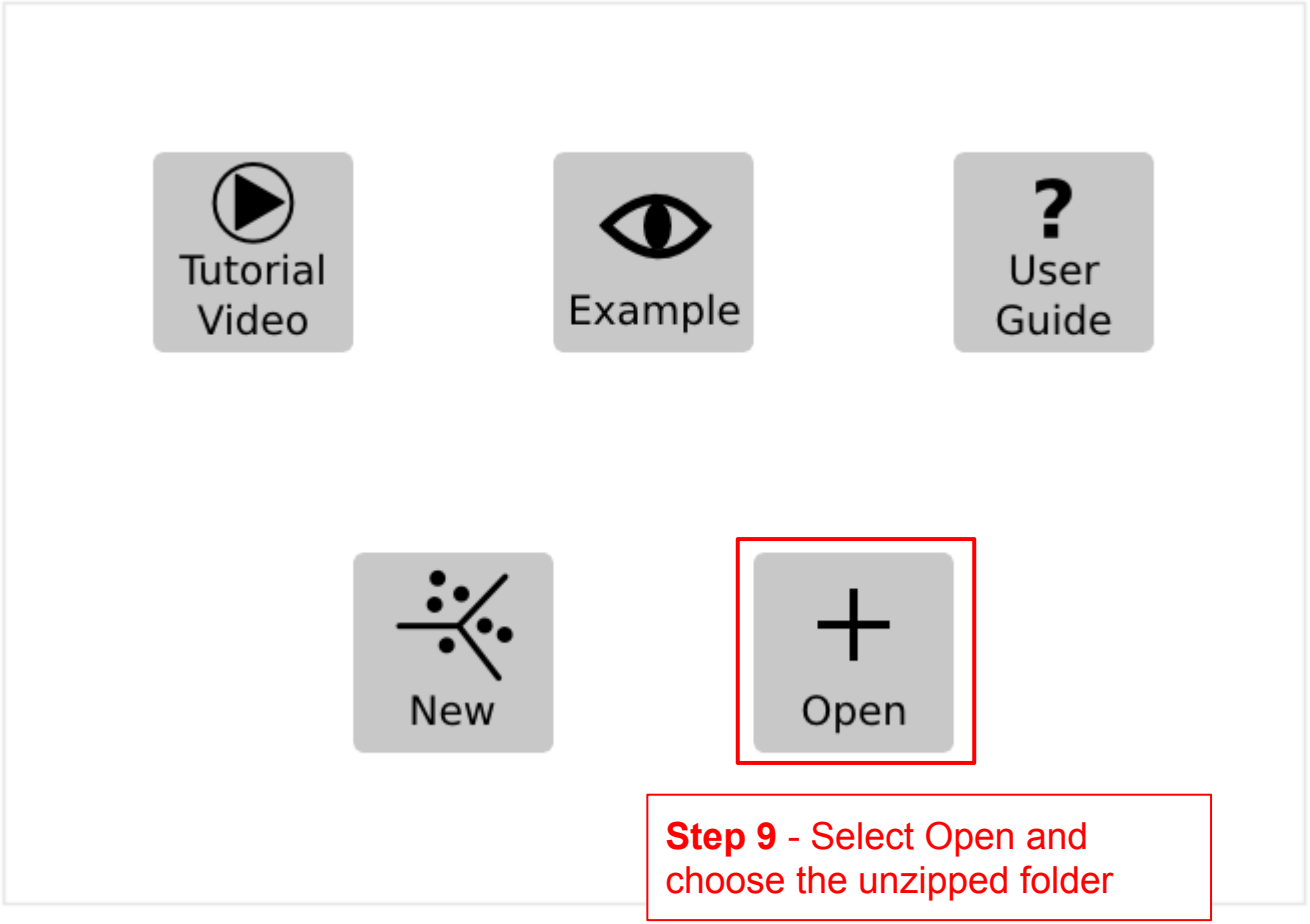
Result File Location in the Genboree Workbench:
(Direct links to files are at the end of this email)
Host: genboree.org
Grp:
Db:
Files Area:
* Spark - Results/
*
*

Result File URLs (click or paste in browser to access file):
FILE: Spark_Myeloid_HSC.zip
URL:
http://genboree.org/java-bin/apiCaller.jsp?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

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

Step 8 - Download Spark Results and UNZIP the Folder

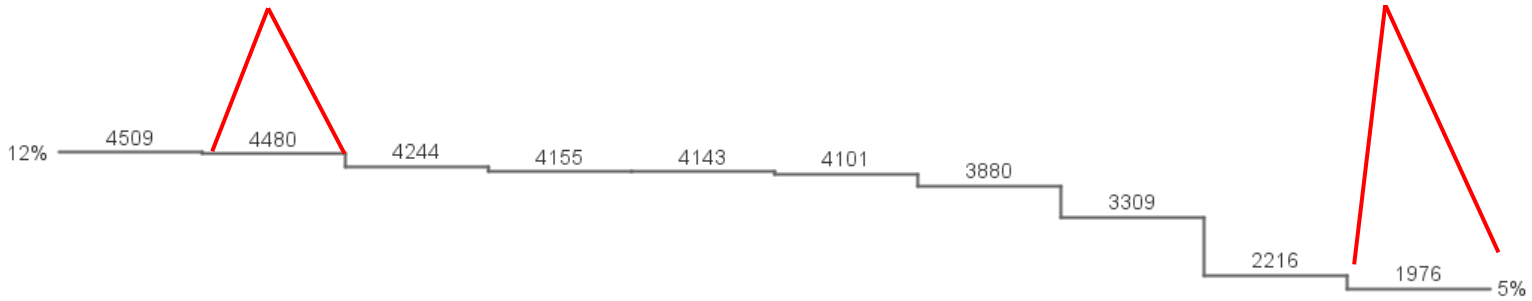
SPARK GUI

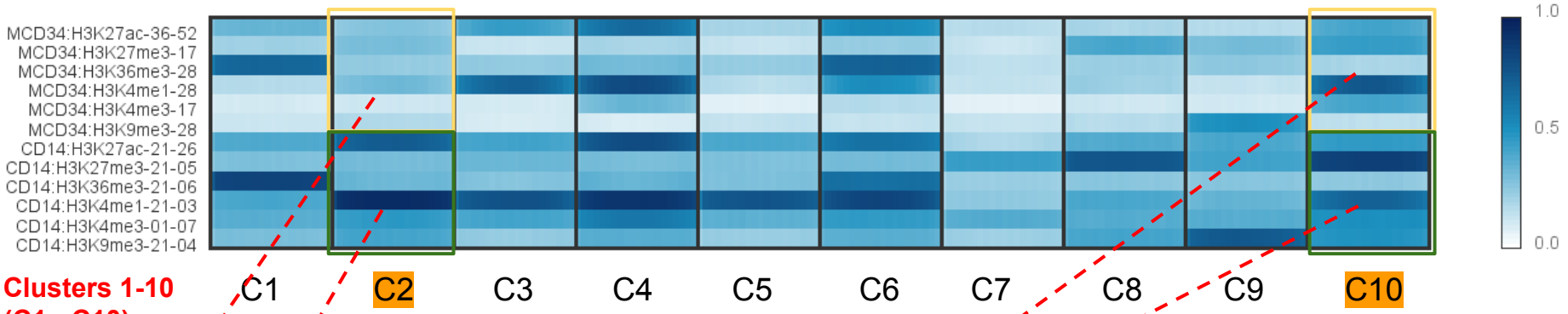

Tutorial Video
Example
User Guide
New
Open
Step 9 - Select Open and choose the unzipped folder

SPARK output results

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

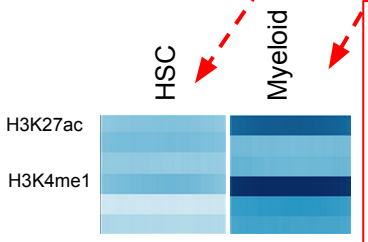
1976 Enhancers (out of ~35000) have this profile of epigenomic marks.



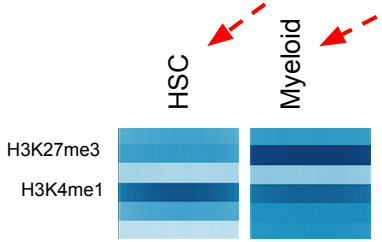


Clusters 1-10 (C1 - C10)

C1 C2 C3 C4 C5 C6 C7 C8 C9 C10

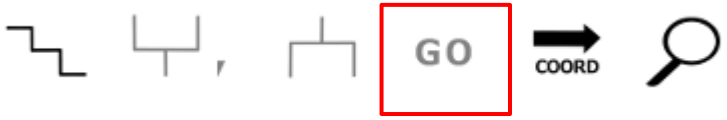


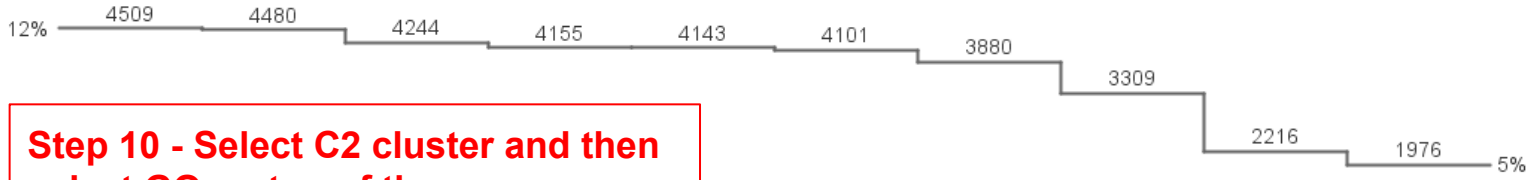
Clustering (C2) reveals a shift from inactive epigenomic marks in human immune stem cells (MD34) to active epigenomic marks in myeloid cells (CD14).



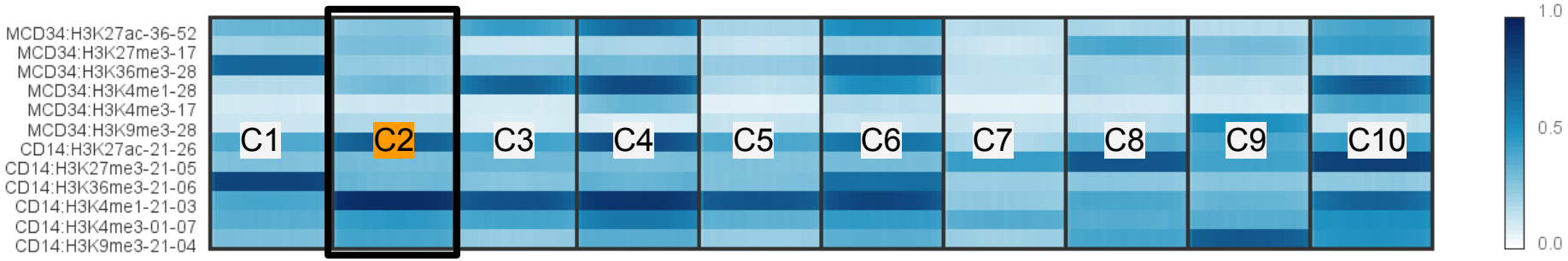
Clustering (C10) reveals a shift from active epigenomic marks in human immune stem cells (MD34) to poised epigenomic marks in myeloid cells.(CD14)

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

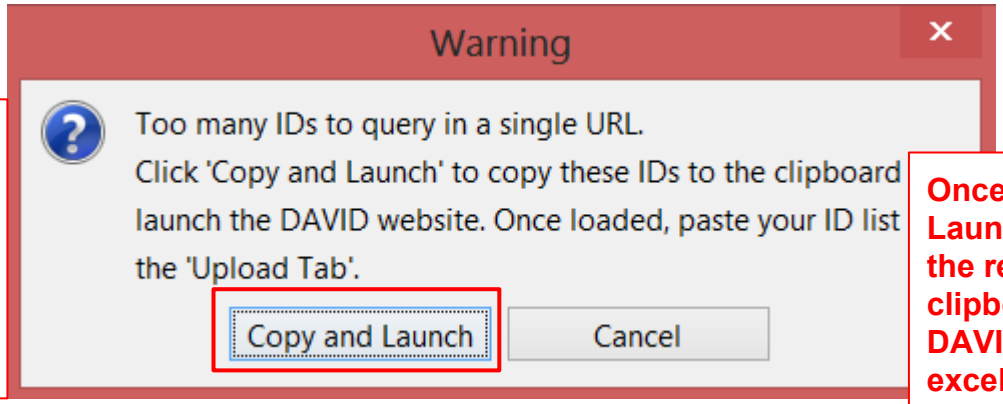




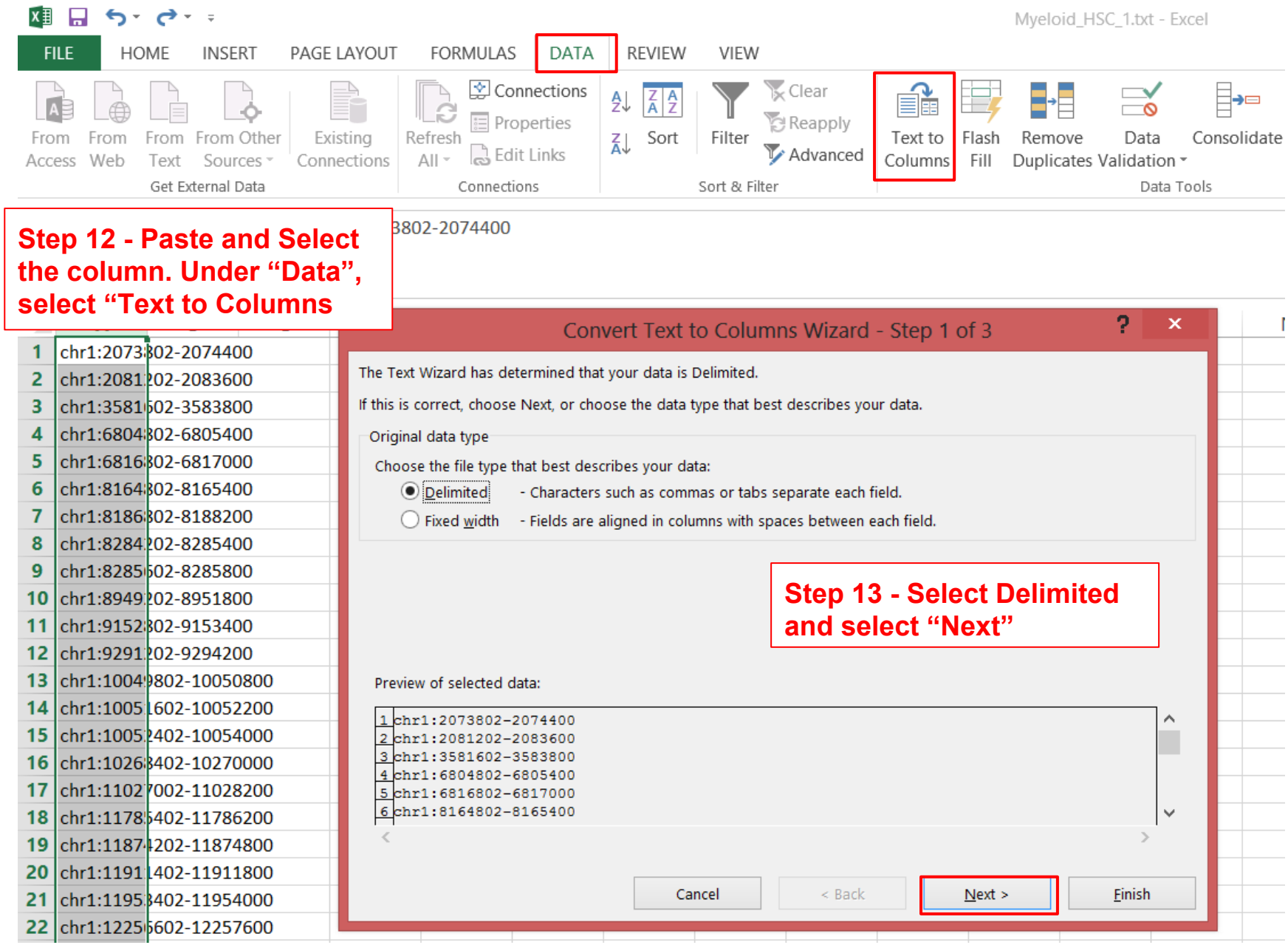
Step 10 - Select C2 cluster and then select GO on top of the page

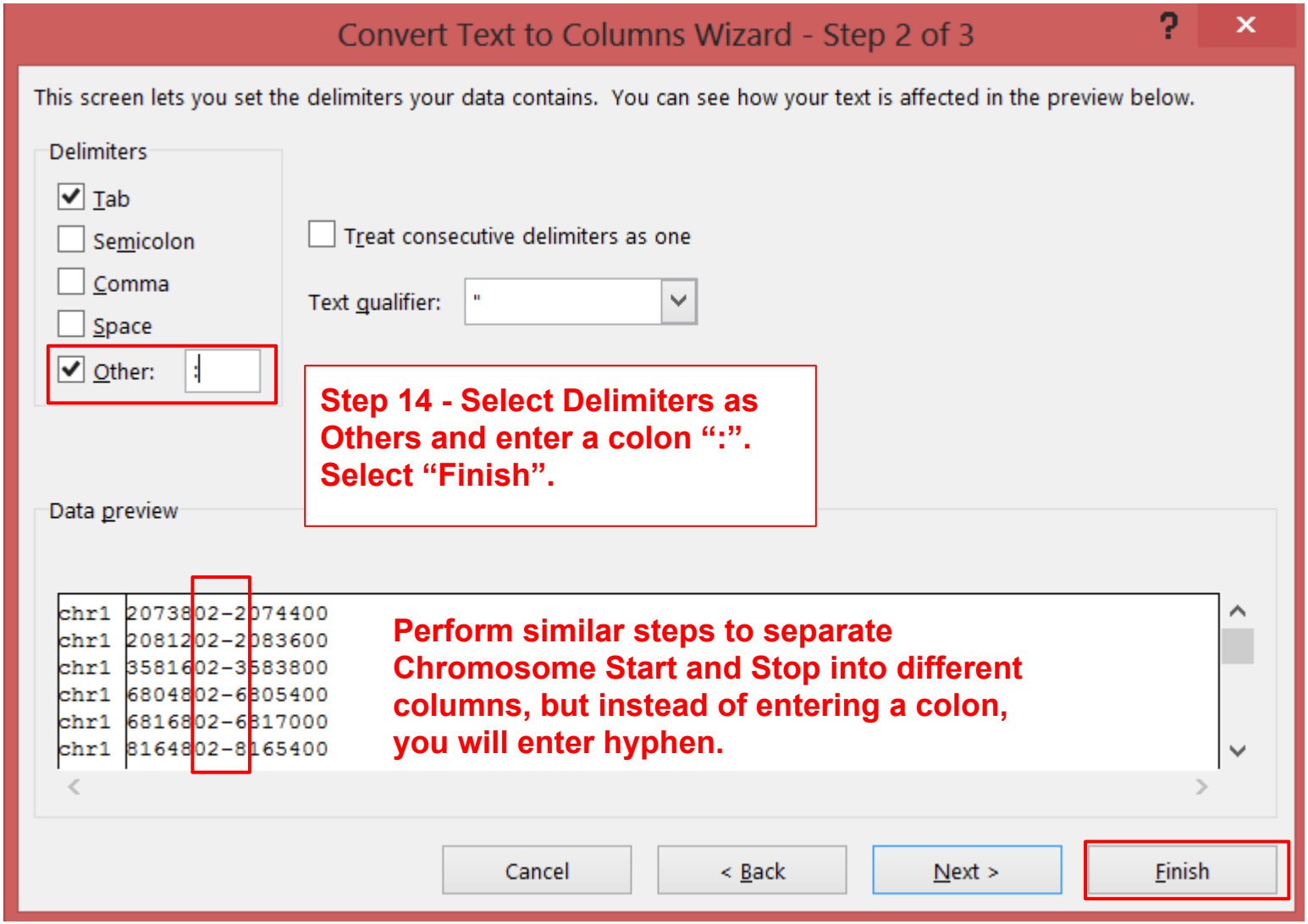


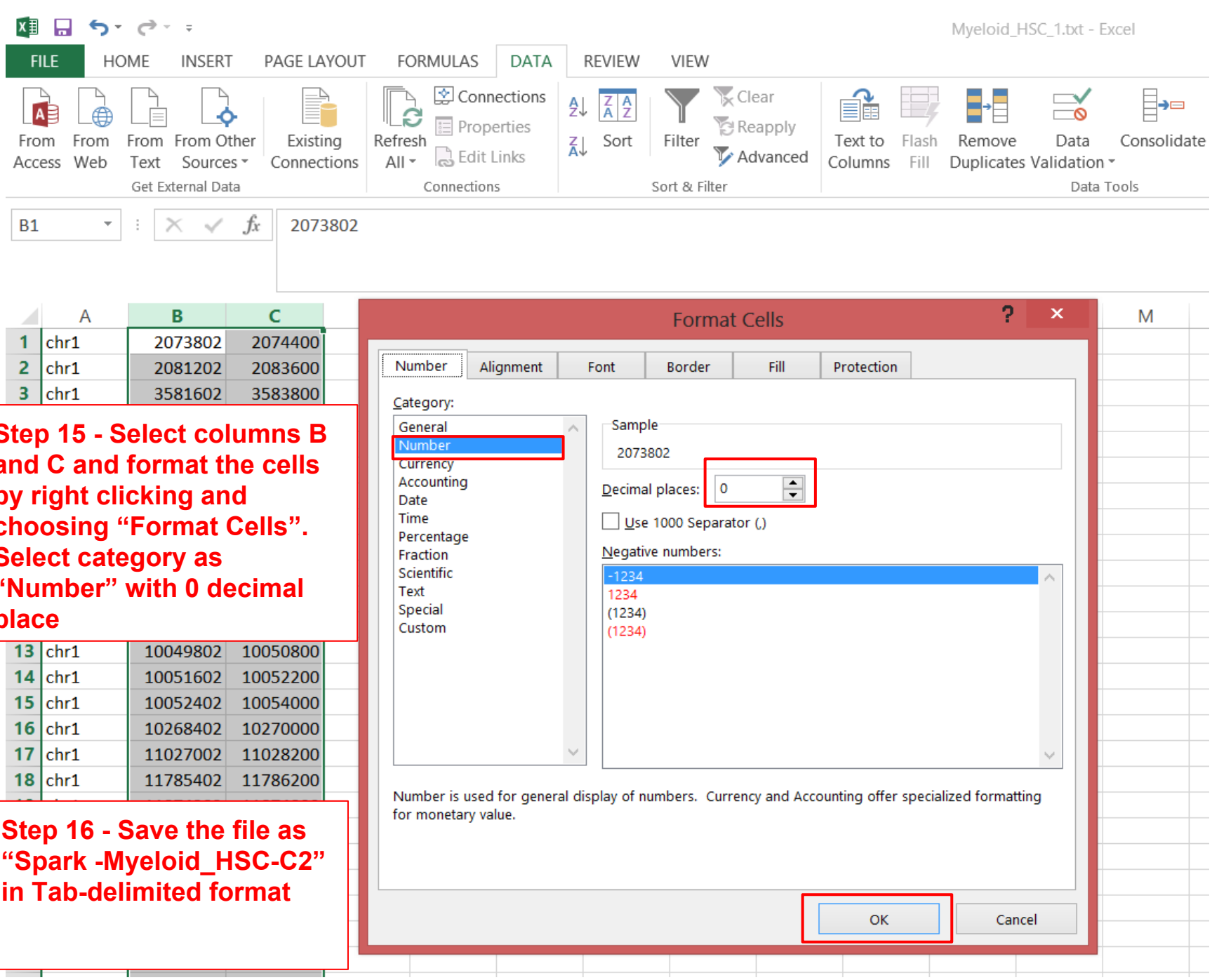
Step 11 - Select Copy and Launch. Web-browser will be launched. We will not use DAVID as regions are enhancers and not genes. We will do Motif and GREAT analysis via Genboree.

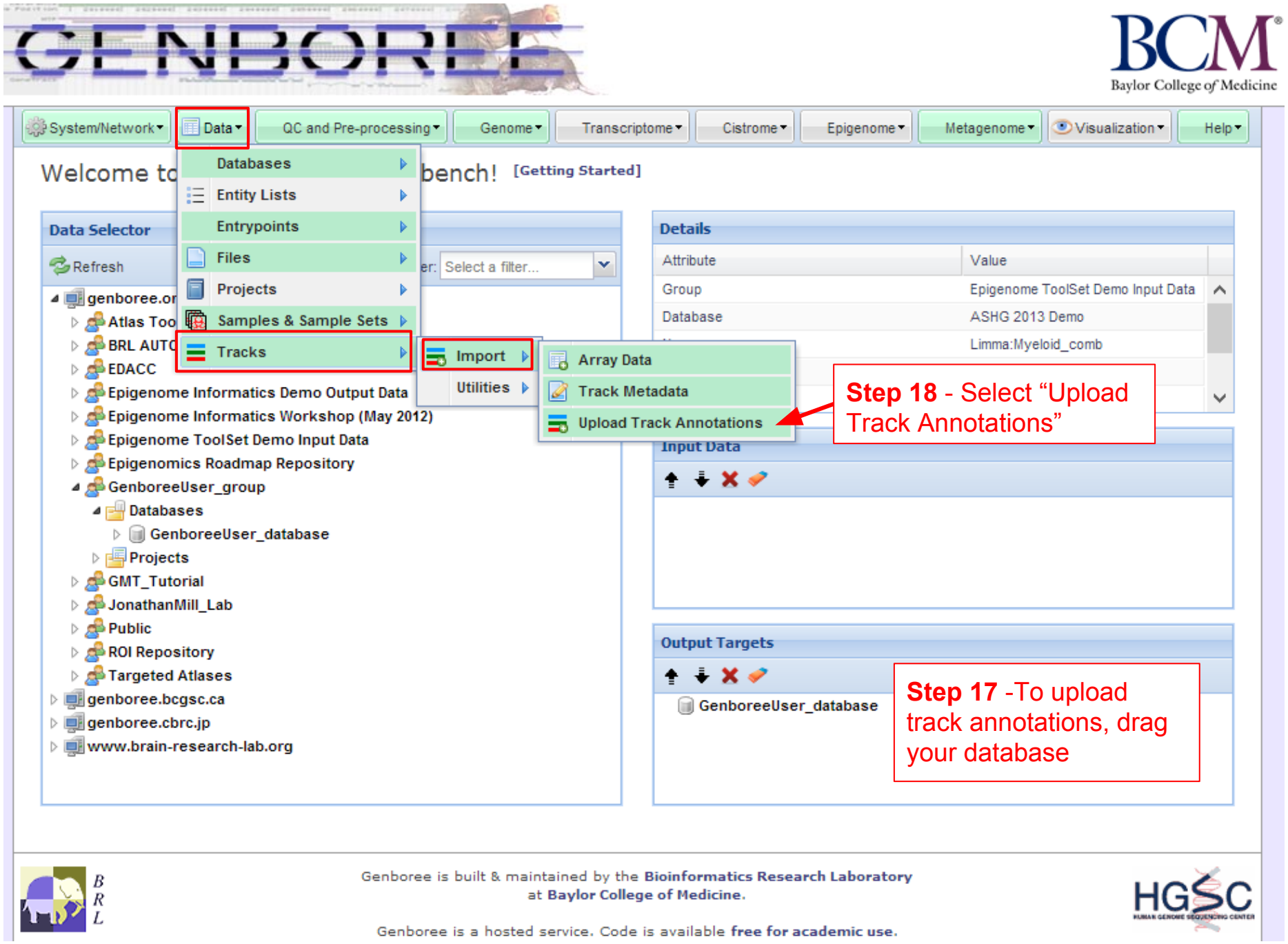


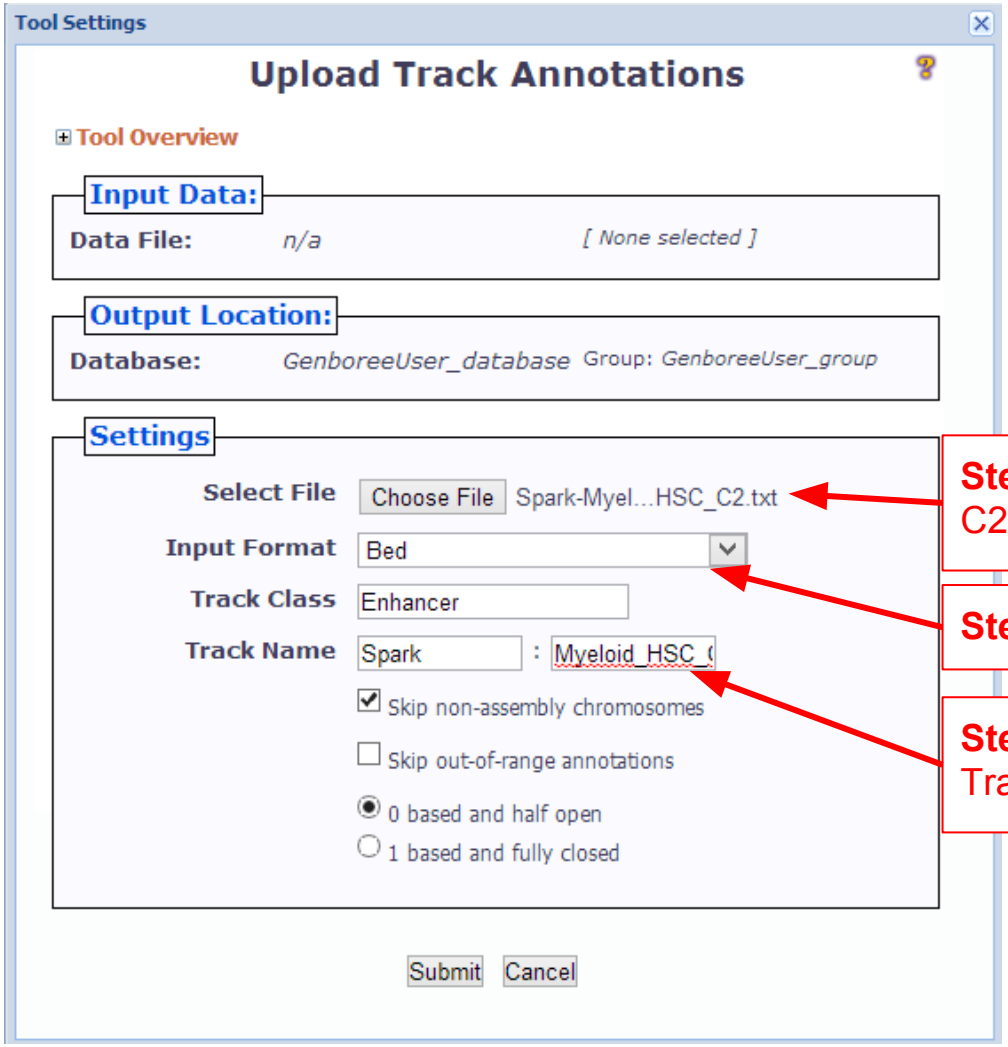
Once selected "Copy and Launch". This action copies the region IDs to your clipboard and launches the DAVID website. Next, open excel and paste the region IDs (which exist on your clipboard) to make a BED file


Myeloid_HSC_1.txt - 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
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.
Step 13 - Select Delimited and select "Next"
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
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 20738 02-2074400
chr1 20812 02-2083600
chr1 35816 02-3583800
chr1 68048 02-6805400
chr1 68168 02-6817000
chr1 81648 02-8165400
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


Myeloid_HSC_1.txt - 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
B1
2073802
A
B
C
M
1 chr1 2073802 2074400
2 chr1 2081202 2083600
3 chr1 3581602 3583800
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 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
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
Step 16 - Save the file as “Spark -Myeloid_HSC-C2” in Tab-delimited format


GENBOREE
BCM
Baylor College of Medicine
System/Network
Data
QC and Pre-processing
Genome
Transcriptome
Cistrome
Epigenome
Metagenome
Visualization
Help
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"
Welcome to
bench!
[Getting Started]
Data Selector
Refresh
Select a filter...
genboree.or
Atlas Too
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
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.
BRL
HGSC

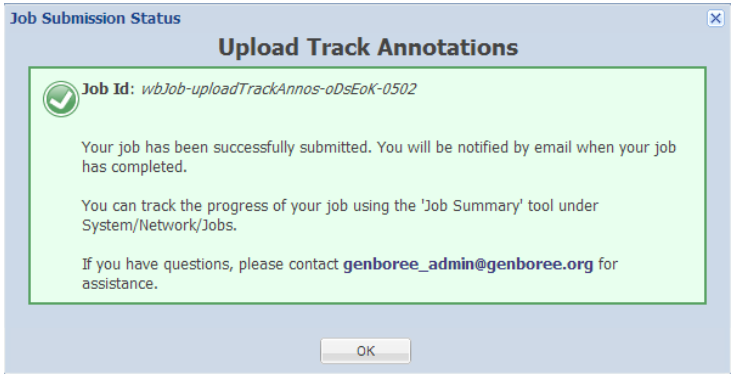


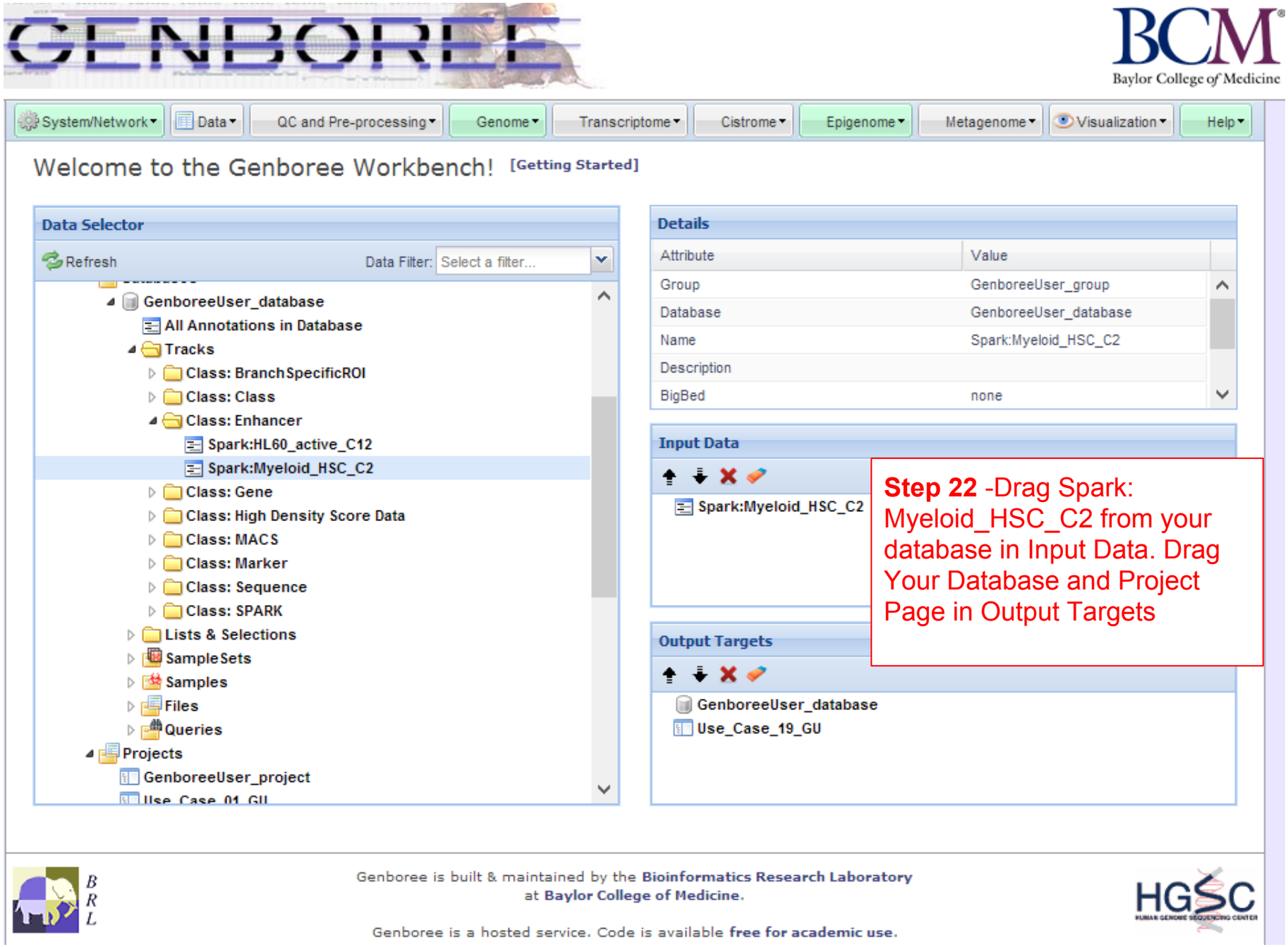
Step 19 - Choose File "Spark-Myeloid_HSC-C2.txt" to upload

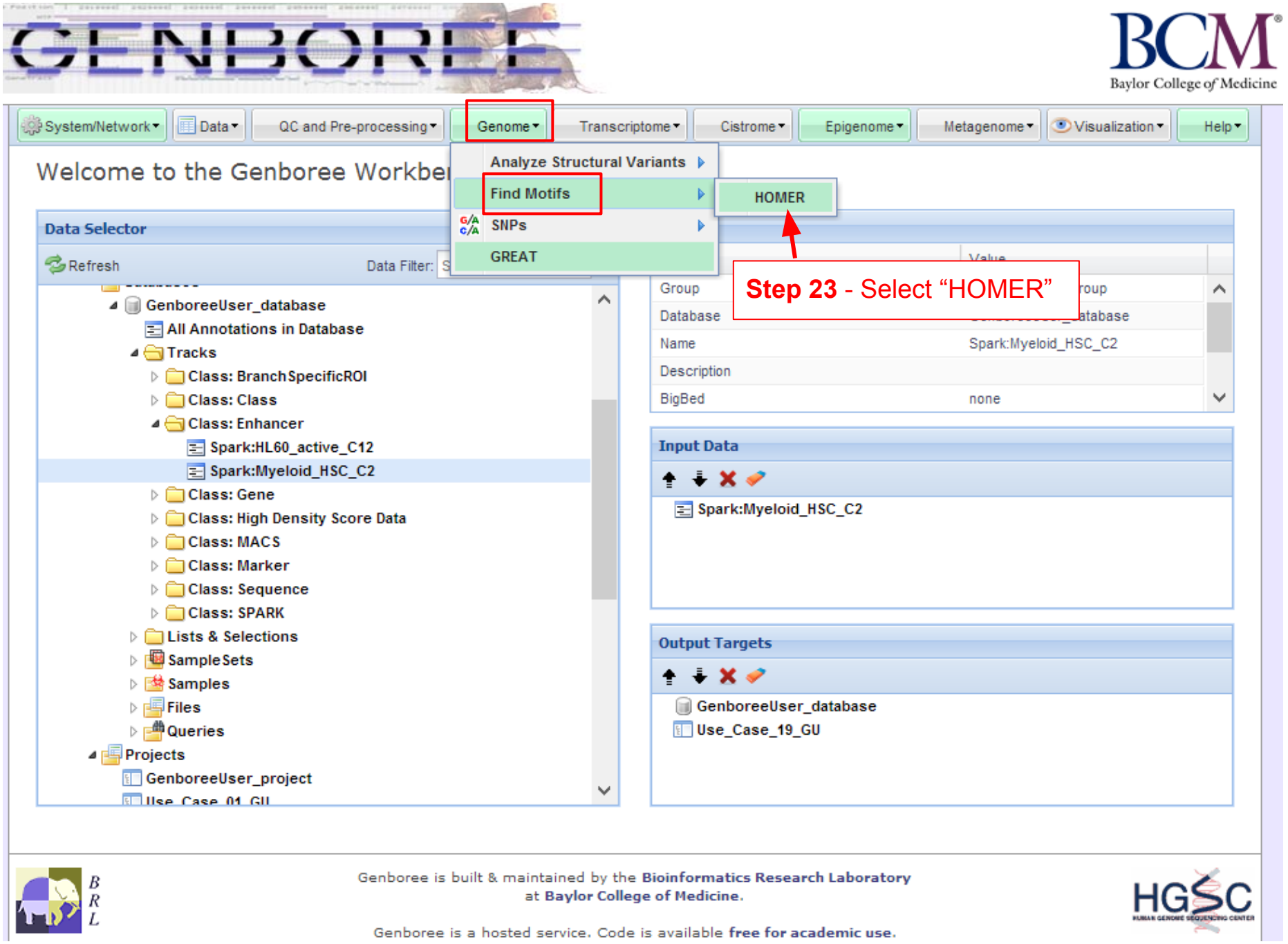
Step 20 - Select Bed

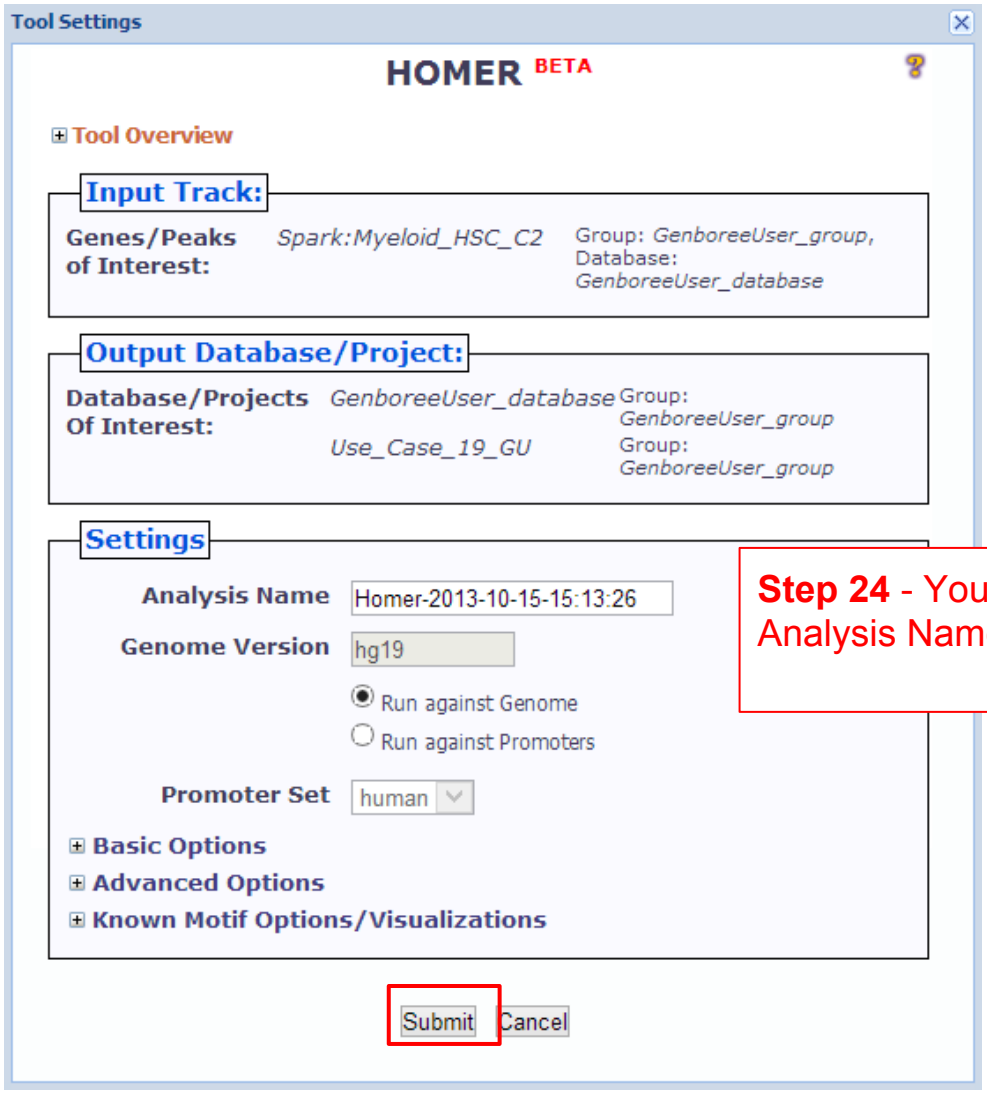
Step 21 - Name Track Class as "Enhancer", Track Name as "Spark:Myeloid_HSC_C2"

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



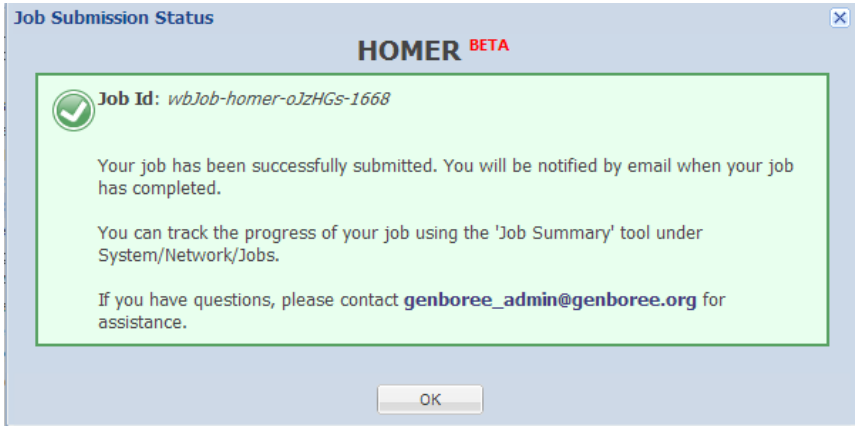

GENBOREE
BCM
Baylor College of Medicine
System/Network
Data
QC and Pre-processing
Genome
Transcriptome
Cistrome
Epigenome
Metagenome
Visualization
Help
Welcome to the Genboree Workbench!
[Getting Started]
Data Selector
Refresh
Data Filter:
Select a filter...
GenboreeUser_database
All Annotations in Database
Tracks
Class: BranchSpecificROI
Class: Class
Class: Enhancer
Spark:HL60_active_C12
Spark:Myeloid_HSC_C2
Class: Gene
Class: High Density Score Data
Class: MACS
Class: Marker
Class: Sequence
Class: SPARK
Lists & Selections
SampleSets
Samples
Files
Queries
Projects
GenboreeUser_project
Use_Case_01_GU
Details
Attribute
Value
Group
GenboreeUser_group
Database
GenboreeUser_database
Name
Spark:Myeloid_HSC_C2
Description
BigBed
none
Input Data
Spark:Myeloid_HSC_C2
Output Targets
GenboreeUser_database
Use_Case_19_GU
Step 22 -Drag Spark: Myeloid_HSC_C2 from your database in Input Data. Drag Your Database and Project Page in 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

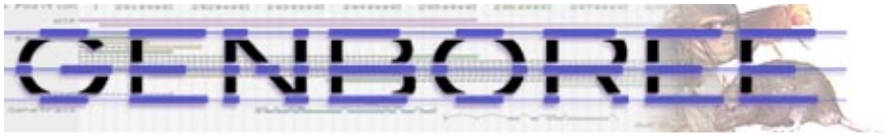

GENBOREE
BCM
Baylor College of Medicine
System/Network
Data
QC and Pre-processing
Genome
Transcriptome
Cistrome
Epigenome
Metagenome
Visualization
Help
Analyze Structural Variants
Find Motifs
SNPs
GREAT
HOMER
Step 23 - Select "HOMER"
Welcome to the Genboree Workbe
Data Selector
Refresh
Data Filter:
GenboreeUser_database
All Annotations in Database
Tracks
Class: BranchSpecificROI
Class: Class
Class: Enhancer
Spark:HL60_active_C12
Spark:Myeloid_HSC_C2
Class: Gene
Class: High Density Score Data
Class: MACS
Class: Marker
Class: Sequence
Class: SPARK
Lists & Selections
SampleSets
Samples
Files
Queries
Projects
GenboreeUser_project
Use_Case_01_GU
Value
Group
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.
BRL
HGSC



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

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





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

Step 26 - Select "GREAT"

Welcome to the Genboree Workbe

Data Selector

Refresh Data Filter:

- GenboreeUser_database
  - All Annotations in Database
  - Tracks
    - Class: BranchSpecificROI
    - Class: Class
    - Class: Enhancer
      - Spark:HL60_active_C12
      - Spark:Myeloid_HSC_C2
    - Class: Gene
    - Class: High Density Score Data
    - Class: MACS
    - Class: Marker
    - Class: Sequence
    - Class: SPARK
  - Lists & Selections
  - SampleSets
  - Samples
  - Files
  - Queries
- Projects
  - GenboreeUser_project
  - Use_Case_01_GU

Group	GenboreeUser_group
Database	GenboreeUser_database
Name	Spark:Myeloid_HSC_C2
Description	
BigBed	none

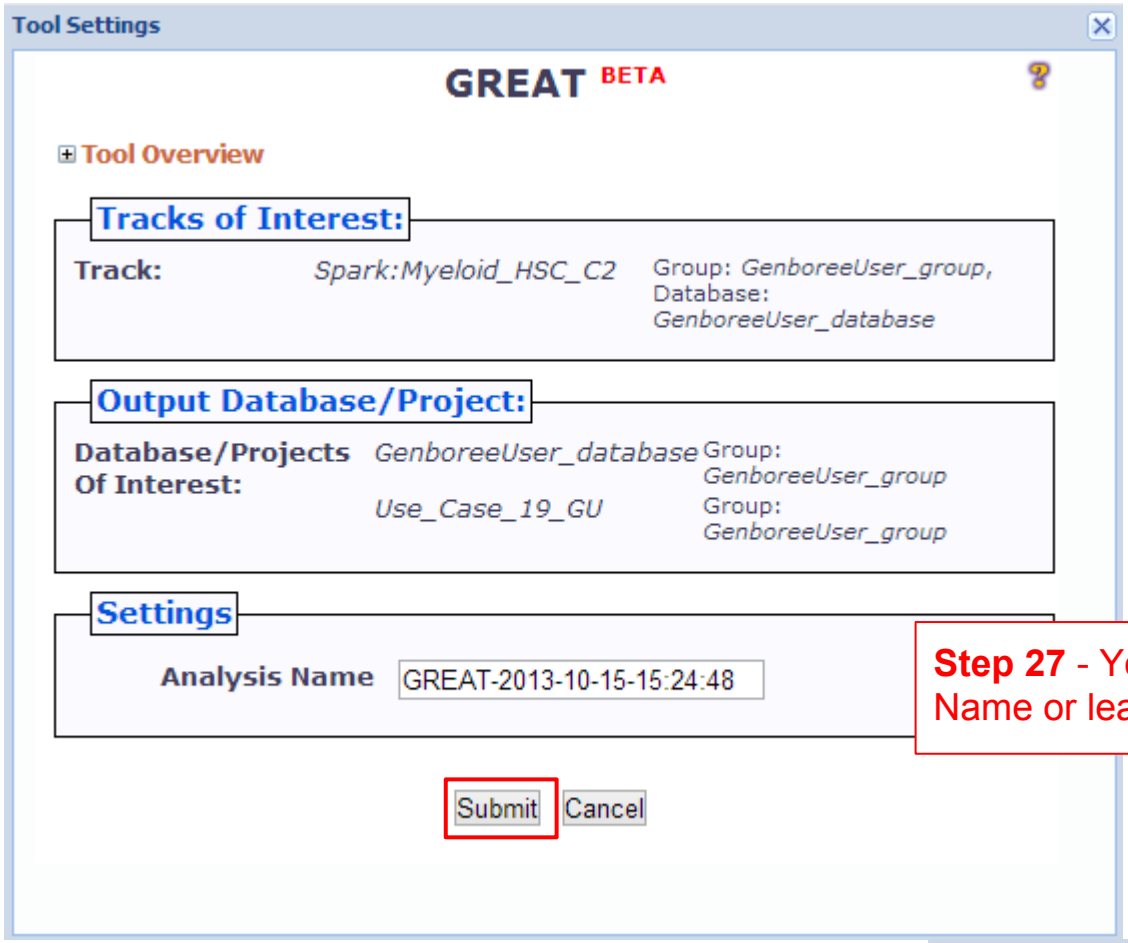
Input Data

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



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

You will see this message upon successful submission of your 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:

2013/10/15: Genboree User ran a GREAT (Genomic Regions Enrichment of Annotations Tool) job (wbJob-great-EmC8lG-0940). Click the link below to perform live analysis with GREAT:

- **Peform GREAT anlaysis**

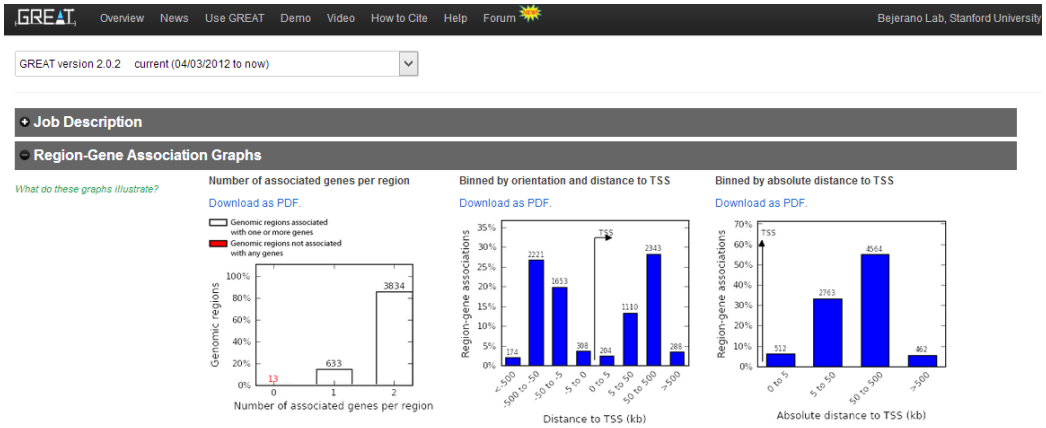
2013/10/15: Genboree User ran a HOMER job (wbJob-homer-oJzHGs-1668) and the results are available at the link below.

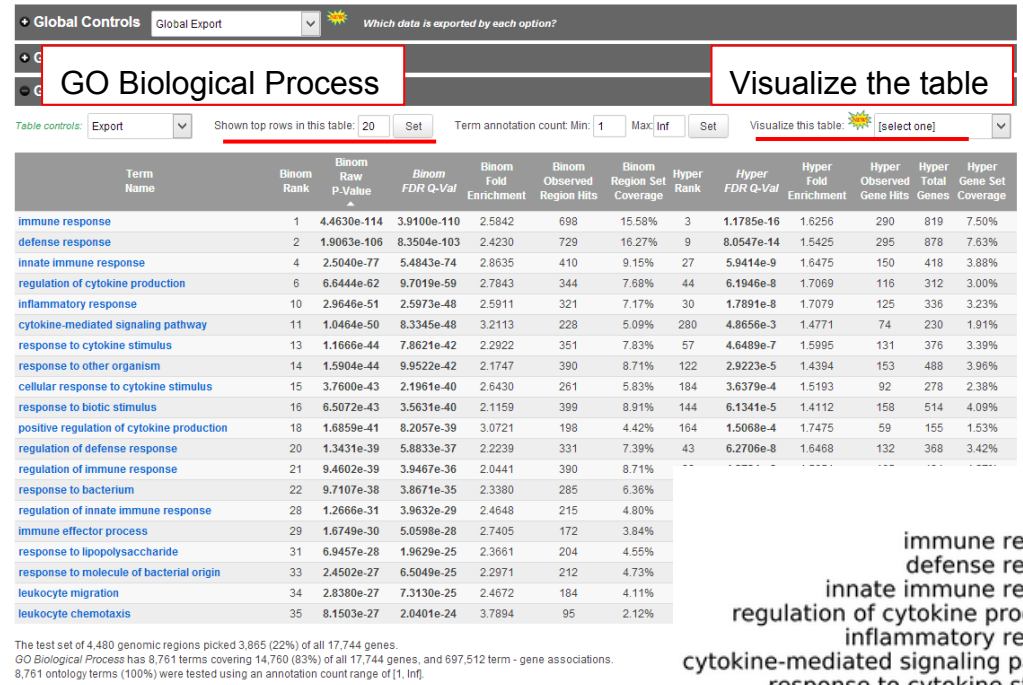
- **Link to Homer Results**

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

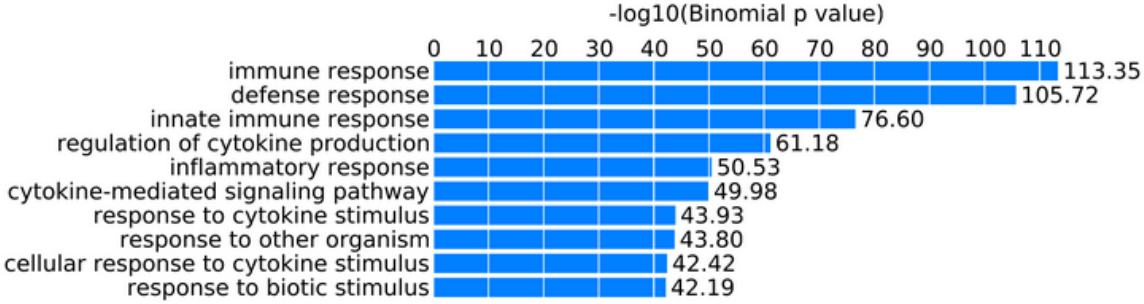




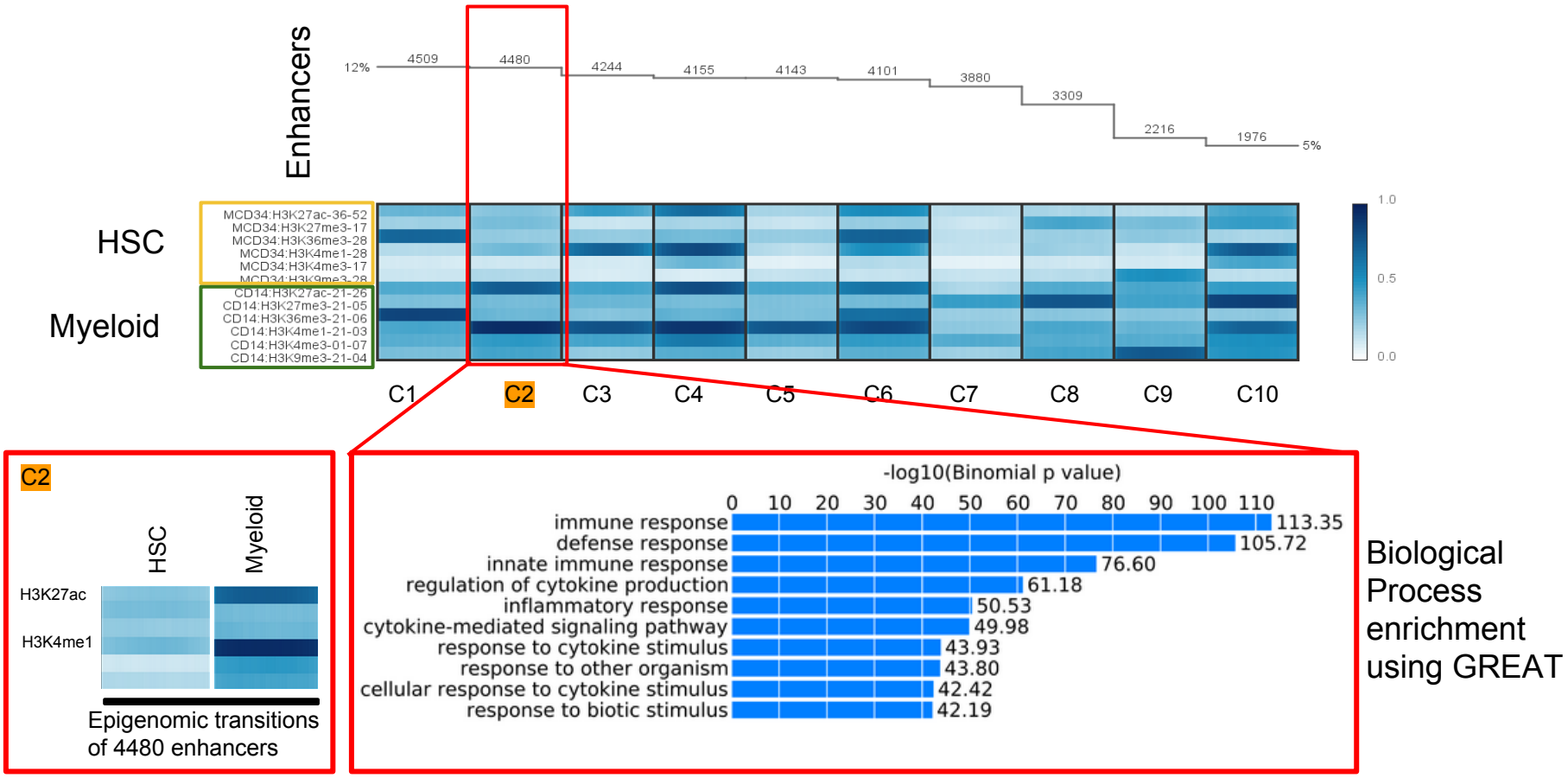
Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Val	Binom Fold Enrichment	Binom Observed Region Hits	Binom Region Set Coverage	Hyper Rank	Hyper FDR Q-Val	Hyper Fold Enrichment	Hyper Observed Gene Hits	Hyper Total Genes	Hyper Gene Set Coverage
immune response	1	4.4630e-114	3.9100e-110	2.5842	698	15.58%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