

Use Case 1: Genomewide patterns of methylation can distinguish between blood, cerebellum, and cortex

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 the following reference:

"Functional annotation of the human brain methylome across brain and blood". Matthew Davies¹, Manuela Volta¹, Abhishek Dixit¹, Simon Lovestone¹, Cristian Coarfa², R. Alan Harris², Aleksandar Milosavljevic², Claire Troakes¹, Safa Al-Sarraj¹, Richard Dobson¹, Leonard C. Schalkwyk¹, Jonathan Mill^{1*} *Genome Biology*, 12:R43, 2012

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

Summary of Davies et al Manuscript

Davies et al: "*Functional annotation of the human brain methylome across brain and blood*".

Background: Dynamic changes in the epigenome play a critical role in establishing and maintaining cellular phenotype during differentiation. However, little is known about normal methylomic differences between functionally distinct areas of the brain. It was therefore of interest to examine intra- and inter-individual variation across multiple regions of the brain. The authors also sought to examine how methylomic differences in the brain correspond to methylation patterns observed in easily accessible peripheral tissues such as blood.

Results: Distinct tissue-specific patterns of DNA methylation were identified, with significant tissue-specific differentially methylated regions observed.

Promoter DNA Methylation in the Human Genome

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

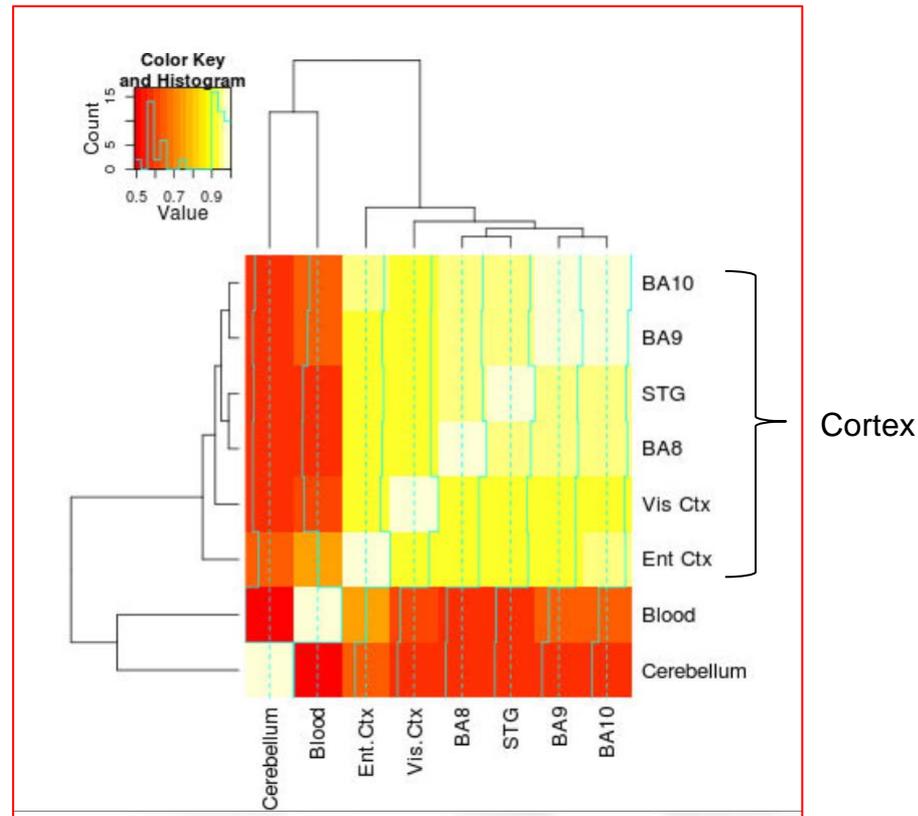
- Enriched methylated DNA from human primary fibroblasts using methylated DNA immunoprecipitation (MeDIP) + microarray detection
- 15,609 promoters evaluated in primary somatic and germline cells
- **HCPs** (high-CpG promoters) – contain 500 bp region with CpG ratio above 0.75 and GC content >55%
- **LCP** (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

Use Case 1: Genomewide patterns of methylation can distinguish between blood, cerebellum, and cortex

Canonical genic DNA methylation profiles do not differ across tissue types (i.e. similar at transcription start site, across the gene body, and 3' end). Davies et al asked if there was more differential methylation genomewide, and if so, could it be used to classify tissue types?

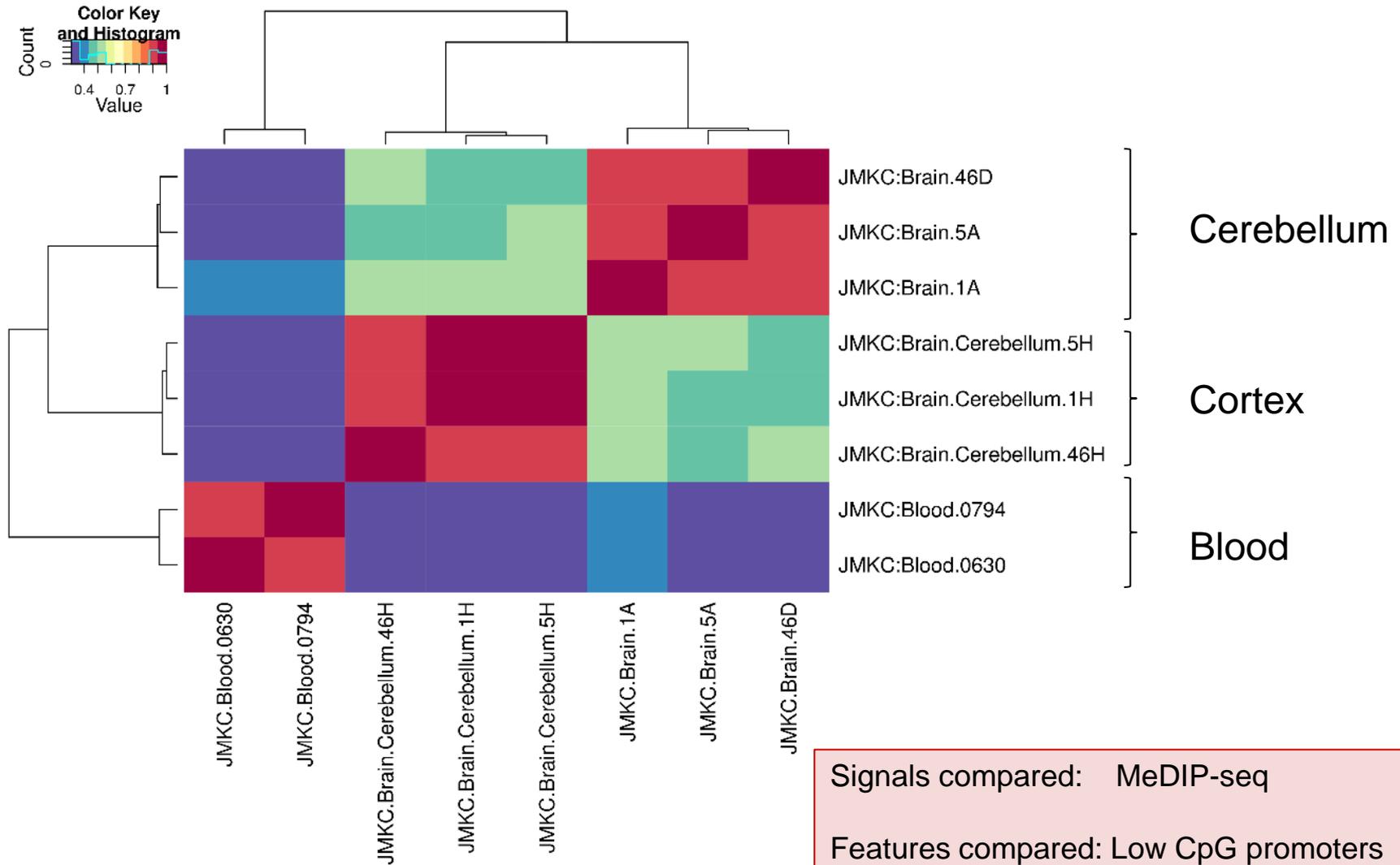
MEDIPS-processed
signal averaged over
500bp windows
genome-wide



Signals compared: MeDIP-seq

Features compared: Low CpG promoters

Genboree clustering mimics Davies et al. clustering



Prior to starting Use Case 1, you should have created a project and database in your Group.

Instructions for creating a Genboree group, project and database are contained in the Genboree Workbench.

<u>Instructions for:</u>	<u>Menu:</u>
Creating a Group:	System/Network → Group
Creating a Database:	Data → Databases
Creating a Project:	Data → Project

Instructions are also available in the Genboree Commons FAQ.

“GenboreeUser_group” is used as a generic placeholder name for any Genboree user group name: GenboreeUser is **you**.

Similarly, “GenboreeUser_database” is used as placeholder name for your database name. Therefore, as you go through the use cases, any place you see “GenboreeUser_group” or “GenboreeUser_database”, you will actually be interacting with your group and database.

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’

Welcome to the Genboree Workbench! [Getting Started]

Data Selector

Refresh

- www.genboree.org
 - Atlas Tools Access
 - EDACC
 - Epigenome Informatics Workshop (May 2012)
 - Epigenome ToolSet Demo Input Data
 - Databases
 - Brain**
 - Breast
 - Breast 450K
 - Disease Epigenome
 - MeDIP and GSEA
 - Peak Calling Demo
 - RNA-Seq Tool Demo
 - Projects
 - Epigenomics Roadmap Repository
 - GenboreeUser_group
 - GMT_Tutorial
 - JonathanMill_Lab
 - paithank_group
 - Public
 - ROI Repository

Input Data

Brain

Output Targets

The “Grid Viewer” 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 (MeDIP here)

Drag

Group	Epigenome ToolSet Demo Input Data
Role	subscriber
Name	Brain
Description	Template for Human Genome, UCSC Build Hg19
Species	Homo sapiens

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 Brair

Grid Title Tracks from Brain

X Label eaAssayType

Y Label eaSampleType

Advanced Settings:

Submit Cancel

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"

to the Genboree Workbench! [Getting Started]

Genboree.org
Tools Access
C
Genome Informatics Work
Genome ToolSet Demo
Databases
Brain
Breast
Breast 450K
Disease Epigenome
MeDIP and GSEA
Peak Calling Demo
RNA-Seq Tool Demo
Objects
Genomics Roadmap Repository
GenboreeUser_group
Tutorial
IlluminaMill_Lab

QC and Pre-processing ▾ Genome ▾ Transcriptome ▾ Cistrome ▾ Epigenome ▾ Metagenome ▾ Visualization ▾

Job Submission Status

View Track Grid

 Please click the link below to launch the grid viewer:

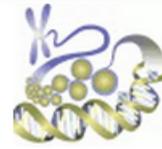
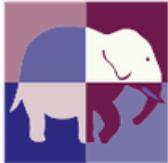
Launch Grid Viewer.

If you have questions, please contact genboree_admin@genboree.org for assistance.

OK

Output Targets

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



- 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
- **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

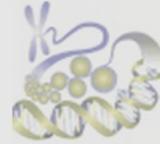
eaAssayType 	
eaAssayType = Epigenome Atlas, Assay Type	
eaSampleType = Epigenome Atlas, Sample Type	
eaSampleType 	
Blood	2
Brain	3
Cerebellum	3

MeDIP-Seq



Step 5. Select the samples of interest (in this case, all eight), by clicking on each cell. Selected cells will turn grey.

Then click on "Save Selections" (see next slide).



- Select cells by **clicking and dragging**, then use the "View Selections in" pulldown menu
- **NOTE:** Some pages may not be accessible over low bandwidth internet connections

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

Step 6. Select your user group ("GenboreeUser" is used here for illustration purposes)

Select a Database:

Choose a database within your group to save to

GenboreeUser_database

Step 7. Select the database that you created earlier

Save Selection as:

Enter a name to identify this set of selections

UseCase1_Brain_A

Step 8. Name this list of tracks

Save Selections

Cancel

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 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

eaAssayType

eaSampleType

Filter: (e.g. "cell line")

Blood

Brain

Cerebellum

Save Track Selections

Choose a group and database to save selections in:

Select a Group:

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"
⇒ "UseCase1_Brain_A"

Enter a name to identify this set of 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 6-10 are repeated here, but with the same set of tracks are given a different name ("UseCase1_B"). The same set of tracks is being used here to compute a similarity matrix for illustration purposes.

Subsequent use cases will perform more meaningful biological comparisons.

The screenshot shows the Bioinformatics Research Laboratory web application interface. A 'Save Track Selections' dialog box is open, overlaid on the main application. The dialog box contains the following sections and fields:

- Choose a group and database to save selections in:**
- Select a Group:** This is the group where your selections will be saved. The text input field contains 'GenboreeUser_group'. A red callout box points to this field with the text: "Step 11. Select your user group".
- Select a Database:** Choose a database within your group to save to. The text input field contains 'GenboreeUser_database'. A red callout box points to this field with the text: "Step 12. Select the database that you created earlier".
- Save Selection as:** Enter a name to identify this set of selections. The text input field contains 'UseCase1_Brain_B'. A red callout box points to this field with the text: "Step 13. Name this list of tracks".
- At the bottom of the dialog box are two buttons: 'Save Selections' and 'Cancel'. A red callout box points to the 'Save Selections' button with the text: "Step 14. Click 'Save Selections'".

The background application shows the 'Tracks from Brain' section with a table of tracks. The table has columns for 'eaSampleType' and 'MeDIP-Seq'. The rows are:

eaSampleType	MeDIP-Seq
Blood	2
Brain	3
Cerebellum	3

- 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 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")

Blood

Brain

Cerebellum

Save Track Selections

Choose a group and database to save selections in:

Select a Group:

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"
⇒ "UseCase1_Brain_B"

OK

Enter a name to identify this set of selections

UseCase1_Brain_B

Save Selections

Cancel

Step 15. Click "OK". You may 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...

- www.genboree.org
 - Atlas Tools Access
 - EDACC
 - Epigenome Informatics Workshop (May 2012)
 - Epigenome ToolSet Demo Input Data
 - Epigenomics Roadmap Repository
 - GenboreeUser_group**
 - Databases
 - GenboreeUser_database
 - All Annotations in Database
 - Tracks
 - Lists & Selections
 - Lists of Tracks
 - UseCase1_Brain_A
 - UseCase1_Brain_B
 - Lists of Files
 - Sample Sets
 - Samples
 - Files
 - Queries
 - Projects
 - GMT_Tutorial

Step 16. Populate "Input Data"
 In "Data Selector" expand ("double click") on your user group
 -Expand "Databases"
 -Expand you user group (i.e. "GenboreeUser_database")
 -Expand "Lists & Selections"
 -Expand "Lists of Tracks"
 -Drag "UseCase1_Brain_A" and "UseCase1_Brain_B" into "Input Data"

Track 5	JMKC:Brain.5A
Track 6	JMKC:Brain.Cerebellum.1H

Input Data

↑ ↓ × ↻

- UseCase1_Brain_A
- UseCase1_Brain_B

Output Targets

↑ ↓ × ↻

Welcome to the Genboree Workbench! [Getting Started]

System/Network | Data | QC and Pre-processing | Genome | Help

Data Selector

Refresh | Data Filter: Select a filter...

- Public
 - ROI Repository
 - Databases
 - ROI Repository - hg18
 - ROI Repository - hg19**
 - All Annotations in Database
 - Tracks
 - Class: Affymetrix
 - Class: Agilent
 - Class: ENCODE
 - 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:LCP
 - Class: Sequence

Table:

Name	Promoters:LCP
Description	
BigBed	none
Display	----

Input Data

- UseCase1_Brain_A
- UseCase1_Brain_B
- Promoters:LCP**

Output Targets

Drag

Step 17. Populate "Input Data"

- Expand "ROI Repository" group
- Expand "Databases"
- Expand "ROI Repository Hg19"
- Expand "Tracks"
- Expand "Class: Regulation"
- Drag "Promoters: LCP" to "Input Data"

Note: the order of the files in the "Input Data" dictates which dataset is displayed on the X and Y-axis. The "Promoters:LCP file should be at the bottom of the list, as shown.

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

Step 18. Populate "Output Targets"

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

-Expand "Databases"

-Drag your database (i.e. "GenboreeUser_database") to "Output Targets"

-Expand "Projects"

-Drag your project ("Use_Case_01_GU" is example) to "Output Targets"

The screenshot shows the Genboree Workbench interface. The top navigation bar includes "System/Network", "Data", "QC and Pre-processing", and "Genome". The main area is titled "Welcome to the Genboree Workbench!".

The **Data Selector** panel on the left shows a tree view of the Genboree.org structure. The "GenboreeUser_group" is expanded, showing "Databases" and "Projects". The "GenboreeUser_database" is selected and highlighted in blue. A red dashed arrow points from it to the "Output Targets" panel. Another red dashed arrow points from the "Use_Case_01_GU" project to the "Output Targets" panel. The word "Drag" is written in red between these two arrows.

The **Details** panel on the right shows the properties of the selected "GenboreeUser_database":

Attribute	Value
Group	GenboreeUser_group
Role	administrator
Name	GenboreeUser_database
Description	Template for Human Genome, UCSC Build Hg19
Species	Homo sapiens

The **Input Data** panel shows a list of data items: "UseCase1_Brain_A", "UseCase1_Brain_B", and "Promoters:LCP".

The **Output Targets** panel shows a list of output targets: "GenboreeUser_database" and "Use_Case_01_GU". These two items are enclosed in a red box.

Note the “Epigenome” menu turns green when “Input Data” and “Output Targets” are properly populated.

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

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

The screenshot displays a software interface with several key components:

- Data Selector:** A tree view on the left showing a hierarchy of data sources. The selected path is: `www.genboree.org` > `GenboreeUser_group` > `Databases` > `GenboreeUser_database`. Other visible items include `Atlas Tools Access`, `EDACC`, `Epigenome Informatics Workshop (May 2012)`, `Epigenome ToolSet Demo Input Data`, `Epigenomics Roadmap Repository`, `All Annotations in Database`, `Tracks`, `Lists & Selections`, `SampleSets`, `Samples`, `Files`, `Queries`, and `Projects` (containing `GenboreeUser_project`, `Use Case 05 GU`, `Use Case 01 GU`, `Use Case 02 GU`, and `Use Case 05 GU`).
- Epigenome Menu:** A dropdown menu at the top center is open, showing a list of tools. The menu is highlighted in green. The selected tool is `Compute Similarity Matrix (heatmap)`. Other tools include `Random Forest`, `QIIME`, `QC`, `Search for Similar Signals by Correlation`, `Analyze Signals`, `Slice Epigenomic Data`, and `Analyze Signals in the Context of Epigenome Atlas`.
- Details Panel:** A panel on the right showing metadata for the selected tool, including `Attribute`, `Group`, `Role`, `Name`, `Description`, and `Species` (set to `Homo sapiens`).
- Input Data:** A section on the right showing a list of input data items: `UseCase1_Brain_A`, `UseCase1_Brain_B`, and `Promoters:LCP`.
- Output Targets:** A section on the right showing a list of output target items: `GenboreeUser_database` and `Use Case 01 GU`.

Step 20. Check that the “Input Files Directory” and “Output Database” and “Project” are correct (based on what you named them). Use the default parameters to begin with, and experiment with changing the parameters in subsequent jobs.

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.

Input Entity Lists(s)/ROI-Track:

Items: UseCase1_Brain_A (Track Entity List)
UseCase1_Brain_B (Track Entity List)
Promoters:LCP (Track)

Output Database/Project:

Database/Projects Of Interest: GenboreeUser_database Group: GenboreeUser_group
Use_Case_01_GU Group: GenboreeUser_group

Epigenomic Experiment Heatmap Tool

Analysis Name EpigenomeExpHeatmap2013.

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

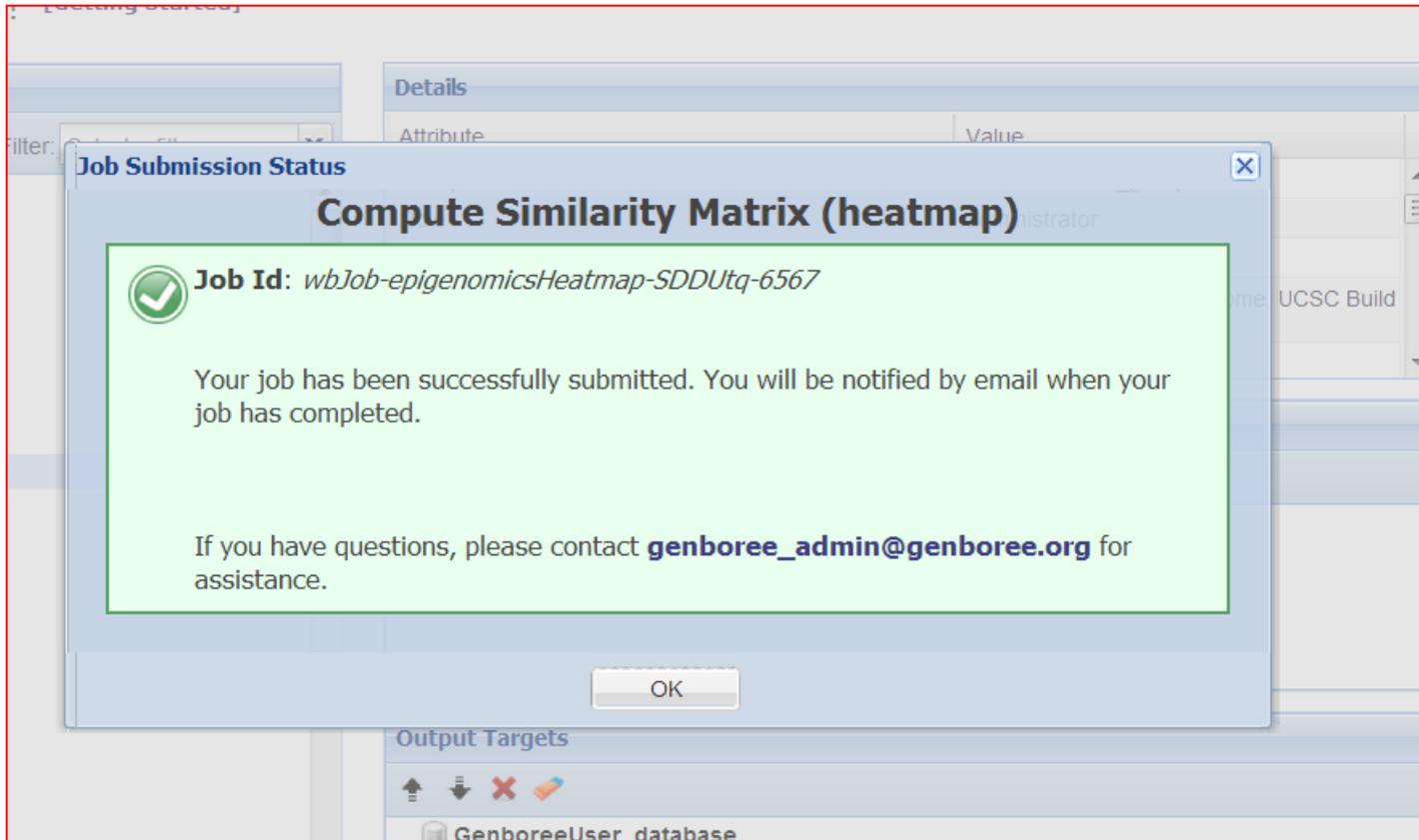
Density Histogram

Dendograms Both

Submit Cancel

Step 21. Click on “Submit”

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



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

Hello Genboree User,

Your job completed successfully.

Job Summary:

JobID - wbJob-epigenomicsHeatmap-SDDUtq-6567

Analysis Name - EpigenomeExpHeatmap2013-02-21-17:14:44

Inputs:

1. Entitylist - UseCase1_Brain_A

2. Entitylist - UseCase1_Brain_B

3. Trk - Promoters%3ALCP

Outputs:

1. Db - GenboreeUser_database

2. Prj - Use_Case_01_GU

Settings:

analysisName - EpigenomeExpHeatmap2013-02-21-17:14:44

color - Spectral

dendograms - both

density - histogram

distfun - dist

hclustfun - hclust

height - 8

key - TRUE

keySize - 0.75

normalization - quant

quantileNormalized - false

removeNoDataRegions - true

spanAggFunction - avg

trace - none

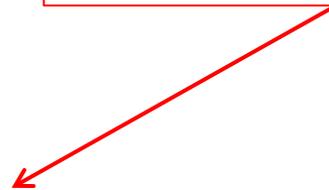
width - 10

- The Genboree Team

Result File Location in the Genboree Workbench:

http://www.genboree.org/java-bin/project.jsp?projectName=Use_Case_01_GU

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



The Genboree Project Page



The screenshot shows the Genboree project page. At the top left is the Genboree logo, which includes the word "GENBOREE" in large blue letters and a small image of a mouse. To the right of the logo is the BCM logo (Baylor College of Medicine). 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_01_GU" in large blue font. Below the heading is a placeholder text: "[[Put description for the project 'Use_Case_01_GU' here]]". A horizontal line separates the heading from the "Project News:" section. The "Project News:" section contains a date "2013/2/21:" followed by a paragraph: "Genboree User ran Epigenomic Heatmap Tool (EpigenomeExpHeatmap2013-02-21-17 14 44) and the results are available at the link below." Below this paragraph is a bulleted list with two items: "Study Name: EpigenomeExpHeatmap2013-02-21-17 14 44" and "Link to results".

Genboree clusters mimic Davies et al. clustering

