

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

## Epigenome Informatics Workshop Bioinformatics Research Laboratory



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<sup>1</sup>, Manuela Volta<sup>1</sup>, Abhishek Dixit<sup>1</sup>, Simon Lovestone<sup>1</sup>, Cristian Coarfa<sup>2</sup>, R. Alan Harris<sup>2</sup>, Aleksandar Milosavljevic<sup>2</sup>, Claire Troakes<sup>1</sup>, Safa Al-Sarraj<sup>1</sup>, Richard Dobson<sup>1</sup>, Leonard C. Schalkwyk<sup>1</sup>, Jonathan Mill<sup>1\*</sup> *Genome Biology*, 12:R43, 2012

<sup>1</sup>Institute of Psychiatry, King's College London. UK. <sup>2</sup>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

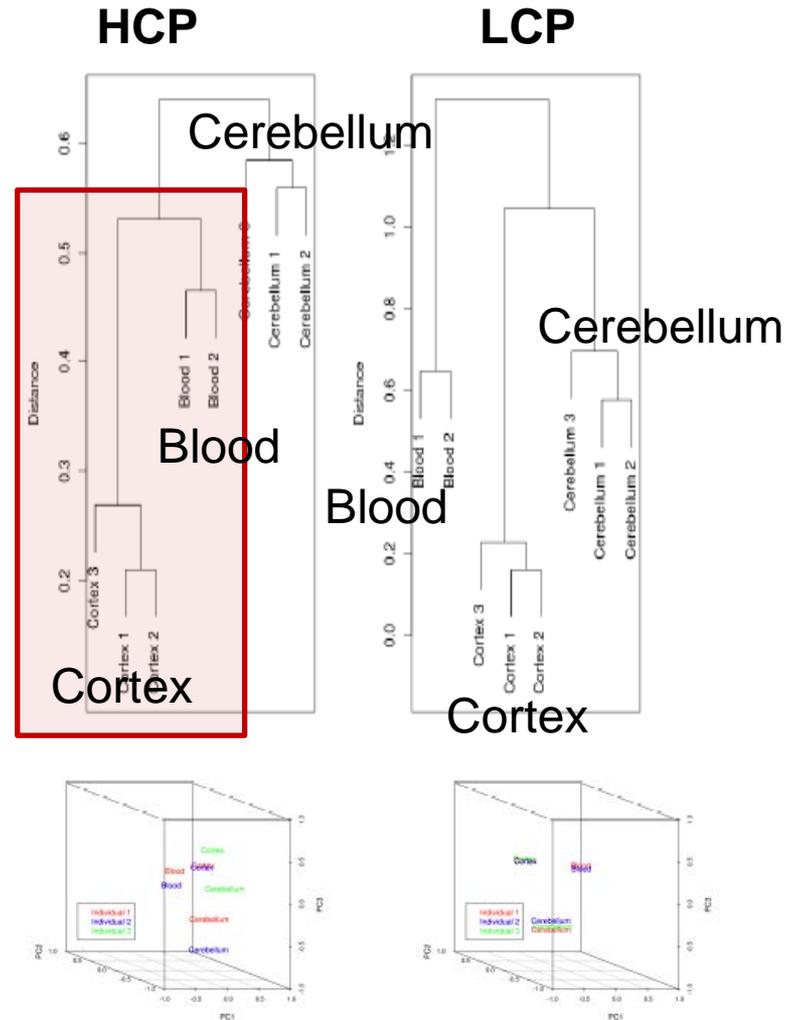
*Source of ROIs found in “Class: Regulation” in the Data Selector*

- 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%

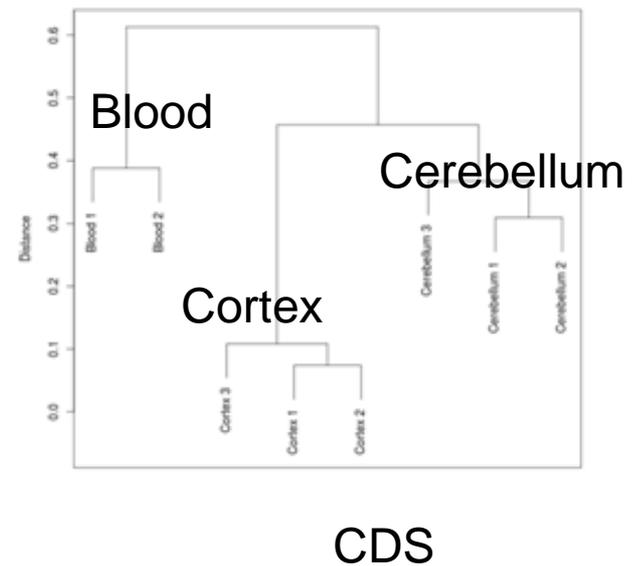
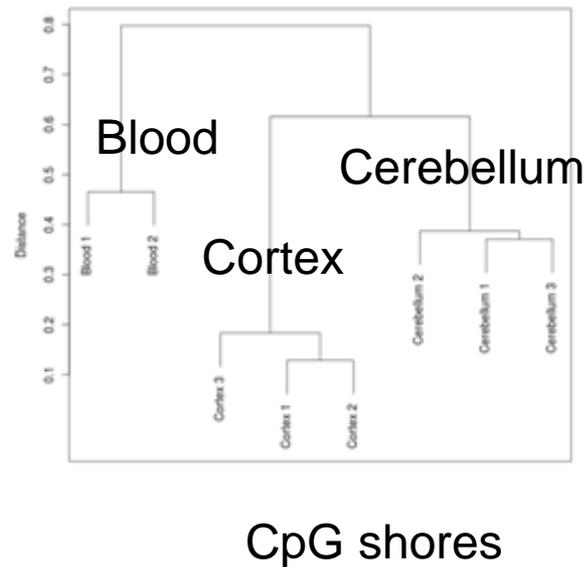
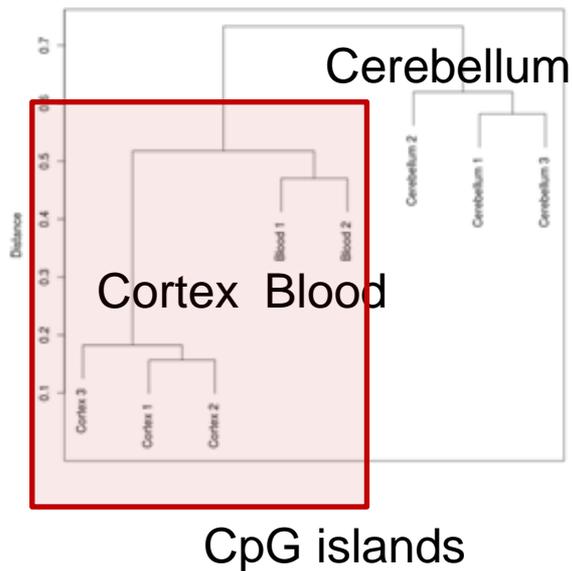
Weber et al, “Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome” *Nature Genetics*, 39 (4), April 2007

# 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)



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, so 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 “Visualization” 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’

“View Track Grid” provides an easy way to visualize and select for analysis, only those tracks and assays from the large number that may be available. The grid partitions the tracks by the type of assay used to generate the track, for example MeDIP in this case

**Drag**

Role	Value
subscriber	subscriber
Name	Brain
Description	Template for Human Genome, UCSC Build Hg19

**Input Data**

- Brain

**Output Targets**

Epigenome ToolSet Demo Input Data

- subscriber
- Brain
- Template for Human Genome, UCSC Build Hg19

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

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

Genboree Workbench! [Getting Started]

### Tool Settings

## View Track Grid

**Tool Overview**

**Databases with tracks of interest:**

**Database:** Brain Group: Epigenome ToolSet  
Demo Input Data

**Settings**

**X-axis attribute:** eaAssayType

**Y-axis attribute:** eaSampleType

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

**Grid Title:** Tracks from Brain

**X Label:** eaAssayType

**Y Label:** eaSampleType

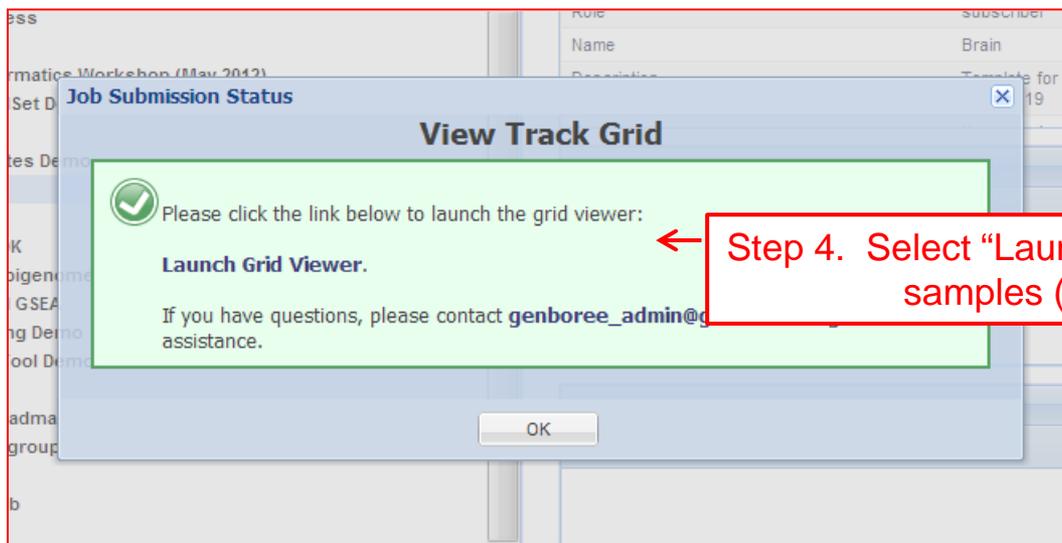
**Advanced Settings:**

Genboree is built & maintained by the **Bioinformatics Research Laboratory**

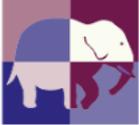
Step 2. Select which attributes you wish to have displayed on the X and Y-axes of the grid.

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

Step 3. Click "Submit"



Step 4. Select "Launch Grid Viewer" to select t  
samples (i.e. tracks of interest)



- Select cells by **clicking and dragging**, then use the "View Selections in" pull-down in the top left corner (below) to view selections in the Atlas Gene Browser or the UCSC Genome Browser
- **NOTE:** Some pages may not be accessible over low bandwidth internet connections. This page has been tested with the following browsers: 

### Tracks from Brain

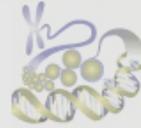
 View Selections In ▾  Clear Selections  Save Selections

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).**



- 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

MeDIP-Seq

eaSampleType

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 be saved

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"



- 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 UC
- 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

Blood

Brain

Cerebellum

#### Save Track Selections

Choose a group and database to save selections in:

Select a Group:

Select a Database:

OK

#### Save successful

Your Selections have been saved!  
View your saved tracks in the [Workbench Data Selector](#) within your database: "GenboreeUser\_database"

"List of Selections"  
⇒ "List of tracks"  
⇒ "UseCase5\_Brain\_A"

OK

Enter a name to identify this group:

UseCase5\_Bra

Save Selections

Cancel

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.

The screenshot shows the 'Epigenome Atlas' interface with a 'Save Track Selections' dialog box open. The dialog box has the following fields and buttons:

- Choose a group and database to save selections in:**
- Select a Group:** This is the group where your selections will be saved. The text box contains 'GenboreeUser\_group'.
- Select a Database:** Choose a database within your group to save to. The text box contains 'GenboreeUser\_database'.
- Save Selection as:** Enter a name to identify this set of selections. The text box contains 'UseCase5\_Brain\_B'.
- Buttons: 'Save Selections' and 'Cancel'.

Four red callout boxes with arrows point to these fields:

- Step 11. Select your user group (points to 'GenboreeUser\_group')
- Step 12. Select your destination database (points to 'GenboreeUser\_database')
- Step 13. Name this list of tracks (points to 'UseCase5\_Brain\_B')
- Step 14. Click “Save Selections” (points to the 'Save Selections' button)

The background interface shows the 'Tracks from Brain' section with a table of tracks:

eaSampleType	MeDIP-Seq
Blood	2
Brain	3
Cerebellum	3

The screenshot shows the Epigenome Atlas web interface. In the foreground, a 'Save successful' dialog box is displayed with the following text:

**Save successful**

Your Selections have been saved!  
View your saved tracks in the [Workbench Data Selector](#) within your database: "GenboreeUser\_database"

"List of Selections"  
⇒ "List of tracks"  
⇒ "UseCase5\_Brain\_B"

OK

In the background, a 'Save Track Selections' dialog box is partially visible, showing the text 'Choose a group and database to save selections in:' and 'Select a Group:'. Below it, a text input field contains 'UseCase5\_Brain\_B' and a 'Save Selections' button is visible.

A red arrow points from a text box at the bottom right to the underlined link 'Workbench Data Selector' in the 'Save successful' dialog.

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]

**Data Selector**

Refresh | Data Filter: Select a filter

- Epigenome Toolset Demo Input data
- 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

**Track 4** JMKC:Brain.46D

**Track 5** JMKC:Brain.5A

**Input Data**

- UseCase5\_Brain\_A
- UseCase5\_Brain\_B

**Output Targets**

**Drag**

**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"

### 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.*

The screenshot shows the Genboree Data Selector interface. On the left, a tree view under 'Data Selector' shows the following structure:

- Public
  - ROI Repository
    - Databases
      - ROI Repository - hg18
      - ROI Repository - hg19
        - All Annotations in Database
          - Tracks
            - Class: Affymetrix
            - Class: Agilent
            - Class: ENCODE
            - Class: ENCODE - T.f. Binding Sites
            - Class: GC
            - Class: Gene
            - Class: Gene Model
            - Class: GeneModel
            - Class: Illumina
            - Class: Marker
            - Class: Regulation
              - Promoters:ALL
              - Promoters:HCP
              - Promoters:HCP (1k subset)
              - Promoters:ICP
              - Promoters:ICP

On the right, the 'Input Data' section is visible, containing a list of datasets: UseCase5\_Brain\_A, UseCase5\_Brain\_B, and Promoters:HCP. A red dashed arrow labeled 'Drag' points from the 'Promoters:HCP' item in the 'Class: Regulation' folder to the 'Input Data' list. Another red dashed arrow points from the 'Promoters:HCP' item in the 'Input Data' list to the 'Output Targets' section.

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

**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

**Input Data**

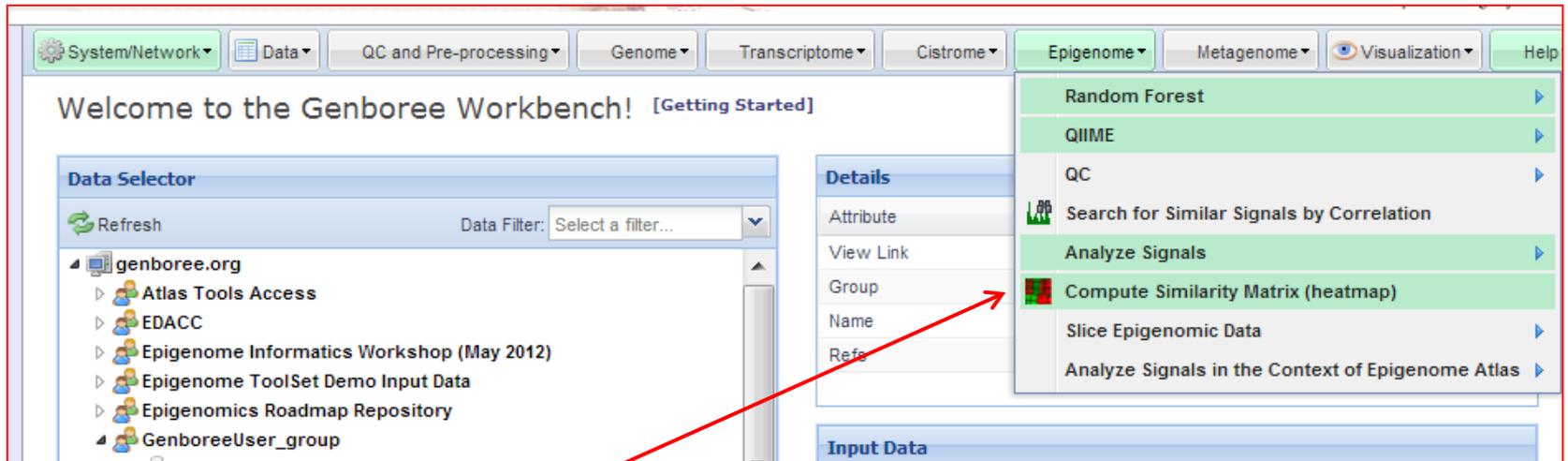
- UseCase5\_Brain\_A
- UseCase5\_Brain\_B
- Promoters:HCP

**Output Targets**

- GenboreeUser\_database
- Use\_Case\_05\_GU

**Drag**

**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"



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).



**Tool Settings**

## Compute Similarity Matrix (heatmap)

**Tool Overview**

**Input Entity Lists(s)/ROI-Track:**

Items: UseCase5\_Brain\_A (Track Entity List)  
UseCase5\_Brain\_B (Track Entity List)  
Promoters:HCP (Track)

**Output Database/Project:**

Database/Projects Of Interest: GenboreeUser\_database Group: GenboreeUser\_group  
Use\_Case\_05\_GU Group: GenboreeUser\_group

**Epigenomic Experiment Heatmap Tool**

Analysis Name: UC5\_EpigenomeExpHeatmap2

Remove No Data Regions?

Normalization: Quantile

Aggregating Function: Avg

Distance Function: dist

Hierarchical Clustering Function: hclust

Key

Key Size: 0.75

Color: Spectral

Height: 8

Width: 10

Trace: None

Density: Histogram

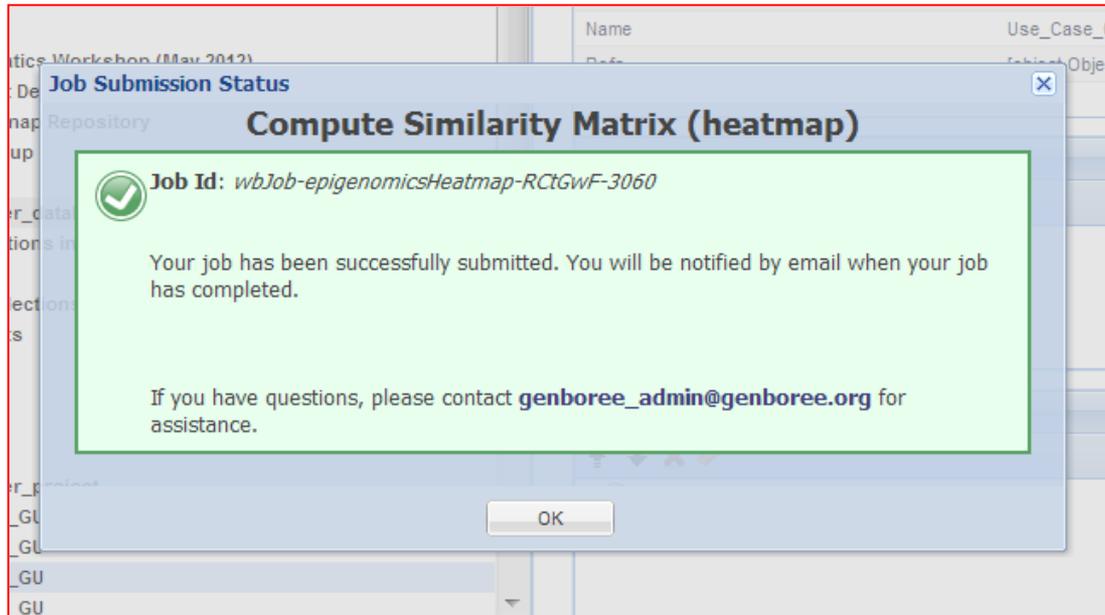
Dendograms: Both

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

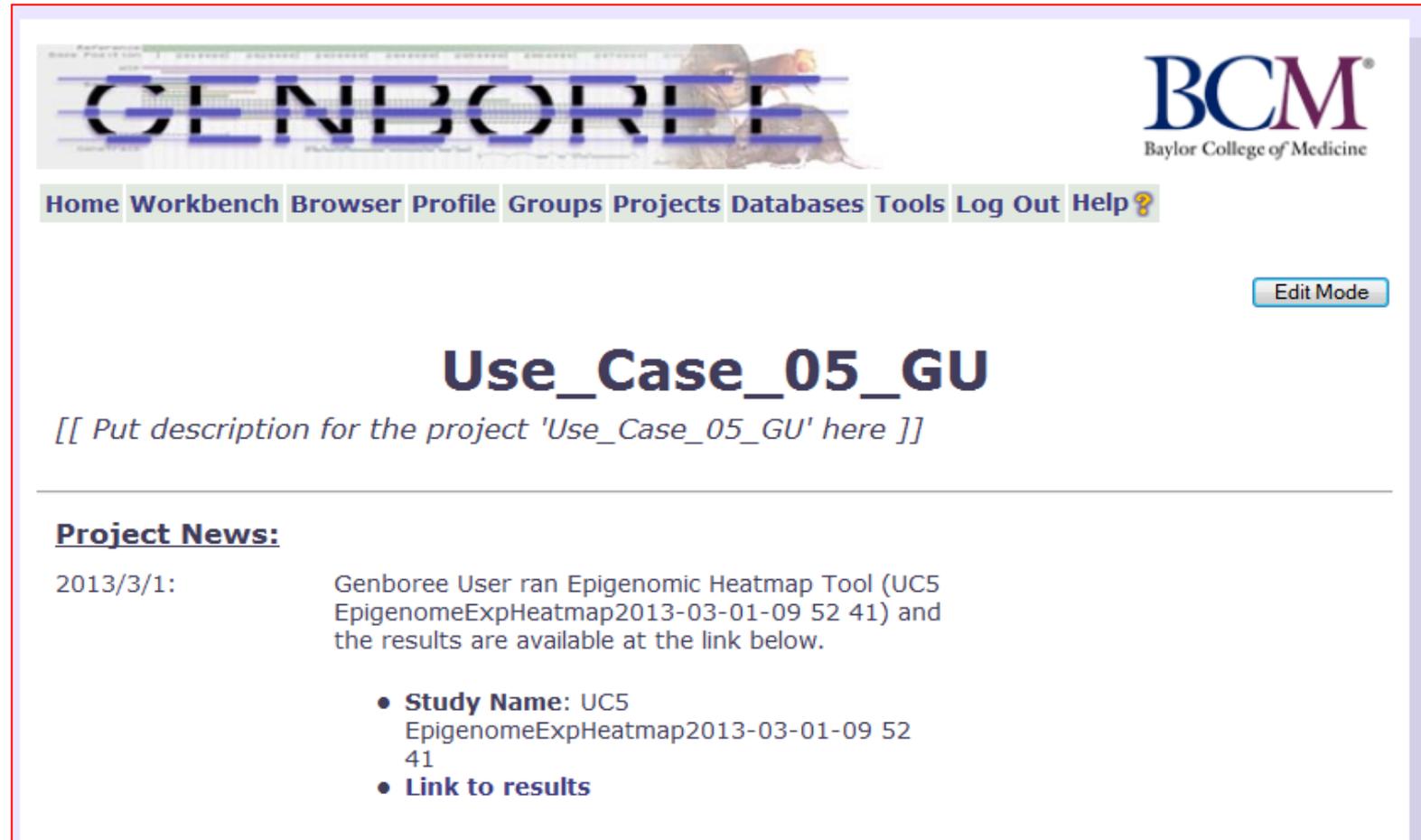
width - 10

- The Genboree Team

Result File Location in the Genboree Workbench:

[http://genboree.org/java-bin/project.jsp?projectName=Use\\_Case\\_05\\_GU](http://genboree.org/java-bin/project.jsp?projectName=Use_Case_05_GU)

# The Genboree Project Page



The screenshot shows the Genboree project page. At the top left is the Genboree logo, which includes a DNA microarray visualization and the word "GENBOREE" in large, stylized letters. To the right of the logo is the BCM (Baylor College of Medicine) logo. Below the logo is a navigation menu with buttons for Home, Workbench, Browser, Profile, Groups, Projects, Databases, Tools, Log Out, and Help. A button labeled "Edit Mode" is located in the top right corner. The main heading is "Use\_Case\_05\_GU". Below the heading is a placeholder text: "[[ Put description for the project 'Use\_Case\_05\_GU' here ]]". A horizontal line separates the heading from the "Project News:" section. The "Project News:" section contains a date "2013/3/1:" followed by a paragraph: "Genboree User ran Epigenomic Heatmap Tool (UC5 EpigenomeExpHeatmap2013-03-01-09 52 41) and the results are available at the link below." Below the paragraph is a bulleted list with two items: "Study Name: UC5 EpigenomeExpHeatmap2013-03-01-09 52 41" and "Link to results".

GENBOREE

BCM  
Baylor College of Medicine

Home Workbench Browser Profile Groups Projects Databases Tools Log Out Help ?

Edit Mode

## Use\_Case\_05\_GU

*[[ Put description for the project 'Use\_Case\_05\_GU' here ]]*

---

**Project News:**

2013/3/1: Genboree User ran Epigenomic Heatmap Tool (UC5 EpigenomeExpHeatmap2013-03-01-09 52 41) and the results are available at the link below.

- **Study Name:** UC5 EpigenomeExpHeatmap2013-03-01-09 52 41
- **Link to results**

# Genboree clusters mimic Davies et al. clustering

