

Use Case 14: ChIP-Seq Analysis Using MACS via the Genboree Workbench

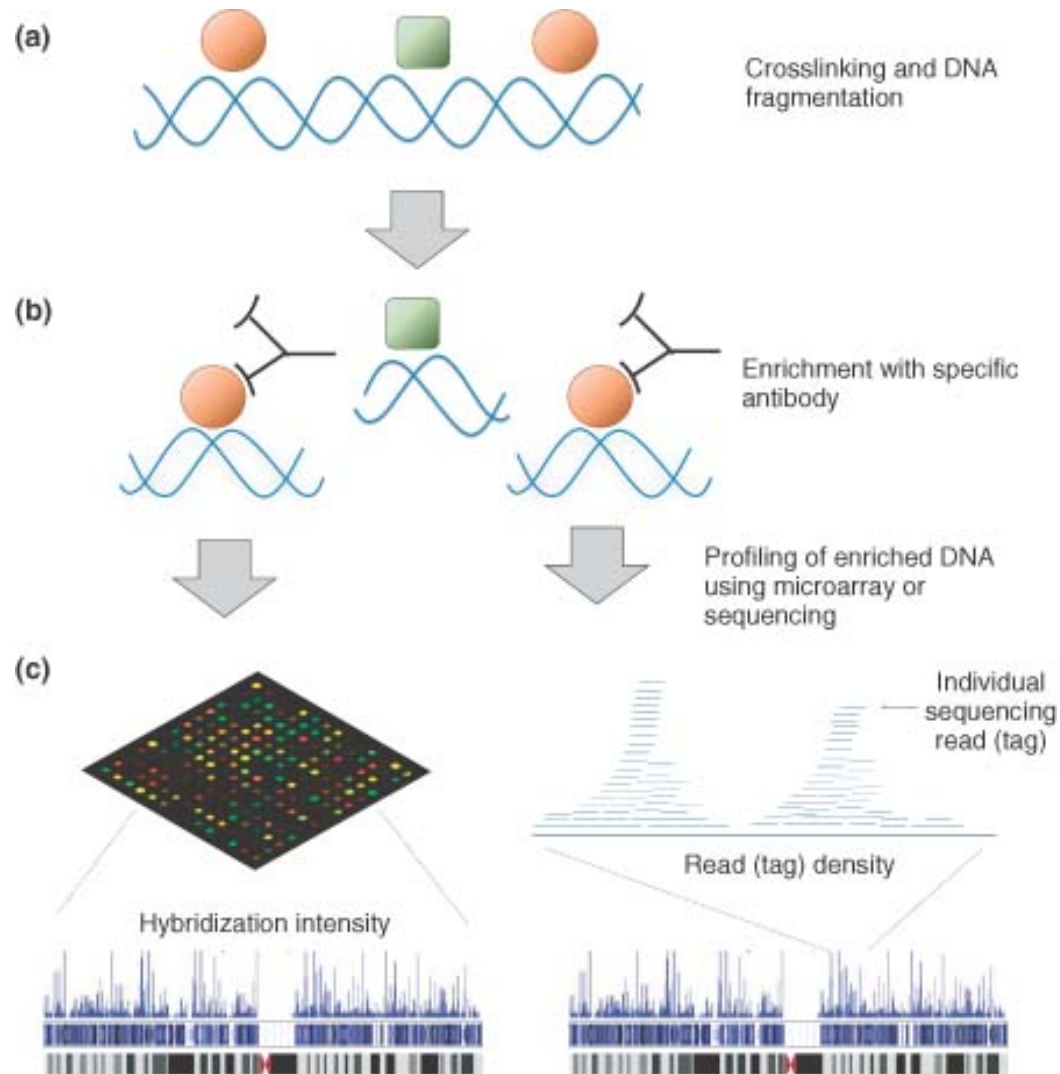
Epigenome Informatics Workshop
Bioinformatics Research Laboratory



Peak Calling

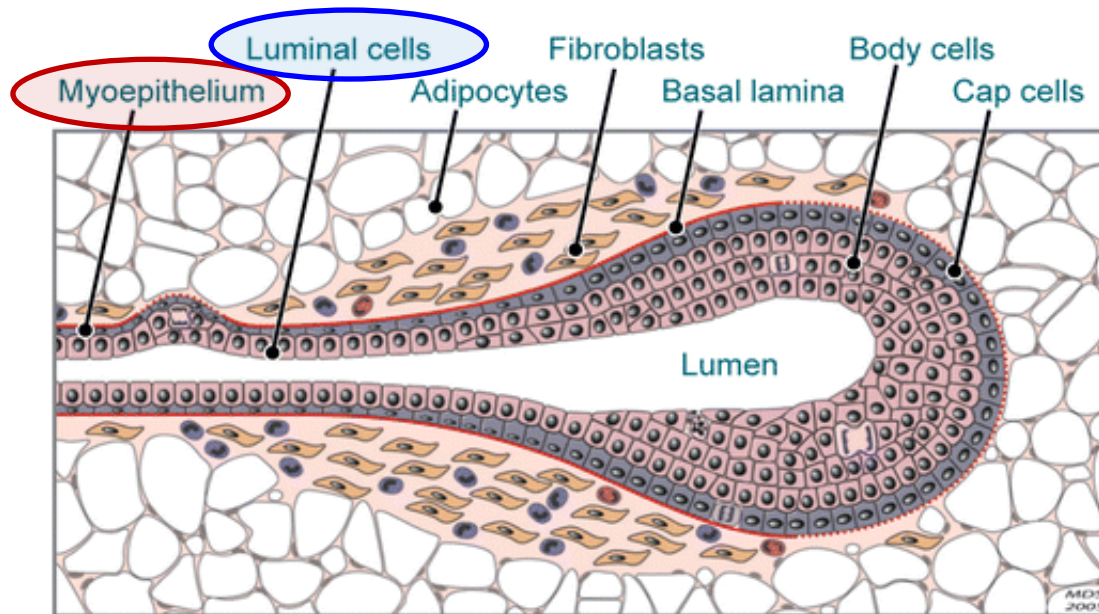
- Transcription Factors
- Histone modifications
 - H3K4me3, H3K27me3, H3K36me3, H3K9me3, H3K4me1, H3K27ac
- DNase hypersensitivity
- Digital Genomic Footprinting
- MeDIP-Seq, MRE-Seq
- Many others

Chromatin Immunoprecipitation Sequencing – ChIP-Seq



Breast Tissue Data from Epigenome Atlas

- **Breast Luminal Epithelial Cell (BLEC): H3K4me1**
 - ❑ H3K4me1 → mono-methylation of lysine 4 of H3 histone protein; associated with enhancers and DNA regions downstream of transcription starts
 - ❑ BRCA1 basal-like breast cancers originate from luminal epithelial progenitors
- **Breast Myoepithelial Cells (BMC): H3K4me3**
 - ❑ H3K4me3 → tri-methylation of lysine 4 of H3 histone protein; associated with promoters which are active or ready to be activated
 - ❑ Myoepithelial cells stabilize normal structure and help limit cancer growth



MACS

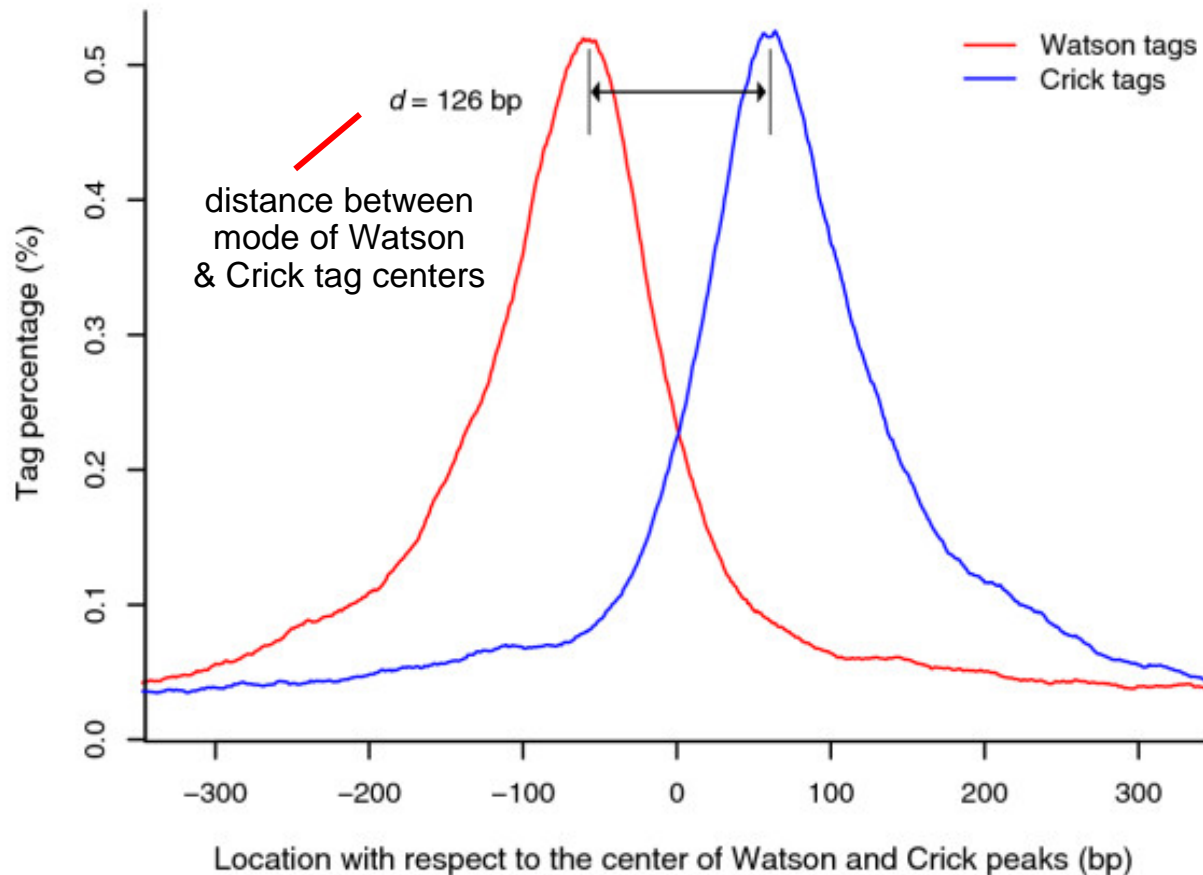
- Model Based Analysis of ChIP-Seq
- Call peaks on both strands
 - Empirical modeling of bandwidth (sonication size)
- Infer local distribution parameters
- Chip-Seq fragments equally likely to be sequenced from both ends (bimodal distribution around binding site)
- Account for local biases in sequencing
- Options
 - Run w/ control data
 - Run w/o control data

Zhang et al, Model-based Analysis of Chip-Seq (MACS) *Genome Biology* 2008, 9:R137

Feng et al, Using MACS to Identify Peaks from Chip-Seq Data, *Current Protocols in Bioinformatics* 2011, 34:2.14.1-2.14.14; John Wiley & Sons, Inc.

MACS

FoxA1 Chip-Seq (3.9 M uniquely mapped tags), Zheng et al



MACS in the Genboree Workbench

- Operates on BED input files
- 6 field format

Chr start stop name score strand

Chromosome	Start	Stop	Read Name	Score	Strand
------------	-------	------	-----------	-------	--------

- Upload BED files using the Workbench UI
 - Data/Files/Transfer File

Upload Your BED File

Please note: we have provided the BED file for you, this slide and the next show you how to upload your own BED files for future reference

The screenshot displays the Genboree web interface. At the top, there are tabs for 'System/Network', 'Data', 'QC and Pre-processing', 'Genome', 'Transcriptome', 'Cistrome', 'Epigenome', and 'Metagenome'. The 'Data' tab is selected. Below the tabs, there is a 'Welcome to Genboree! [Getting Started]' message. On the left, a 'Data Selector' panel shows a tree view of the database structure. A red arrow points from the 'Files' option in the 'Data Selector' to the 'Transfer File' option in the 'Data' menu. The 'Data' menu is open, showing options: 'Databases', 'Entity Lists', 'Entrypoints', 'Files', 'Projects', 'Samples & Sample Sets', and 'Tracks'. The 'Files' option is highlighted, and a sub-menu is open showing 'Transfer File', 'Process File', 'Copy/Move File', and 'Remove File(s)'. The 'Transfer File' option is highlighted. A tooltip for 'Transfer File' is visible, stating: 'Transfer a raw file to Genboree for storage, sharing, and/or subsequent analysis. The file will not be automatically processed following transfer. It will be stored in Genboree as-is.' Below the 'Data Selector', the 'GenboreeUser_database' is expanded, showing a tree view of the database structure. The 'Files' folder is highlighted. The 'Input Data' and 'Output Targets' panels are visible on the right side of the interface.

Select "Files" → "Transfer File"

Select Your BED File

the Genboree Workbench! [Getting Started]

Tool Settings

We recommend transferring gzipped-compressed files which will: (a) help avoid browser limitations, (b) decrease file transfer time, (c) not occupy excessive amounts of shared Genboree storage.

Additional details about the output files are provided in the Addendum at the bottom of the Help dialog.

Output Folder:

Group: GenboreeUser_group
Database: GenboreeUser_database
Sub-Folder:

Database File Settings

Select File No file chosen

Unpack/Extract File ☐

Convert to Unix? ☐

Create in Sub-Folder

File Description

Click to select BED file

Prepare to Run MACS in the Genboree Workbench (populate Input Data)

Welcome to the Genboree Workbench! [Getting Started]

Data Selector

Refresh Data Filter: Select a filter...

- Epigenome ToolSet Demo Input Data
 - Databases
 - Brain
 - Breast
 - Breast 450K
 - Disease Epigenome
 - MeDIP and GSEA
 - Peak Calling Demo
 - All Annotations in Database
 - Tracks
 - Lists & Selections
 - SampleSets
 - Samples
 - Files
 - macsInput
 - BLEC.H3K4me1.bed.gz
 - BLEC.Input.bed.gz
 - BMC.H3K4me3.bed.gz**

Details

Attribute	Value
Download	Click to Download File
Group	Epigenome ToolSet Demo Input Data
Database	Peak Calling Demo
Description	
Name	macsInput/BMC.H3K4me3.bed.gz
	Mon May 14 2012
	Tue May 15 2012

Input Data

⬆ ⬇ ✖

BMC.H3K4me3.bed.gz

Output Targets

⬆ ⬇ ✖

Drag the input BED file into "Input Data"

Prepare to Run MACS in the Genboree Workbench (populate Output Targets)

The screenshot shows the Genboree Workbench interface. At the top is a navigation bar with tabs: System/Network, Data, QC and Pre-processing, Genome, Transcriptome, Cistrome, Epigenome, Metagenome, Visualization, and Help. Below the navigation bar is a welcome message: "Welcome to the Genboree Workbench! [Getting Started]".

The main interface is divided into four panels:

- Data Selector:** Contains a tree view of data sources. A red dashed arrow points from the "GenboreeUser_database" entry under the "GenboreeUser_group" to the "Output Targets" panel. A red box with the word "Drag" is placed over the arrow.
- Details:** A table showing user information for "GenboreeUser_group".
- Input Data:** A section for uploading input data, currently showing "BMC.H3K4me3.bed.gz".
- Output Targets:** A section for designating output targets, currently showing "GenboreeUser_database".

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

Input Data
BMC.H3K4me3.bed.gz

Output Targets
GenboreeUser_database

Designate the database where you want the MACS results to be deposited by dragging the it into "Output Targets"

Invoke MACS in the Genboree Workbench

The screenshot displays the Genboree Workbench interface. At the top, a navigation bar includes tabs for System/Network, Data, QC and Pre-processing, Genome, Transcriptome, Cistrome, Epigenome, Metagenome, Visualization, and Help. Below this, a 'Call ChIP-Seq Peaks by' dropdown menu is open, showing 'MACS' as the selected option. A red arrow points from a red box labeled 'Invoke MACS' to the 'MACS' button. The main area is divided into three panels: 'Data Selector' on the left, 'Details' in the top right, and 'Input Data' and 'Output Targets' in the bottom right. The 'Data Selector' panel shows a tree view of data sources, with 'GenboreeUser_group' and 'GenboreeUser_database' selected. The 'Details' panel shows a table of user information. The 'Input Data' panel shows a file 'BMC.H3K4me3.bed.gz' and the 'Output Targets' panel shows 'GenboreeUser_database'.

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
 - Databases
 - Brain
 - Breast
 - Breast 450K
 - Disease Epigenome
 - MeDIP and GSEA
 - Peak Calling Demo
 - RNA-Seq Tool Demo
 - Projects
 - Epigenomics Roadmap Repository
 - GenboreeUser_group
 - Databases
 - GenboreeUser_database
 - Projects
 - GMT_Tutorial

Details

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

Input Data

↑ ↓ ✕

BMC.H3K4me3.bed.gz

Output Targets

↑ ↓ ✕

GenboreeUser_database

Invoke MACS

Select Settings in the MACS Dialogue Window

The screenshot shows the 'Tool Settings' window for the MACS tool. It is divided into three main sections: 'Input Files', 'Output Location', and 'MACS Settings'. The 'Input Files' section shows the database as 'Peak Calling Demo', group as 'Epigenome ToolSet Demo Input Data', and file as 'macsInput/BMC.H3K4me3.bed.gz'. The 'Output Location' section shows the database as 'GenboreeUser_database' and group as 'GenboreeUser_group'. The 'MACS Settings' section contains various parameters: 'Analysis Name' (MACS-2013-3-1-18:49:19), 'Results Name' (UC14_MACS-2013-3-1-18:49:1), 'Small Local' (1000), 'Large Local' (10000), 'P Value' (1e-5), 'M Fold' (10 30), 'Upload Results ?' (checked), 'Track Name' (BMC : H3K4me3), 'No Lambda' (unchecked), 'Off Auto' (unchecked), 'No Model' (unchecked), 'MACS Version' (1.4), and 'File Format' (BED). Red arrows point from external text boxes to specific fields: 'p-value cutoff to call peaks' points to the 'P Value' field, 'min, max fold change' points to the 'M Fold' field, 'LFF track name (these boxes will appear after checking "Upload Results")' points to the 'Track Name' field, and 'After selecting settings, click Submit.' points to the 'Submit' button.

Tool Settings

MACS

Tool Overview

Input Files:

Database: Peak Calling Demo
Group: Epigenome ToolSet Demo Input Data
File: macsInput/BMC.H3K4me3.bed.gz

Output Location:

Database: GenboreeUser_database Group: GenboreeUser_group

MACS Settings

Analysis Name: MACS-2013-3-1-18:49:19
Results Name: UC14_MACS-2013-3-1-18:49:1
Small Local: 1000
Large Local: 10000
P Value: 1e-5
M Fold: 10 30
Upload Results ? ☒
Track Name: BMC : H3K4me3
No Lambda ☐
Off Auto ☐
No Model ☐
MACS Version: 1.4
File Format: BED

Submit Cancel

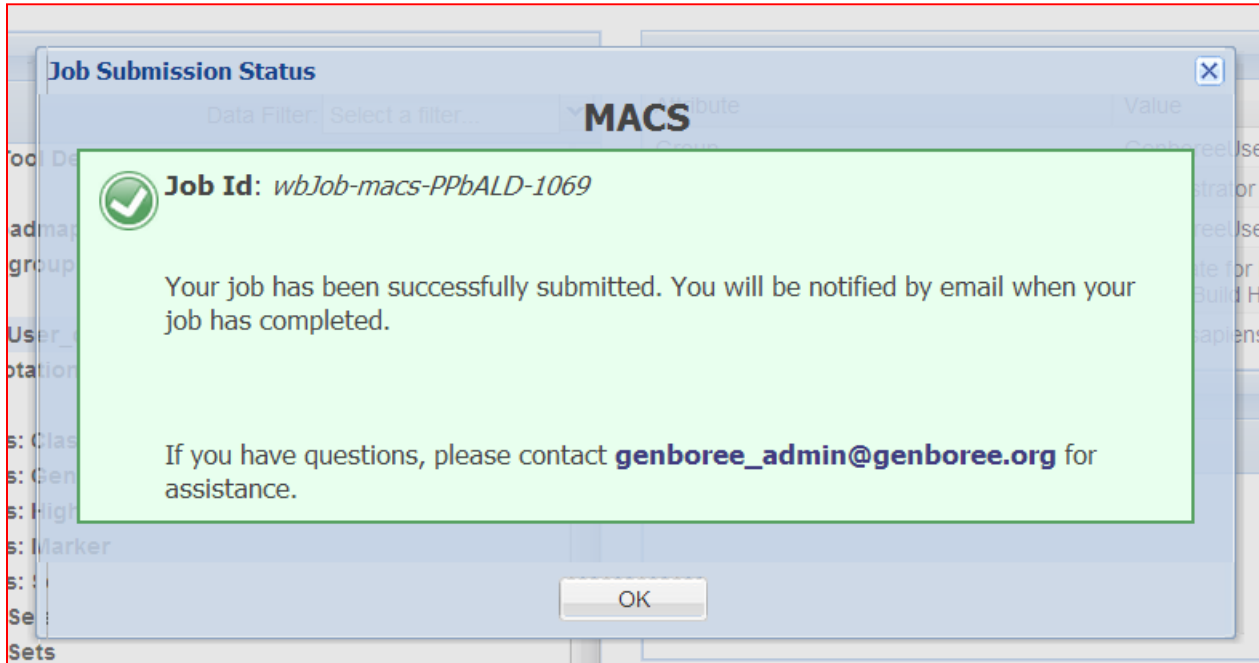
p-value cutoff
to call peaks

min, max
fold change

LFF track name
(these boxes will appear
after checking "Upload
Results")

After selecting settings,
click Submit.

MACS Job Submission Status



Genboree email Notifying You MACS Job is Complete

Hello Genboree User,

Your job completed successfully.

Job Summary:

JobID - wbJob-macs-PPbALD-1069

Additional Info:

Database: 'GenboreeUser_database'

Group: 'GenboreeUser_group'

You can download result files from the 'MACS-2013-2-18-14:53:22' folder under the 'MACS' directory.

MACS called 23943 peaks. The breakdown of the peaks with respect to the gene model follows:

Gene TSS +/- 3K Peaks	13693	57.18%
Gene Exon Peaks	868	3.62%
Gene Intron Peaks	2707	11.3%
Intergenic Peaks	6675	27.87%

- The Genboree Team

Retrieve MACS Results from “Files” in Your Designated Database

Welcome to the Genboree Workbench! [Getting Started]

Data Selector

Refresh Data Filter: Select a filter...

- GenboreeUser_database
 - All Annotations in Database
 - Tracks
 - Lists & Selections
 - Sample Sets
 - Samples
 - Files
 - Epigenome Slice
 - EpigenomicExpHeatmap
 - MACS
 - NIH1_MACS-2012-5-14-12:38:42
 - UC14_MACS-2013-3-1-19:41:12
 - BMC:H3K4me3.lff.zip
 - MACS-2013-3-1-19:41:12_peaks.xls
 - jobFile.json
 - raw
 - Raw Data Files
 - Spark - Results
 - EpigenomeCompLIMMA
 - Epigenome_Limma
 - Cuffdiff

Details

Attribute	Value
Download	Click to Download File
Group	GenboreeUser_group
Database	GenboreeUser_database
Description	
Name	MACS/UC14_MACS-2013-3-1-19:41:12

Input Data

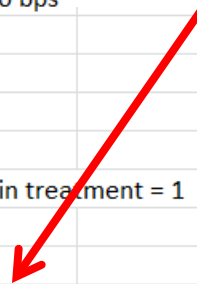
Clicking on a file in the Data Selector will highlight it in Details. Then click on the link in Details to download the file.

The MACS results file will be located within Files, within the database that you designated.

MACS Results Downloaded in Excel

	A	B	C	D	E	F	G	H	I
1	# This file is generated by MACS version 1.4.1 20110627								
2	# ARGUMENTS LIST:								
3	# name = BMC_H3K4Me3_MACS-2013-2-18-14%3A53%3A22								
4	# format = AUTO								
5	# ChIP-seq file = ./9211e5b36b69c5ae404791db02a1d981af29e13e/file_0_macsinpout%2FBMC.H3K4me3.bed								
6	# control file = None								
7	# effective genome size = 2.70e+09								
8	# band width = 300								
9	# model fold = 10,30								
10	# pvalue cutoff = 1.00e-05								
11	# Small dataset will be scaled towards larger dataset.								
12	# Range for calculating regional lambda is: 10000 bps								
13									
14	# tag size is determined as 49 bps								
15	# total tags in treatment: 8184971								
16	# tags after filtering in treatment: 8184971								
17	# maximum duplicate tags at the same position in treatment = 1								
18	# Redundant rate in treatment: 0.00								
19	# d = 200								
20	chr	start	end	length	summit	tags	-10*log10	fold_enrichment	
21	chr1	9861	10677	817	360	292	2034.63	24.91	
22	chr1	713307	715489	2183	1297	83	250.89	11.7	
23	chr1	724760	727160	2401	2202	92	223.86	16.37	
24	chr1	761468	763266	1799	865	133	463.34	14.42	
25	chr1	833096	834002	907	577	20	96.47	11.54	
26	chr1	839750	840203	454	262	15	77.67	11.4	
27	chr1	859636	861418	1783	741	37	83.16	5.71	
28	chr1	875788	876527	740	305	13	50.73	7.95	

Summit = location with highest
fragment pileup, predicted as
precise binding location



MAC Results Uploaded to Genboree

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

GENBOREE

BCM
Baylor College of Medicine

System/Network Data UC and Pre-processing Genome Transcriptome Cistrome Epigenome Metag

Welcome to the Genboree Workbench! [Getting Started]

Data Selector

Refresh Data Filter: Select a filter...

- Epigenomics Roadmap Repository
- GenboreeUser_group
 - Databases
 - GenboreeUser_database
 - All Annotations in Database
 - Tracks
 - Lists & Selections
 - SampleSets
 - Samples
 - Files
 - EpigenomeSlice
 - EpigenomicExpHeatmap
 - MACS
 - NIH1_MACS-2012-5-14-12:38:42
 - UC14_MACS-2013-3-1-19:41:12
 - BMC:H3K4me3.Iff.zip
 - MACS-2013-3-1-19:41:12_peaks.xls
 - jobFile.json
 - raw
 - Raw Data Files
 - Spark - Results

Details

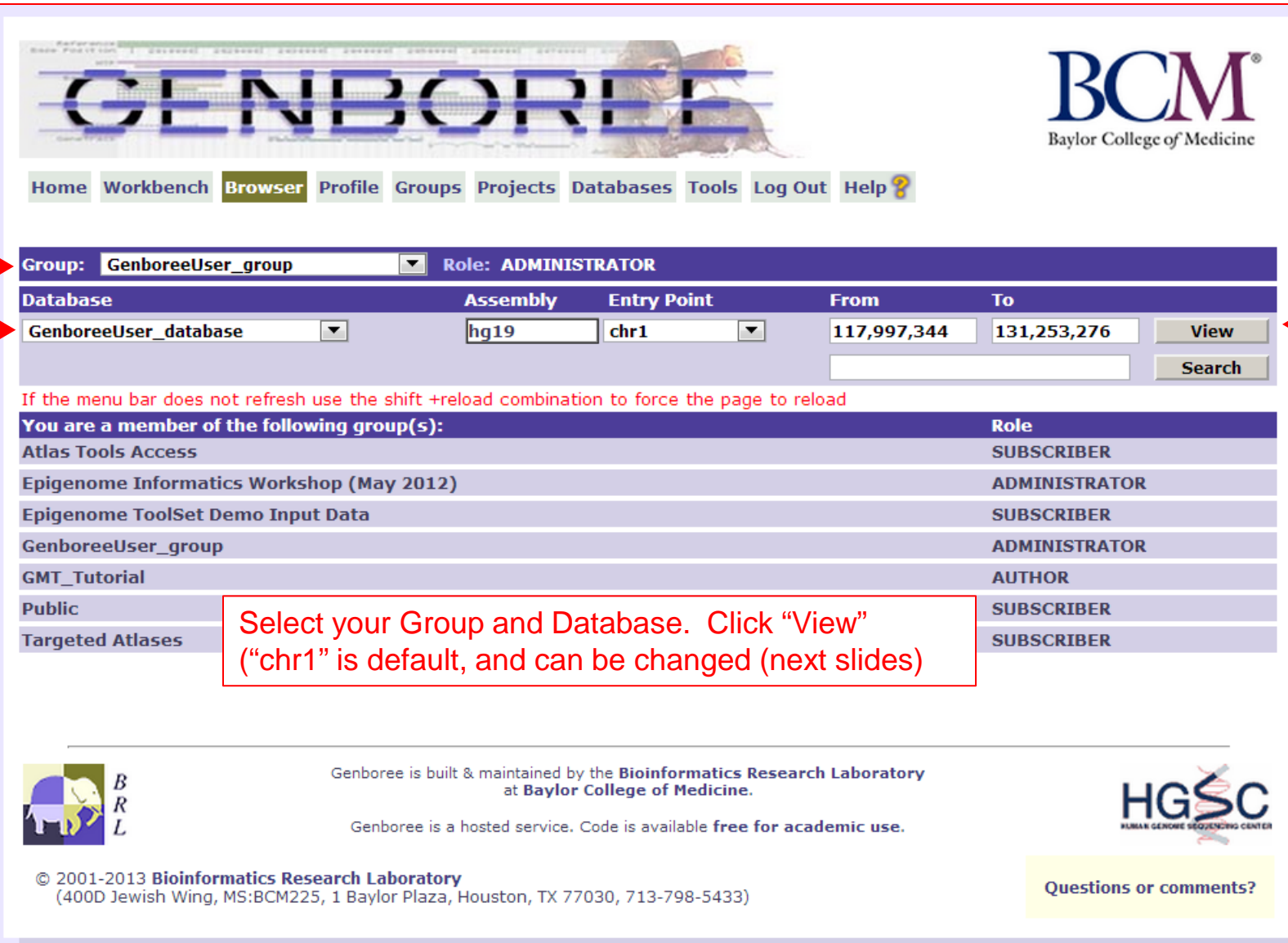
Attribute	Value
Download	Click to Download File
Group	Epigenome ToolSet Demo

Input Data

Output Targets

We now wish to visualize the peaks called by MACs in the context of genomic and/or epigenomic data via the Genboree Browser

Tell the Genboree Browser What You Wish to View



The screenshot shows the Genboree Browser interface. At the top, there is a navigation bar with links: Home, Workbench, **Browser**, Profile, Groups, Projects, Databases, Tools, Log Out, and Help. The BCM Baylor College of Medicine logo is in the top right. Below the navigation bar, a purple header bar displays 'Group: GenboreeUser_group' and 'Role: ADMINISTRATOR'. Below this, a form contains dropdowns for 'Database' (GenboreeUser_database) and 'Assembly' (hg19), followed by an 'Entry Point' dropdown (chr1) and 'From'/'To' range inputs (117,997,344 to 131,253,276). A 'View' button and a 'Search' button are on the right. A red box with text is overlaid on the bottom part of the page.

Group: GenboreeUser_group **Role:** ADMINISTRATOR

Database	Assembly	Entry Point	From	To	
GenboreeUser_database	hg19	chr1	117,997,344	131,253,276	View
					Search

If the menu bar does not refresh use the shift +reload combination to force the page to reload

You are a member of the following group(s):

	Role
Atlas Tools Access	SUBSCRIBER
Epigenome Informatics Workshop (May 2012)	ADMINISTRATOR
Epigenome ToolSet Demo Input Data	SUBSCRIBER
GenboreeUser_group	ADMINISTRATOR
GMT_Tutorial	AUTHOR
Public	SUBSCRIBER
Targeted Atlases	SUBSCRIBER

Select your Group and Database. Click "View" ("chr1" is default, and can be changed (next slides))

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**.

© 2001-2013 **Bioinformatics Research Laboratory**
(400D Jewish Wing, MS:BCM225, 1 Baylor Plaza, Houston, TX 77030, 713-798-5433)

HGSC
HUMAN GENOME SEQUENCING CENTER

Questions or comments?

View MACS Results in the Genboree Browser

Data Selector

- GenboreeUser_group
 - Databases
 - GenboreeUser_database
 - All Annotations in Database
 - Tracks
 - Lists & Selections
 - SampleSets
 - Samples
 - Files
 - EpigenomeSlice
 - EpigenomicExpHeatmap
 - MACS
 - NIH1_MACS-2012-5-14-12:38:42
 - UC14_MACS-2013-3-1-19:41:12
 - BMC:H3K4me3.Iff.zip
 - MACS-2013-3-1-19:41:12 peaks.xls
 - jobFile.json
 - raw

Genboree Browser

Role: ADMINISTRATOR

Assembly: hg19 Entry Point: chr1 From: 117,997,344 To: 131,253,276

Zoom: 1.5x 2x 3x 5x 10x Base Zoom Out 1.5x 2x 3x 5x 10x Full

Tracks:

- Track "BMC:H3K4me3" (23,943 annotations)
- Track "ESCs:Rad21_Nanog" (218 annotations)
- Track "Read:Density_BodySite" (1,552 annotations)
- Track "Read:Density_Individual" (248 annotations)
- Track "Cyto:Band" (862 annotations)
- Track "RefSeq" (345,623 annotations)
- Track "Genes" (244,393 annotations)
- Track "Known:Gene" (733,510 annotations)

The MACS files contained in the Data Selector are now displayed in the Genboree Browser

Change View of BMC:H3Kme3 Data

Select chr17 and type in new "From" and "To" coordinates to generate this view

The screenshot displays the Genboree web interface for viewing genomic data. At the top, the user is logged in as 'GenboreeUser_group' with the role of 'ADMINISTRATOR'. The interface shows the following configuration:

- Database:** GenboreeUser_database
- Assembly:** hg19
- Entry Point:** chr17
- From:** 37,841,179
- To:** 37,892,422

Navigation and zoom controls are available below the configuration fields:

- Extend:** A range of 2,000 bp is shown with navigation arrows.
- Zoom In:** Buttons for 1.5x, 2x, 3x, 5x, 10x, and Base.
- Zoom Out:** Buttons for 1.5x, 2x, 3x, 5x, 10x, and Full.

The main view displays several genomic tracks for chromosome 17 (chr17) in the specified region:

- BMC:H3K4me3:** A track showing H3K4me3 enrichment with black bars. A red arrow points to this track. Annotations include chr17_3784384... and chr17_3785566....
- ESCs:Rad21_Nanog:** A track showing Nanog binding sites with vertical bars.
- Cyto:Band:** A track showing cytobands, with a label '17q12'.
- Gene:RefSeq:** A track showing gene models with blue exons and introns.
- CCDS:Genes:** A track showing gene models with purple exons and introns.
- Known:Gene:** A track showing gene models with green exons and introns.

Each track is labeled with its name and the number of annotations (e.g., BMC:H3K4me3 has 23,943 annotations).

Explore Visualization Settings in Genboree Browser

The screenshot shows the Genboree Browser main interface. At the top, it displays the user group as 'GenboreeUser_group' and the role as 'ADMINISTRATOR'. Below this, there are dropdown menus for 'Database' (GenboreeUser_database), 'Assembly' (hg19), 'Entry Point' (chr17), and a range of coordinates (37,841,179 to 37,892,422). A 'View' button is next to the coordinates. Below the dropdowns, there are navigation controls including 'Extend' with left and right arrows and a value of 2,000, and 'Zoom In' and 'Zoom Out' buttons with various magnification levels (1.5x, 2x, 3x, 5x, 10x, Base, Full). A 'Search' button is also present. The main area shows a genomic track for chr17 with a scale from 37,850,000 to 37,890,000. Two tracks are visible: 'BMC:H3K4me3' and 'ESCs:Rad21_Nanog'. A red box highlights the 'Change settings to view data of interest' text.

Change settings to view data of interest

The screenshot shows the Genboree Browser settings panel. At the top, there are 'Reset' and 'Apply' buttons. Below them, there is a 'Set Visibility For All Tracks' dropdown and a 'Picture width' field set to 620. There are three checkboxes: 'Display empty tracks' (unchecked), 'Indicate continuations?' (unchecked), and 'Display track descriptions' (checked). The settings are organized into sections: 'Class', 'Gene', 'MACS', 'Marker', and 'SPARK'. Each section has a collapse/expand icon on the left and a green arrow icon on the right. The 'Class' section includes 'EpigenomicLimmaComp:Analysis' (Compact), 'Read:Density_BodySite' (Compact), and 'Read:Density_Individual' (Compact). The 'Gene' section includes 'CCDS:Genes' (Compact) and 'Gene:RefSeq' (Compact). The 'MACS' section includes 'BMC:H3K4me3' (Expand with Comments). The 'Marker' section includes 'Cyto:Band' (Compact). The 'SPARK' section includes 'ESCs:Rad21_Nanog' (Compact). At the bottom, there are buttons for 'Back', 'Apply', 'Save', 'Load Default', and 'Save As Default'.

Peak evaluation

- H3K4me3
 - Actively transcribed genes
 - Poised genes
 - co-occurring with H3K27me3
 - Overlap with gene promoters
- H3K4me1
 - Mark of distal enhancers
- Expect different distribution H3K4me3
- Simple check
 - Summary of peak distribution
 - With respect to gene elements

	H3K4me3		H3K4me1	
TSS+/-3k	13693	57.2 %	5381	17.7%
Exons	868	3.6%	2154	7.1%
Introns	2707	11.3%	8416	27.7%
Intergenic	6675	27.9 %	14429	47.5%

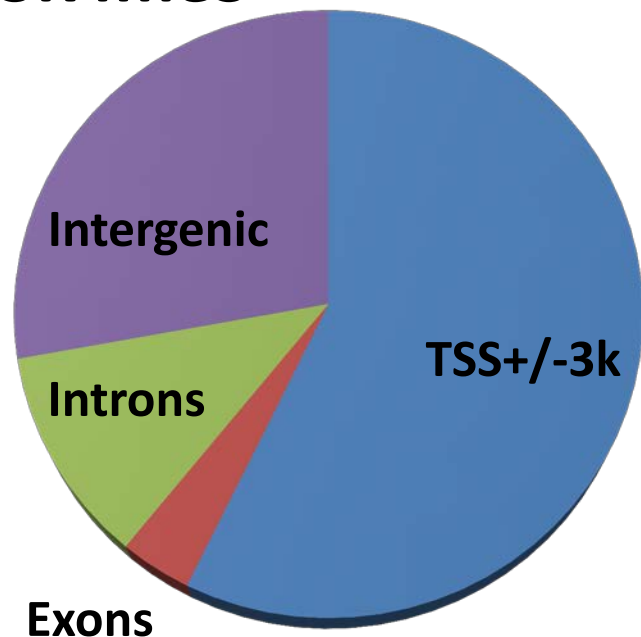
Total

23943

30380

*Please note: Genboree does not generate these visuals
(data is exported and pie charts are generated in Excel)*

H3K4me3



H3K4me1

