Use Case 5: Methylation of some regions discriminate tissue type better than others

Epigenome Informatics Workshop
Bioinformatics Research Laboratory

BCM Baylor College of Medicine

The data for this use case was kindly provided by Dr. Jonathan Mill (King's College London, UK), and is taken from the following reference:

"Functional annotation of the human brain methylome across brain and blood". Matthew Davies[1], Manuela Volta[1], Abhishek Dixit[1], Simon Lovestone[1], Cristian Coarfa[2], R. Alan Harris[2], Aleksandar Milosavljevic[2], Claire Troakes[1], Safa Al-Sarraj[1], Richard Dobson[1], Leonard C. Schalkwyk[1], Jonathan Mill[1]* Genome Biology, 12:R43, 2012

[1]Institute of Psychiatry, King's College London. UK. [2]Baylor College of Medicine, Houston, Texas. USA. *Corresponding Author: Dr. Jonathan Mill, Address: Institute of Psychiatry, SGDP Centre, De Crespigny Park, Denmark Hill, London.

Use Case 5: Methylation of some regions discriminate tissue type better than others

This use case is similar to Use Cases 1 & 2 but with different annotation tracks with different regions of interest (ROIs) to illustrate how some ROIs produce better clustering.

Background: Recent methylomic analyses of various tissues indicate that differential methylation across low CpG-content promoters (LCPs) is associated with tissue-specific gene expression in somatic cells. Based on this observation, Davies et al sought to compare high CpG-content promoters (HCPs) with LCPs across brain regions and blood as a first step in understanding how such regions may impact gene regulatory networks.

Results: LCPs appear to be a major location for tissue-specific DNA methylation signatures across regions in brain and in blood. Hierarchical clustering of both HCP and LCP DNA methylation can distinguish between tissues, although the Euclidean distance between tissues is much larger in the case of LCPs. Principle component analysis of MeDIP-Seq data shows a much stronger classification based upon LCP methylation.

Promoter DNA Methylation in the Human Genome

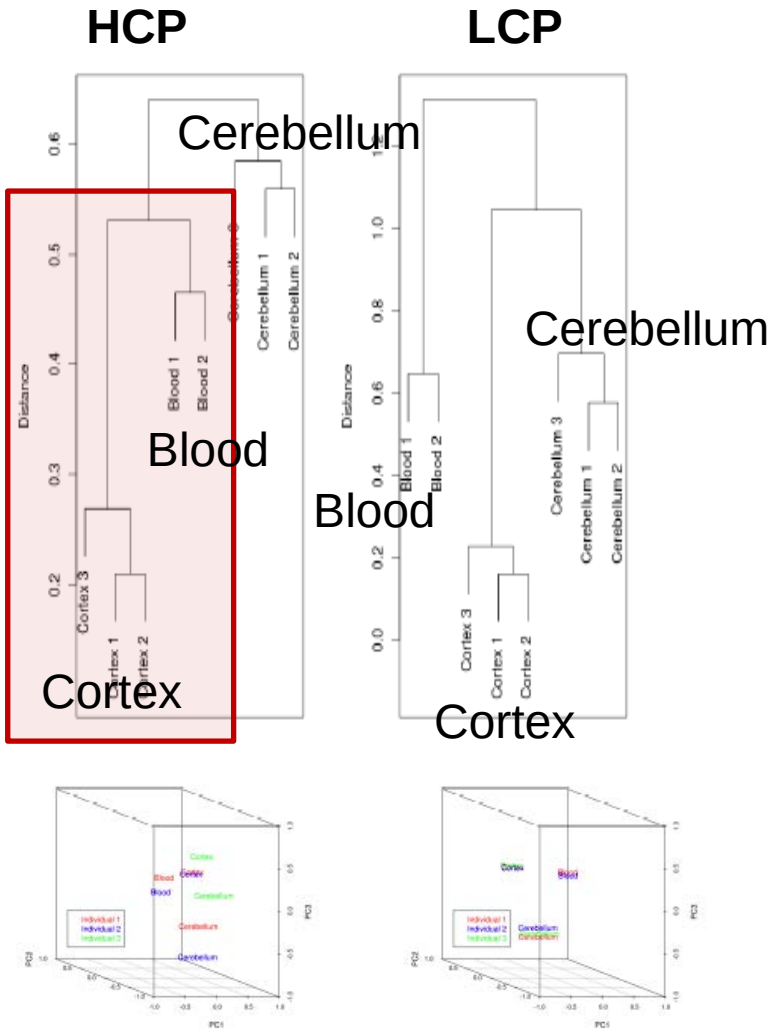
Source of ROIs found in "Class: Regulation" in the Data Selector

- Enriched methylated DNA from human primary fibroblasts using methylated DNA immunoprecipitation (MeDIP) + microarray detection
- 15,609 promoters evaluated in primary somatic and germline cells
- **HCPs** (high-CpG promoters) – contain 500 bp region with CpG ratio above 0.75 and GC content >55%
- **LCP** (low-CpG promoters) – do not contain a 500 bp region with a CpG ratio above 0.48
- **ICP** (intermediate CpG promoters) – are neither HCPs or LCPs. ICP class contains many "subthreshold" CpG islands, meaning small CpG islands (<500 bp), moderate CpG richness and/or GC content <55%

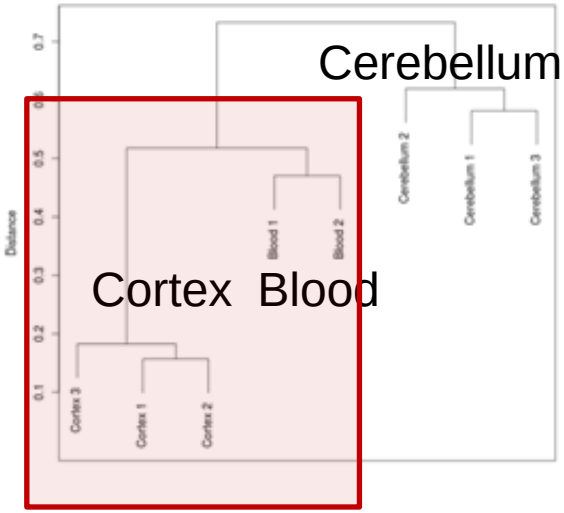
Weber et al, "Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome" *Nature Genetics*, 39 (4), April 2007

Methylation of Low-CpG Promoters (LCP) vs. High-CpG Promoters (HCP)

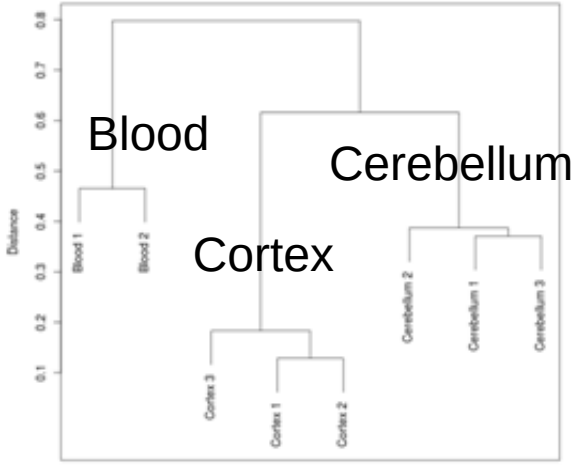
Methylation of LCPs conveys more information about tissue type than methylation of HCPs.



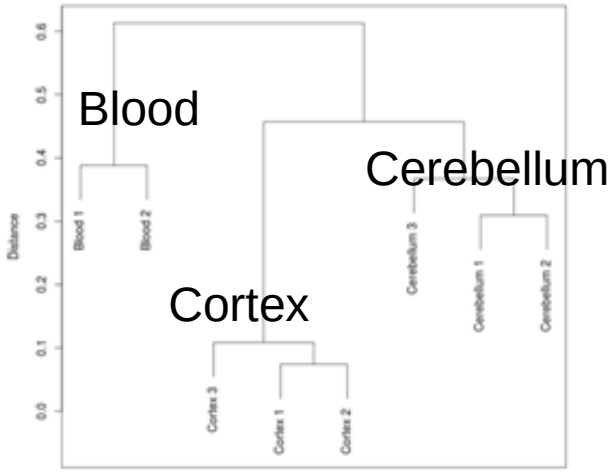
Some features discriminate tissue type better than others (cont'd)



CpG islands



CpG shores



CDS

The following slides walk you through the process of generating the clustering results displayed in the previous slide (from Davies et al).

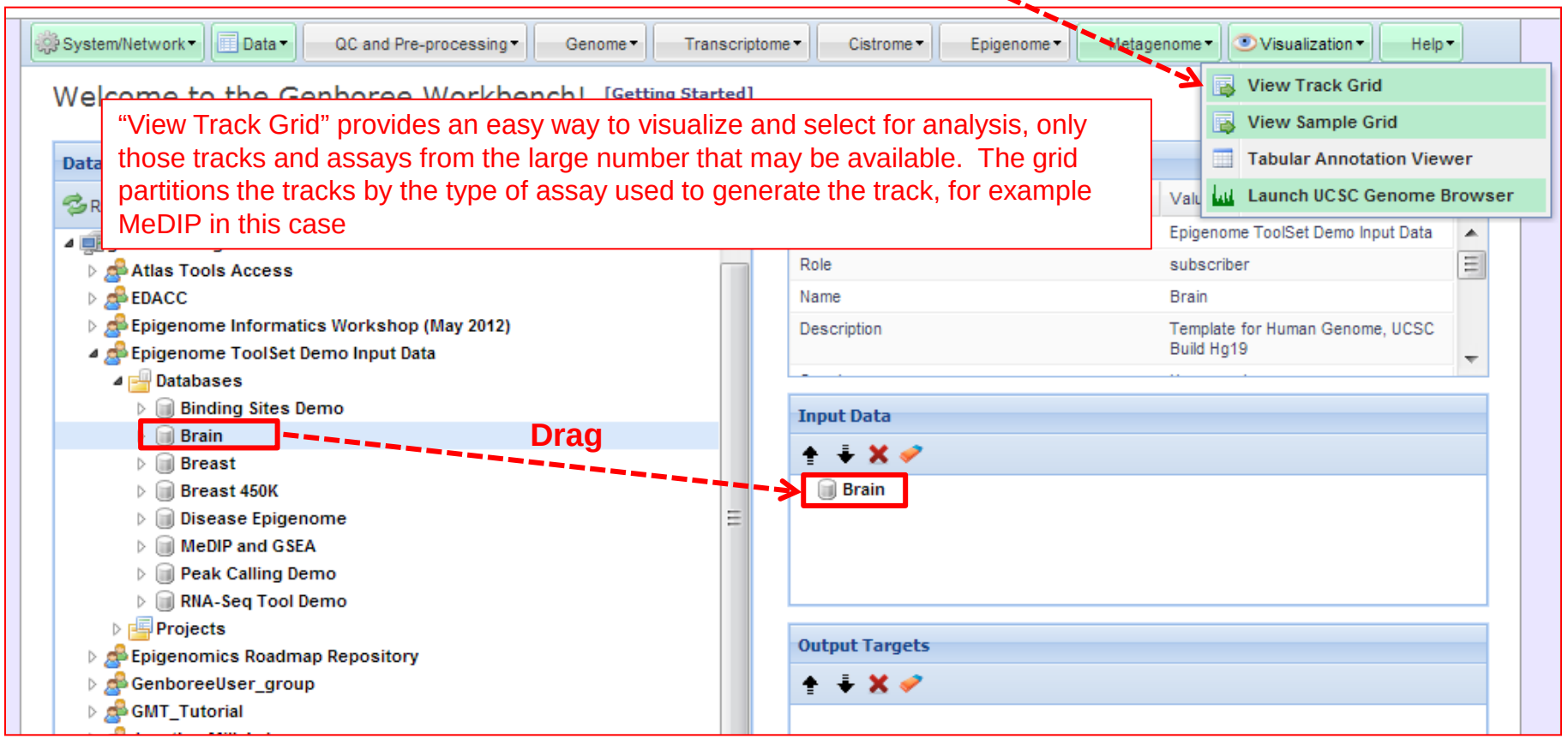
This will be similar to Use Cases 1 & 2 but with the comparison over different features (annotation tracks with different regions of interest - ROIs), and observing that some features produce better clustering in the heatmap.

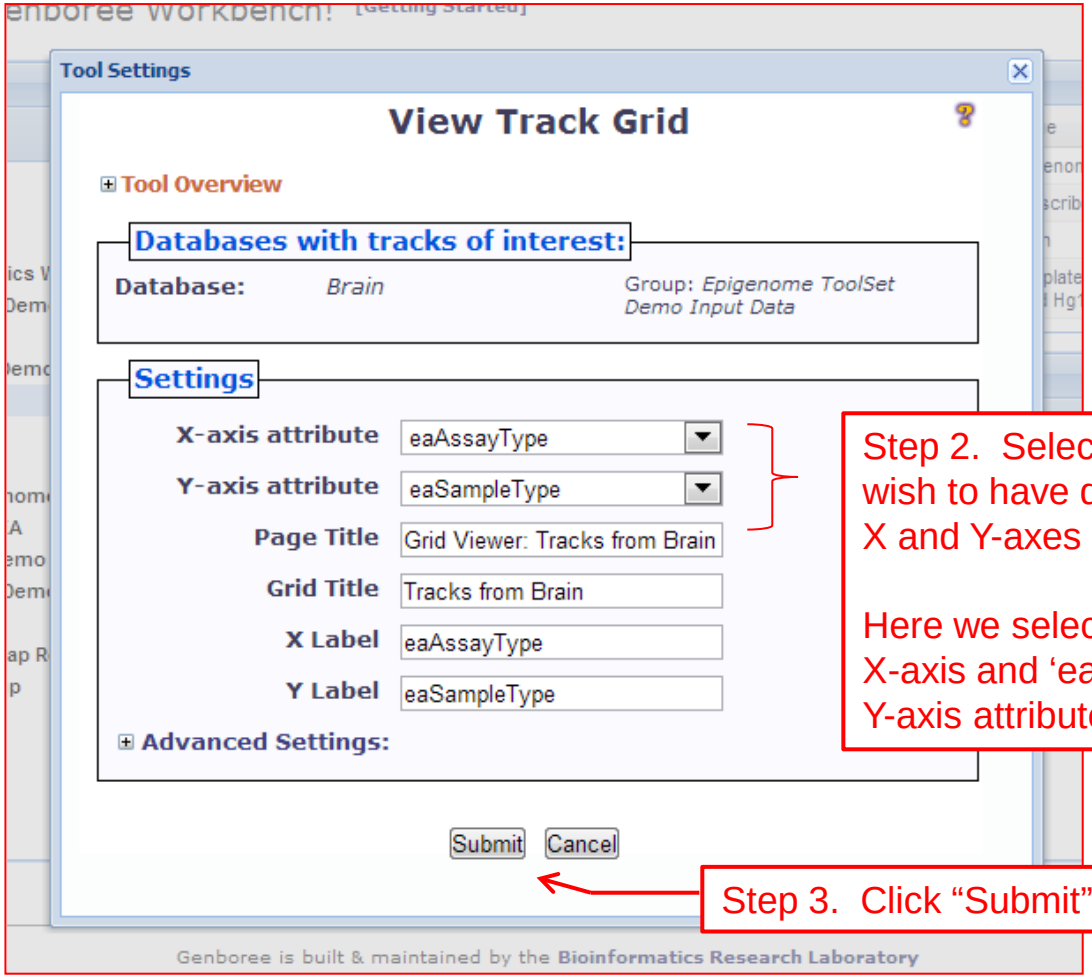
You should have already created a Project and Database in earlier use cases, to you will not need to do that again. The results of this analysis will be part of the same Project and be deposited in the same Database.

The next step is to select the samples that to analyze.

Step 1. Drag the “Brain” database into the “Input Data”.
This will cause the “Visualizaton” menu to turn green, meaning a tool(s) within that menu is active. A tool is active when “Input Data” and “Output Targets” have been populated with the appropriate data/tracks/files/databases required for that tool to operate.

-Click ‘Visualization’ and then ‘View Track Grid’

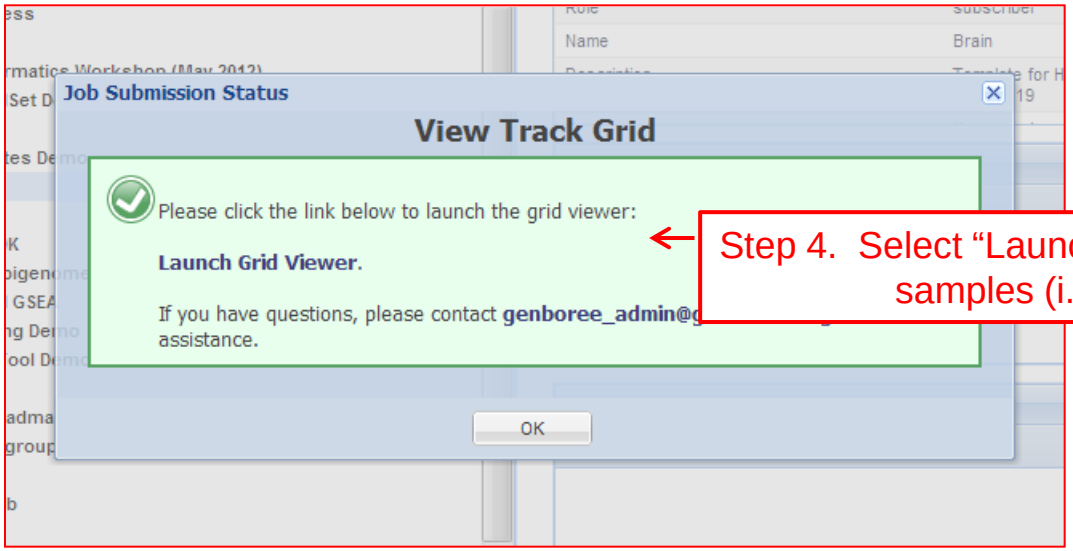


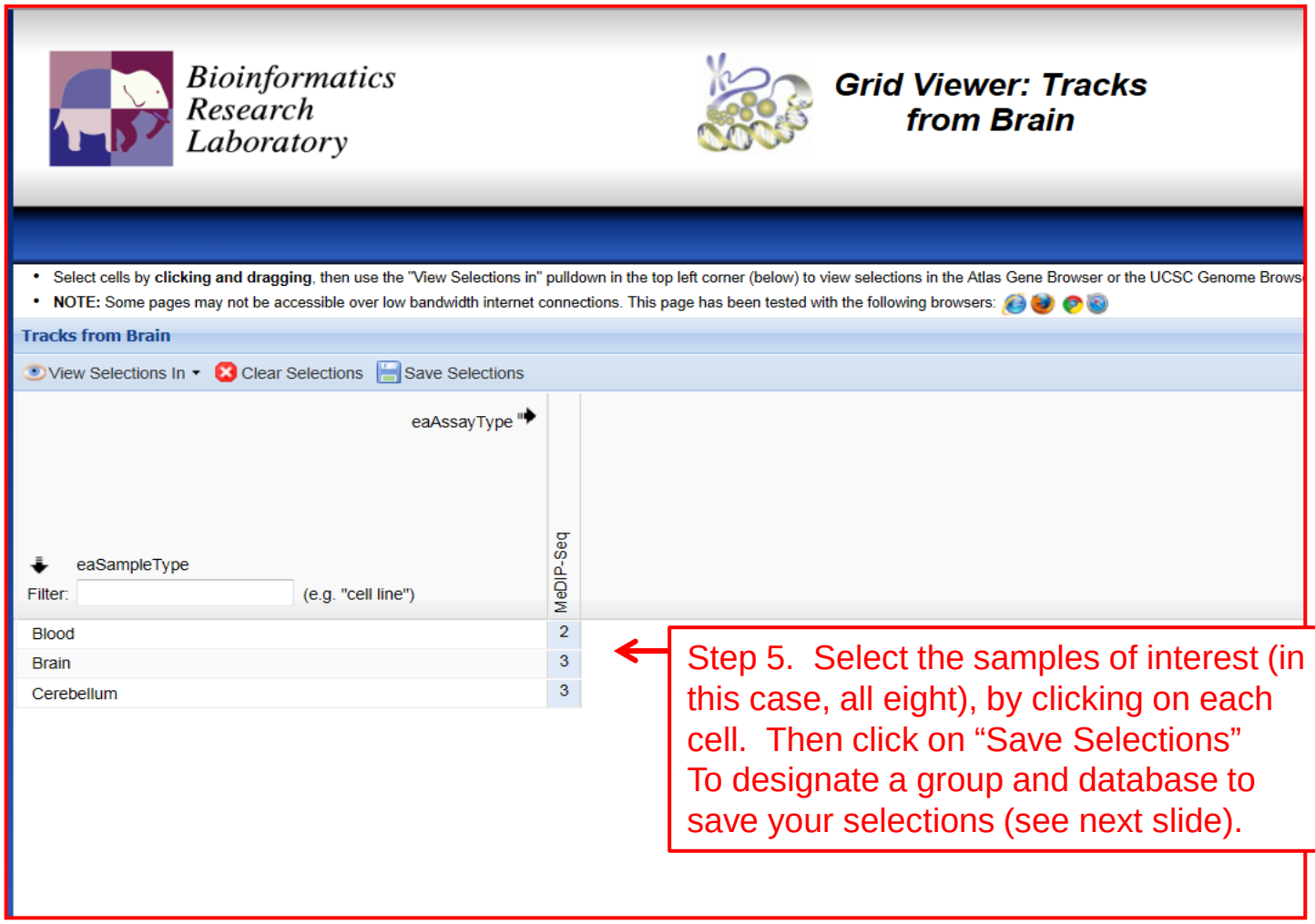


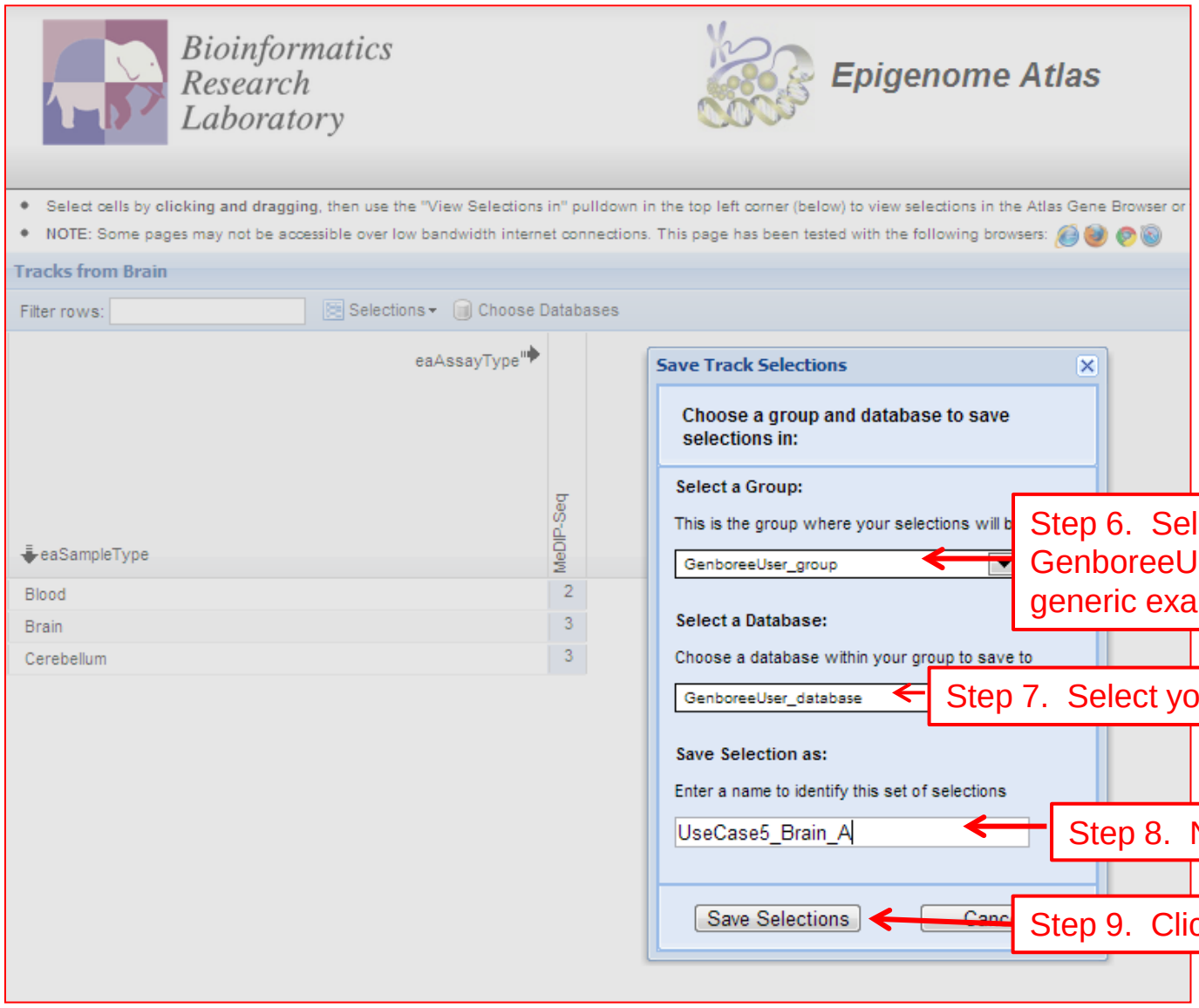
Step 2. Select which attributes you wish to have displayed on the X and Y-axes of the grid.

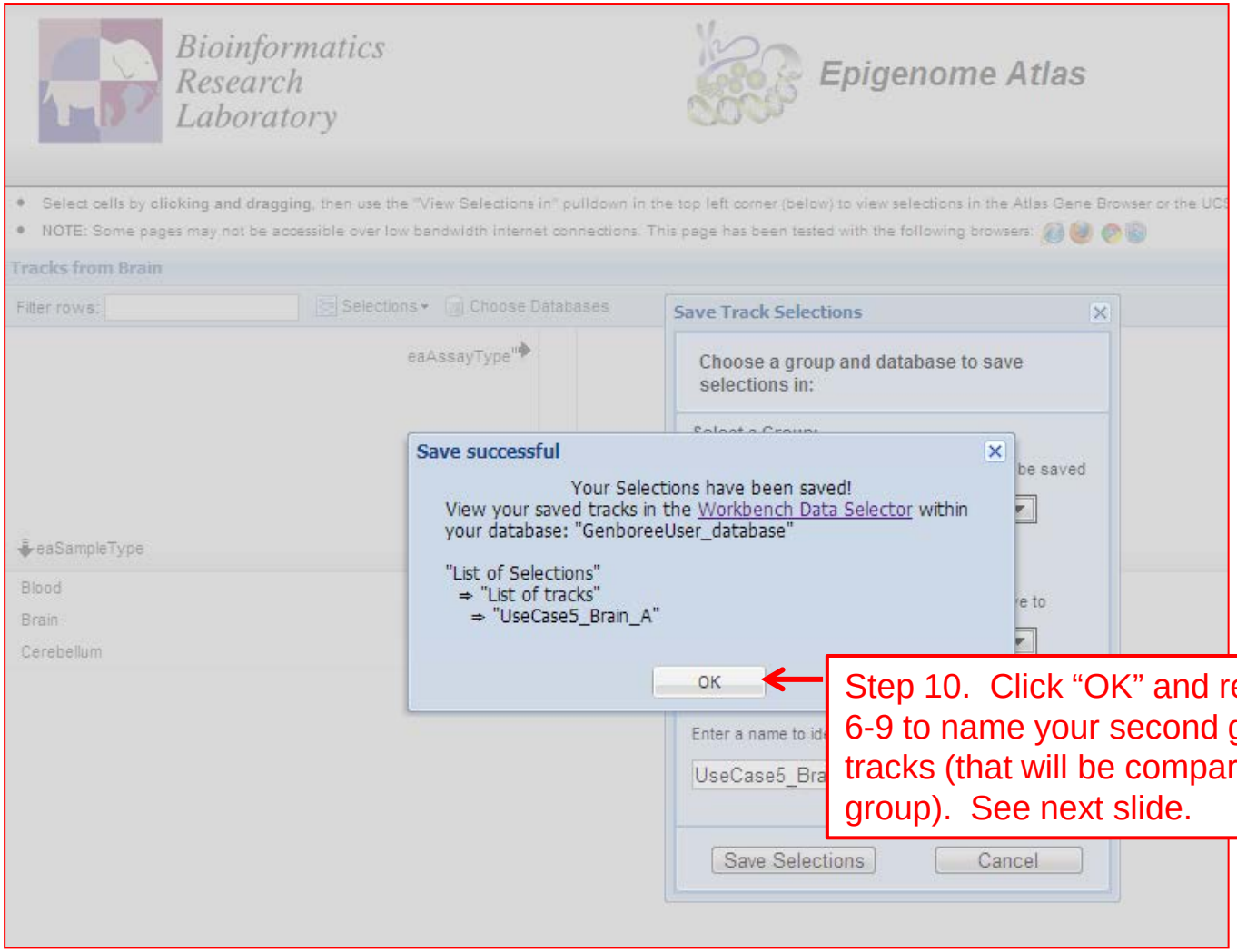
Here we select 'eaAssayType' for the X-axis and 'eaSampleType' for the Y-axis attributes.

Step 3. Click "Submit"


Job Submission Status
View Track Grid
Please click the link below to launch the grid viewer:
Launch Grid Viewer.
If you have questions, please contact genboree_admin@g
assistance.
OK
Name
Brain
Step 4. Select “Launch Grid Viewer” to select t
samples (i.e. tracks of interest)

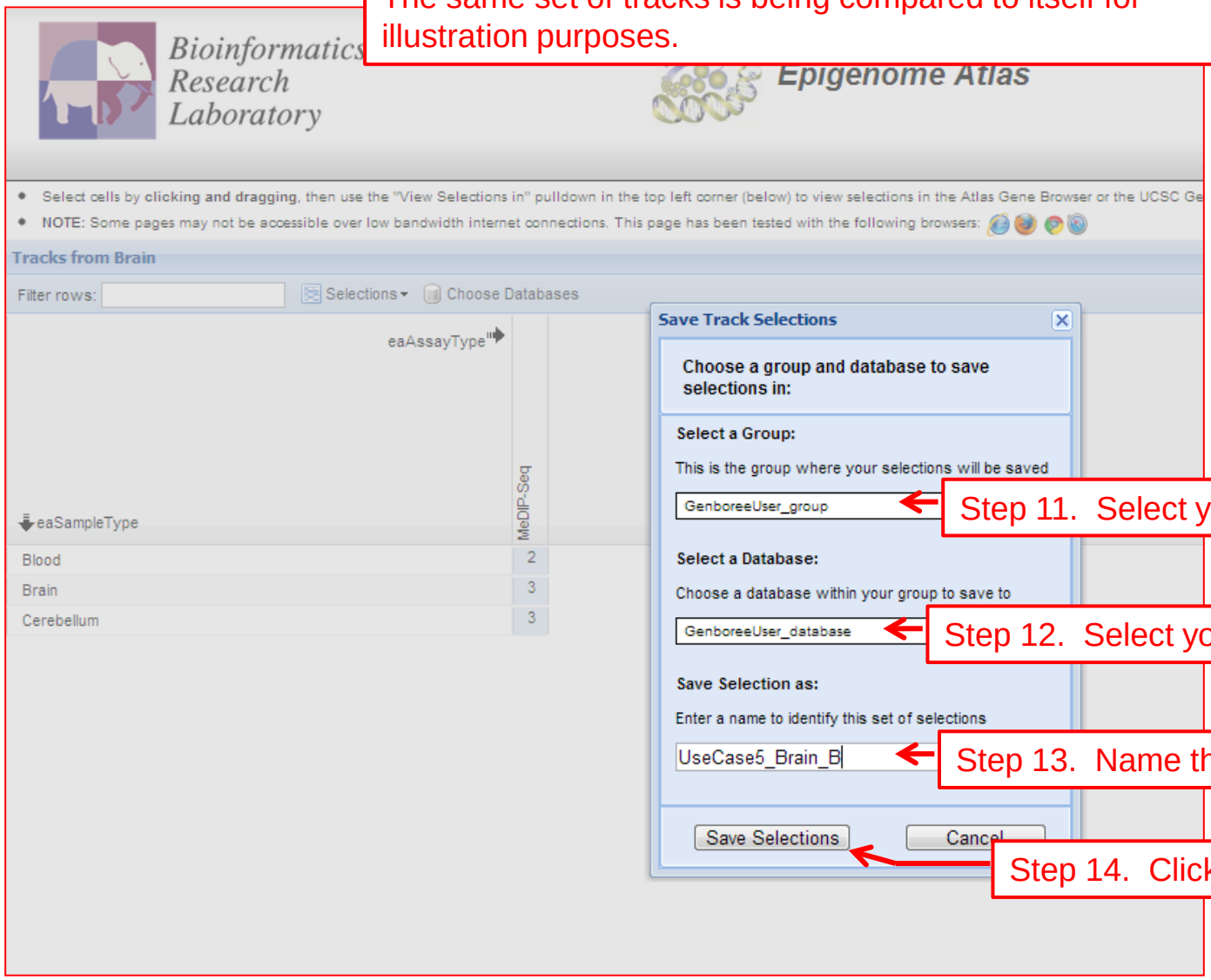

Bioinformatics Research Laboratory
Grid Viewer: Tracks from Brain
Select cells by clicking and dragging, then use the "View Selections in" pulldown in the top left corner (below) to view selections in the Atlas Gene Browser or the UCSC Genome Brows
NOTE: Some pages may not be accessible over low bandwidth internet connections. This page has been tested with the following browsers:
Tracks from Brain
View Selections In
Clear Selections
Save Selections
eaAssayType
eaSampleType
Filter:
(e.g. "cell line")
MeDIP-Seq
Blood 2
Brain 3
Cerebellum 3
Step 5. Select the samples of interest (in this case, all eight), by clicking on each cell. Then click on “Save Selections” To designate a group and database to save your selections (see next slide).


Bioinformatics Research Laboratory
Epigenome Atlas
Select cells by clicking and dragging, then use the "View Selections in" pulldown in the top left corner (below) to view selections in the Atlas Gene Browser or
NOTE: Some pages may not be accessible over low bandwidth internet connections. This page has been tested with the following browsers:
Tracks from Brain
Filter rows:
Selections
Choose Databases
eaAssayType
eaSampleType
MeDIP-Seq
Blood 2
Brain 3
Cerebellum 3
Save Track Selections
Choose a group and database to save selections in:
Select a Group:
This is the group where your selections will b
GenboreeUser_group
Select a Database:
Choose a database within your group to save to
GenboreeUser_database
Save Selection as:
Enter a name to identify this set of selections
UseCase5_Brain_A
Save Selections
Step 6. Select your user group. GenboreeUser_group is used as generic example here.
Step 7. Select your destination database
Step 8. Name this list of tracks
Step 9. Click “Save Selections”



Step 10. Click "OK" and repeat steps 6-9 to name your second group of tracks (that will be compared to the first group). See next slide.

Steps 6-9 are repeated here, but with the same set of tracks given a different name (“UseCase5_Brain_B”). The same set of tracks is being compared to itself for illustration purposes.

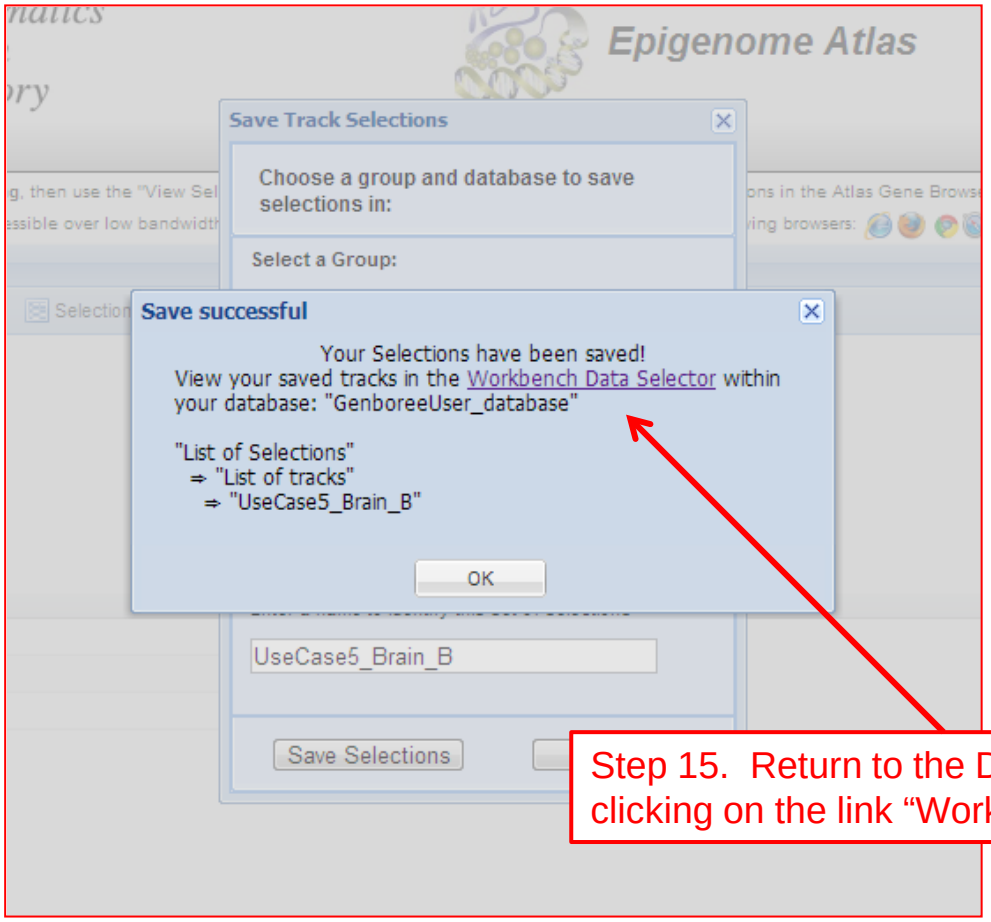


Step 11. Select your user group

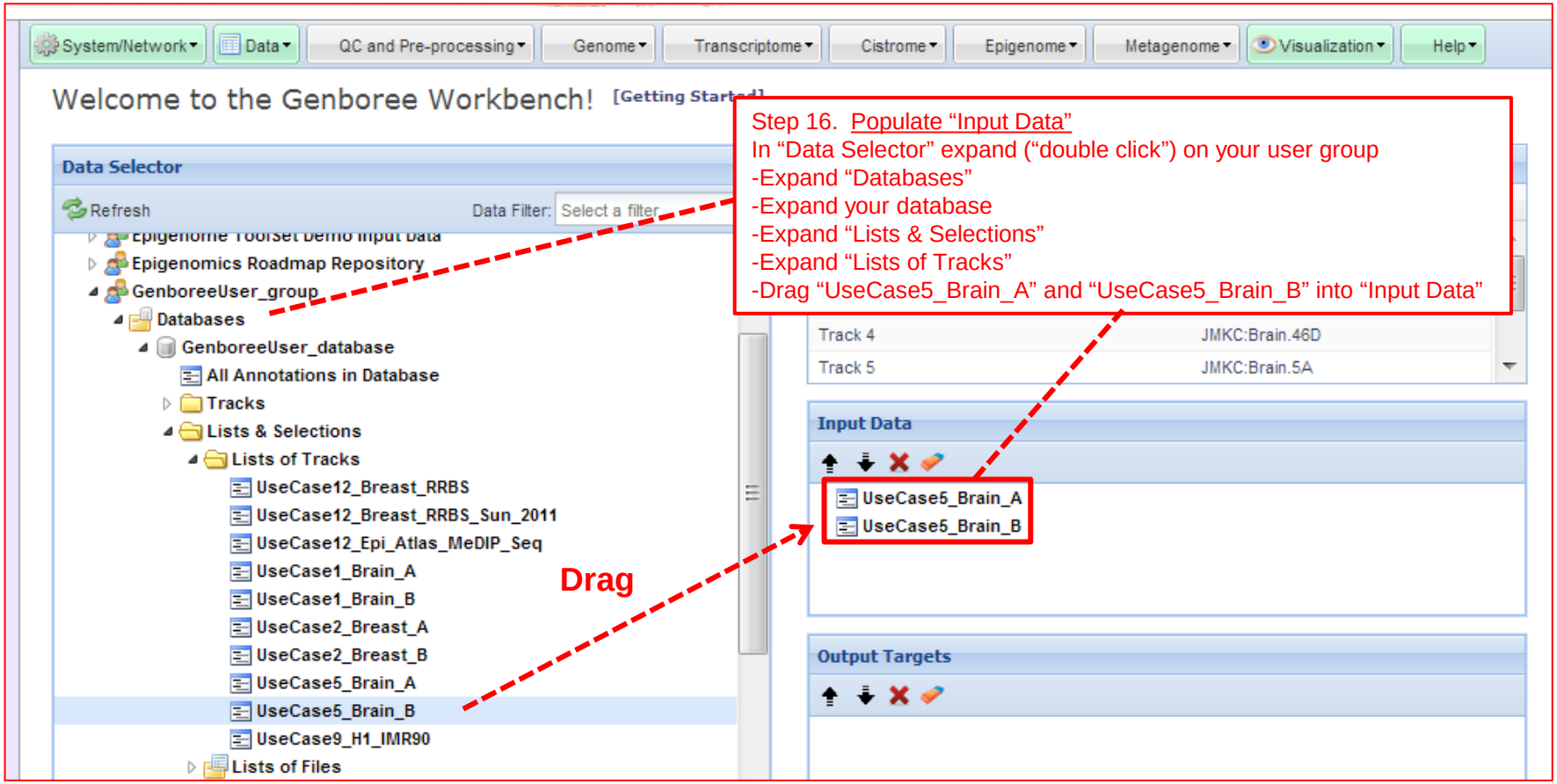
Step 12. Select your destination database

Step 13. Name this list of tracks

Step 14. Click “Save Selections”



Step 15. Return to the Data Selector by clicking on the link “Workbench Data Selector”


System/Network
Data
QC and Pre-processing
Genome
Transcriptome
Cistrome
Epigenome
Metagenome
Visualization
Help
Welcome to the Genboree Workbench!
[Getting Started]
Step 16. Populate "Input Data"
In "Data Selector" expand ("double click") on your user group
-Expand "Databases"
-Expand your database
-Expand "Lists & Selections"
-Expand "Lists of Tracks"
-Drag "UseCase5_Brain_A" and "UseCase5_Brain_B" into "Input Data"
Data Selector
Refresh
Data Filter:
Select a filter
Epigenomics Roadmap Repository
GenboreeUser_group
Databases
GenboreeUser_database
All Annotations in Database
Tracks
Lists & Selections
Lists of Tracks
UseCase12_Breast_RRBS
UseCase12_Breast_RRBS_Sun_2011
UseCase12_Epi_Atlas_MeDIP_Seq
UseCase1_Brain_A
UseCase1_Brain_B
UseCase2_Breast_A
UseCase2_Breast_B
UseCase5_Brain_A
UseCase5_Brain_B
UseCase9_H1_IMR90
Lists of Files
Drag
Track 4
JMKC:Brain.46D
Track 5
JMKC:Brain.5A
Input Data
UseCase5_Brain_A
UseCase5_Brain_B
Output Targets

Step 17. Populate “Input Data”

In “Data Selector” expand (“double click”) on your user group

-Expand “ROI Repository”

-Expand “Databases”

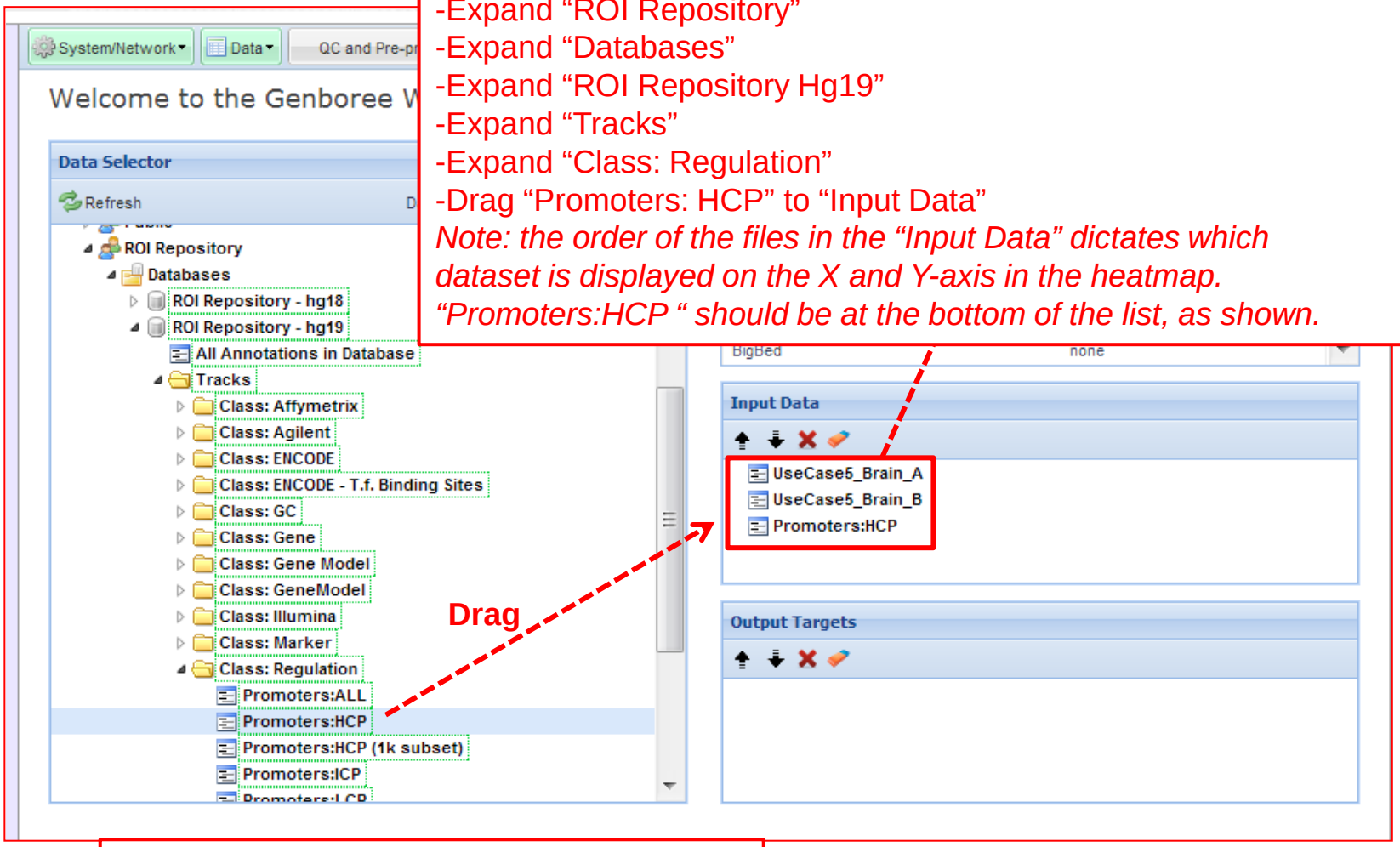
-Expand “ROI Repository Hg19”

-Expand “Tracks”

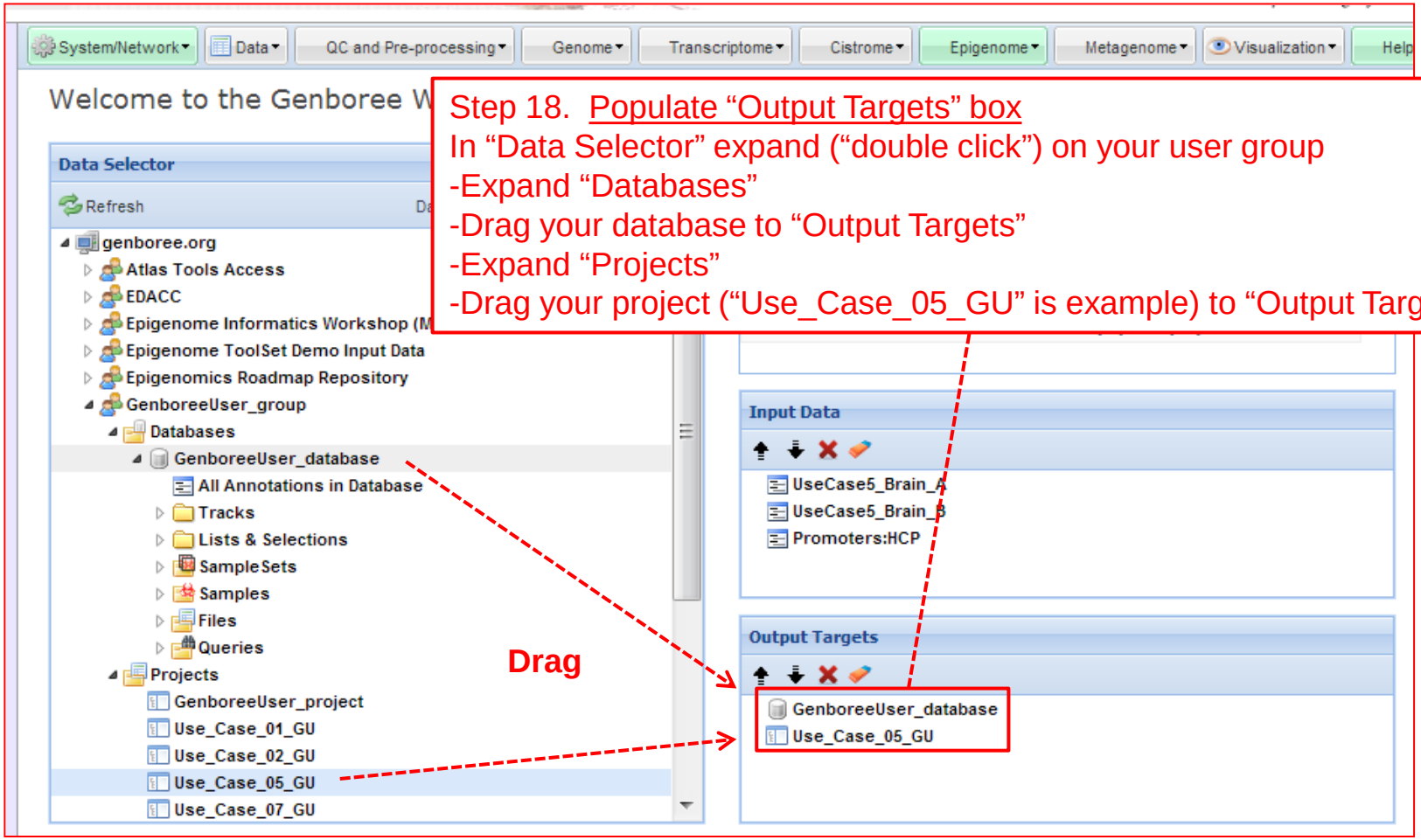
-Expand “Class: Regulation”

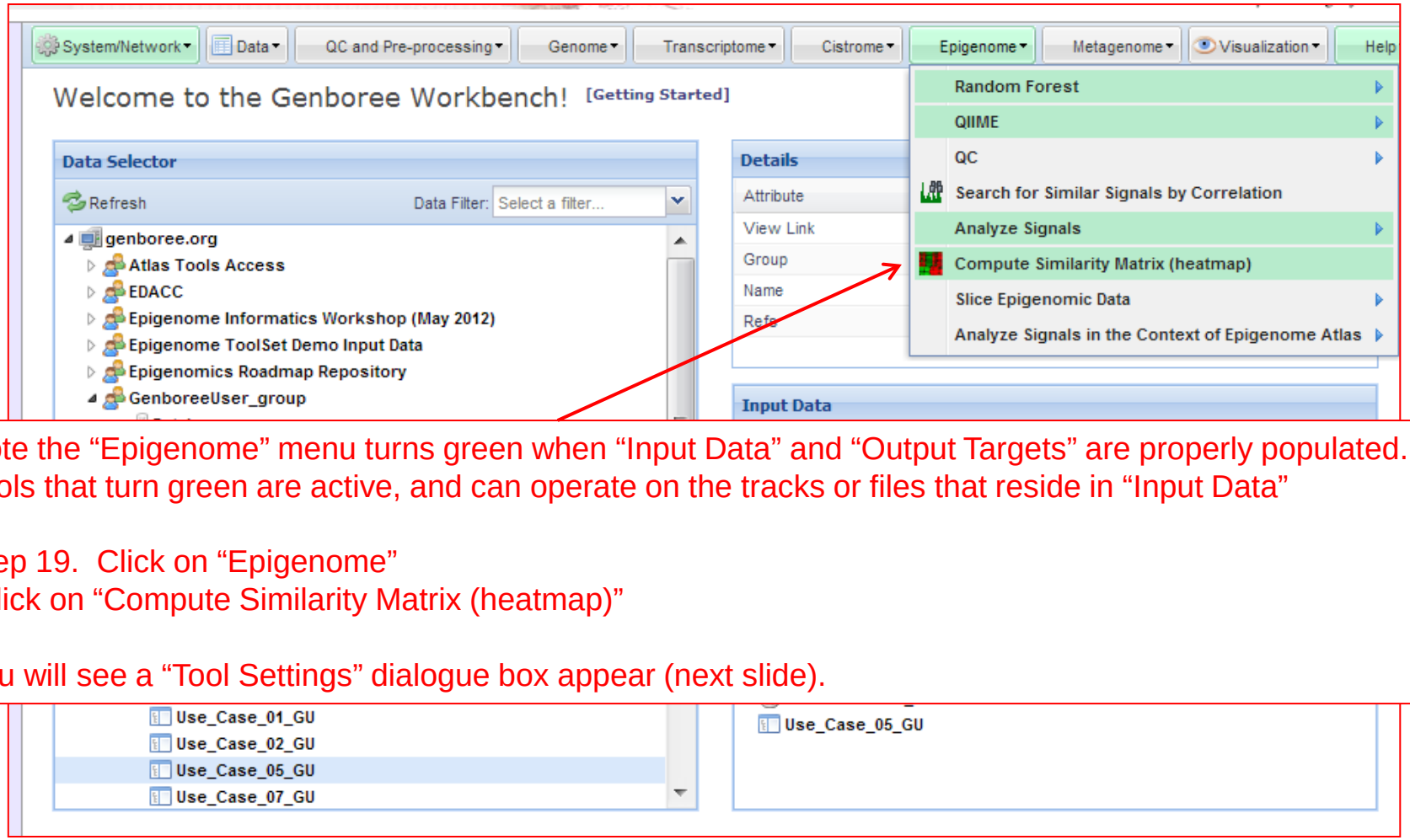
-Drag “Promoters: HCP” to “Input Data”

Note: the order of the files in the “Input Data” dictates which dataset is displayed on the X and Y-axis in the heatmap. “Promoters:HCP “ should be at the bottom of the list, as shown.



Please note: The green rectangles just indicate a certain level of access, and are not important for completing the use case (i.e. they can be ignored).

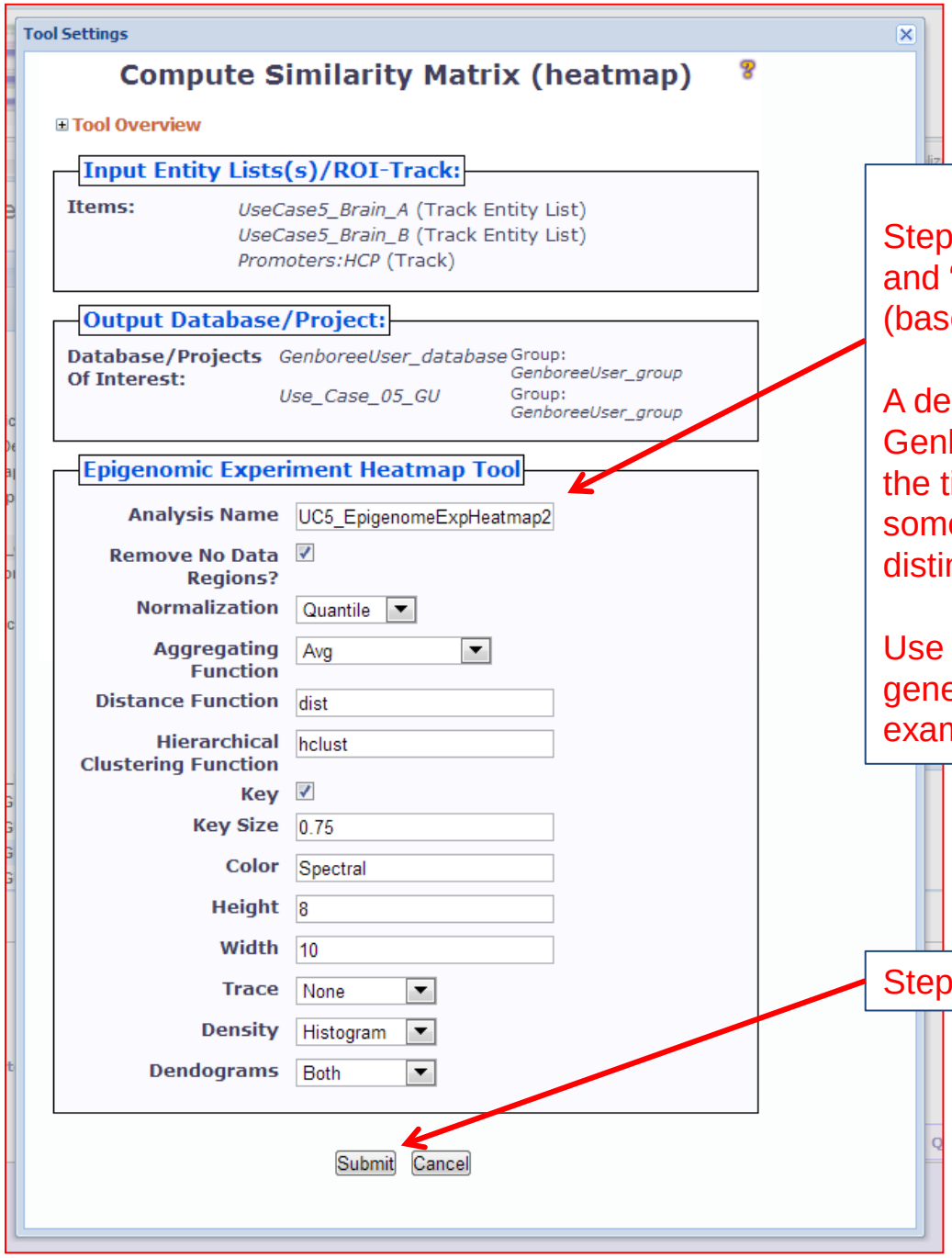

System/Network
Data
QC and Pre-processing
Genome
Transcriptome
Cistrome
Epigenome
Metagenome
Visualization
Help
Welcome to the Genboree W
Step 18. Populate "Output Targets" box
In "Data Selector" expand ("double click") on your user group
-Expand "Databases"
-Drag your database to "Output Targets"
-Expand "Projects"
-Drag your project ("Use_Case_05_GU" is example) to "Output Targets"
Data Selector
Refresh
genboree.org
Atlas Tools Access
EDACC
Epigenome Informatics Workshop (M
Epigenome ToolSet Demo Input Data
Epigenomics Roadmap Repository
GenboreeUser_group
Databases
GenboreeUser_database
All Annotations in Database
Tracks
Lists & Selections
SampleSets
Samples
Files
Queries
Projects
GenboreeUser_project
Use_Case_01_GU
Use_Case_02_GU
Use_Case_05_GU
Use_Case_07_GU
Drag
Input Data
UseCase5_Brain_A
UseCase5_Brain_B
Promoters:HCP
Output Targets
GenboreeUser_database
Use_Case_05_GU



Note the “Epigenome” menu turns green when “Input Data” and “Output Targets” are properly populated. Tools that turn green are active, and can operate on the tracks or files that reside in “Input Data”

Step 19. Click on “Epigenome”
-Click on “Compute Similarity Matrix (heatmap)”

You will see a “Tool Settings” dialogue box appear (next slide).

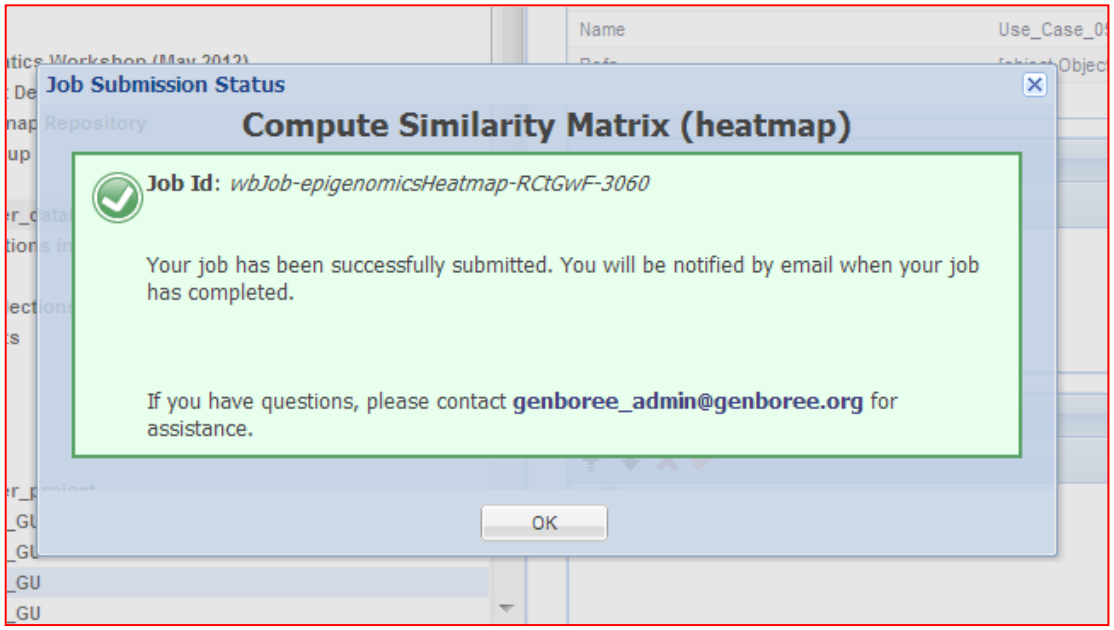


Step 20. Check that the "Input Files Directory" and "Output Database/Project" are correct (based on what you named them).

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.

Use the default parameters so that the heatmap generated will match the heatmap shown in this example.

Step 20. Click on "Submit"



You will see the message above upon successful submission of a job. Click OK. You will then receive an email alerting you to the status of your job.

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

Hello Genboree User,

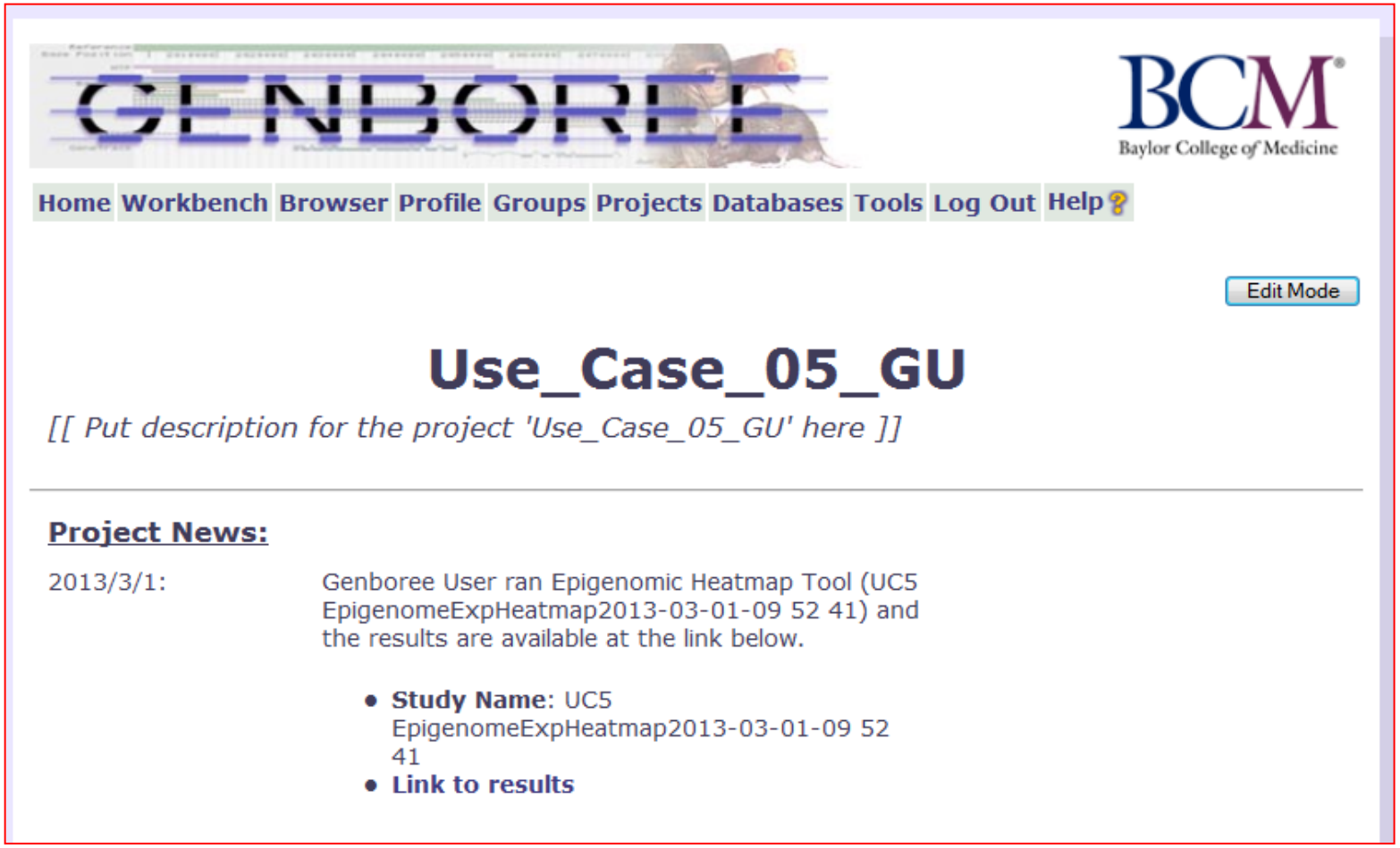
Your Compute Similarity Matrix (heatmap) job completed successfully.

Job Summary:
JobID - wbJob-epigenomicsHeatmap-RCtGwF-3060
Analysis Name - UC5_EpigenomeExpHeatmap2013-03-01-09:52:41
Inputs:
1. Entitylist - UseCase5_Brain_A
2. Entitylist - UseCase5_Brain_B
3. Trk - Promoters%3AHCP
Outputs:
1. Db - GenboreeUser_database
2. Prj - Use_Case_05_GU
Settings:
analysisName - UC5_EpigenomeExpHeatmap2013-03-01-09:52:41
color - Spectral
dendograms - both
density - histogram
distfun - dist
hclustfun - hclust
height - 8
key - TRUE
keySize - 0.75
normalization - quant
removeNoDataRegions - true
spanAggFunction - avg
trace - none
width - 10

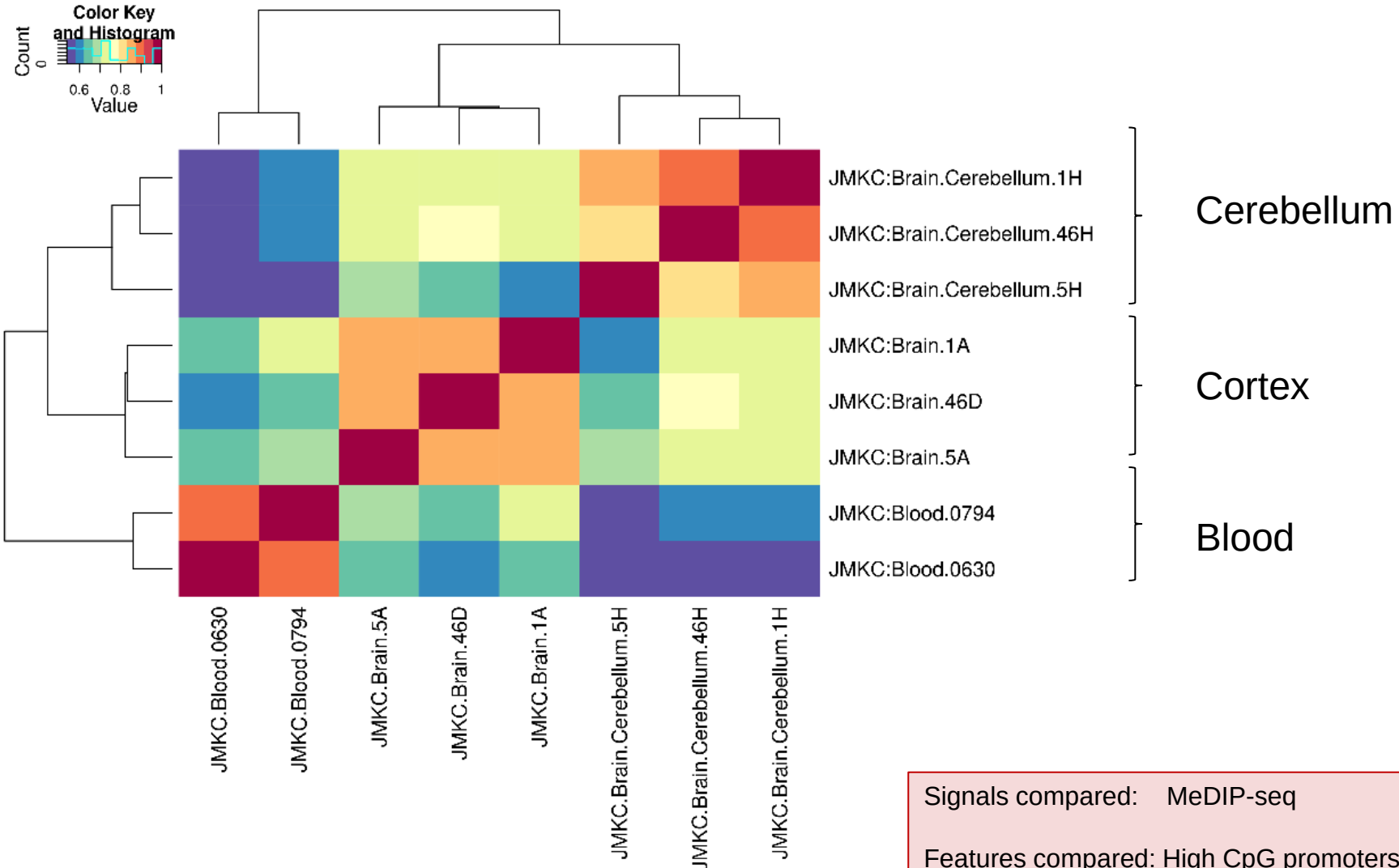
- The Genboree Team

Result File Location in the Genboree Workbench:
http://genboree.org/java-bin/project.jsp?projectName=Use_Case_05_GU

The Genboree Project Page



Genboree clusters mimic Davies et al. clustering



Signals compared: MeDIP-seq

Features compared: High CpG promoters