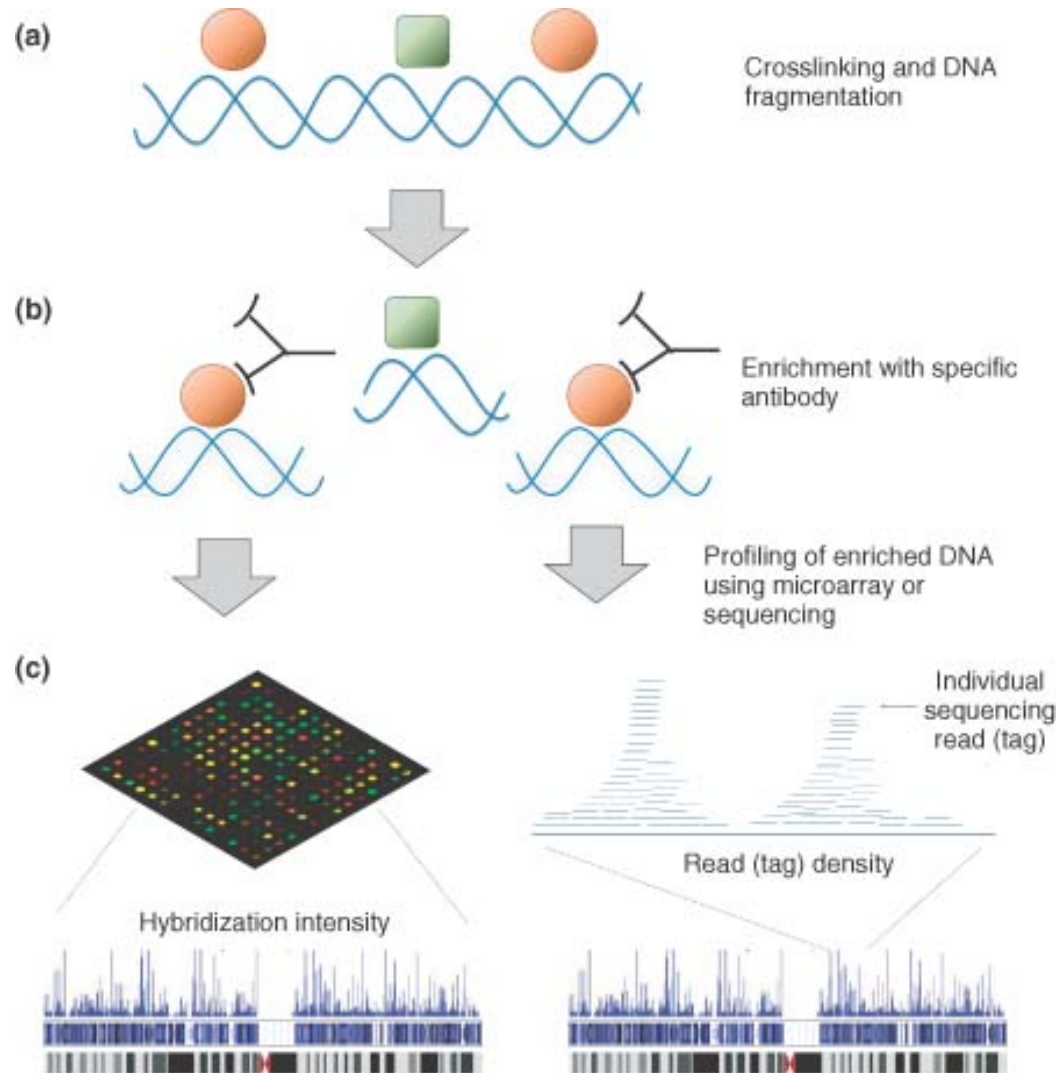


ChIP-Seq, mRNA-Seq, &
Resequencing
via the Genboree Workench

Chromatin Immunoprecipitation Sequencing – ChIP-Seq



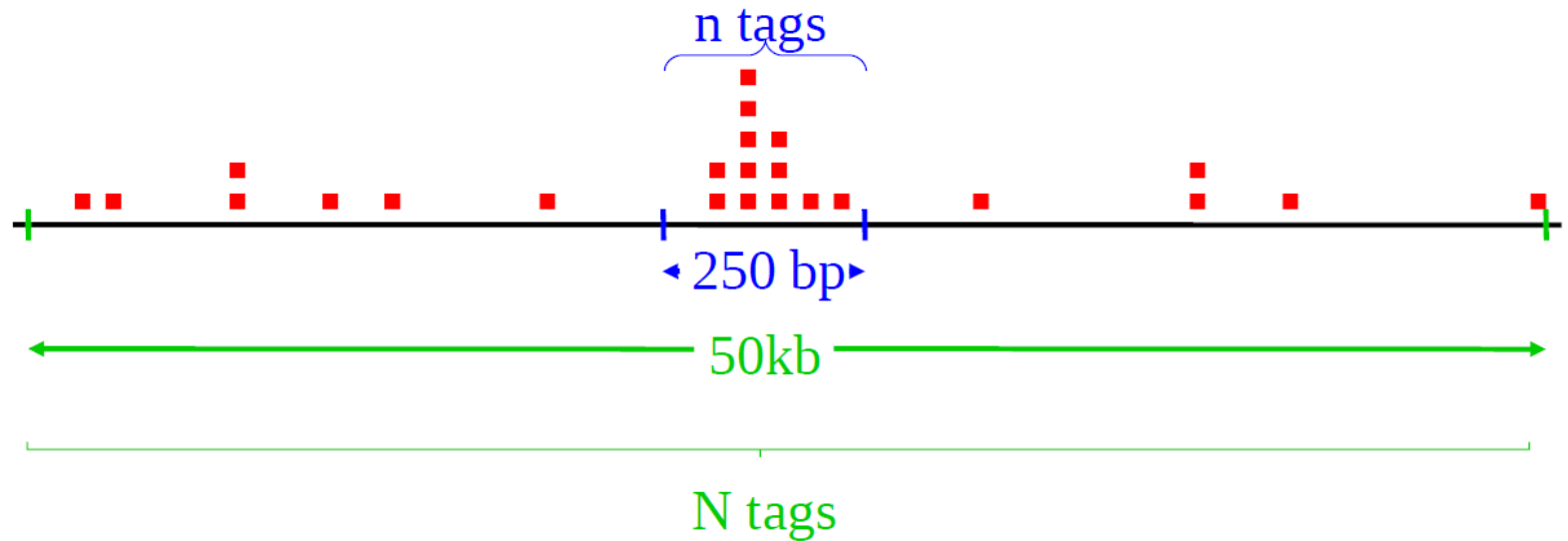
Peak Calling

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

Peak Calling

- Input from
 - NIH Epigenomic Roadmap mapping centers
 - ENCODE analysis group
- Deployed at EDACC/BRL
 - MACS (Harvard)
 - FindPeaks (BCGSC/UCSF)
 - HotSpot (UW)
- Tools under consideration
 - Scripture and PostProcessing (BI)
- Future tools
 - PeakSeq
 - MRE/MeDIP integration tools from WashU/UCSF

HotSpot



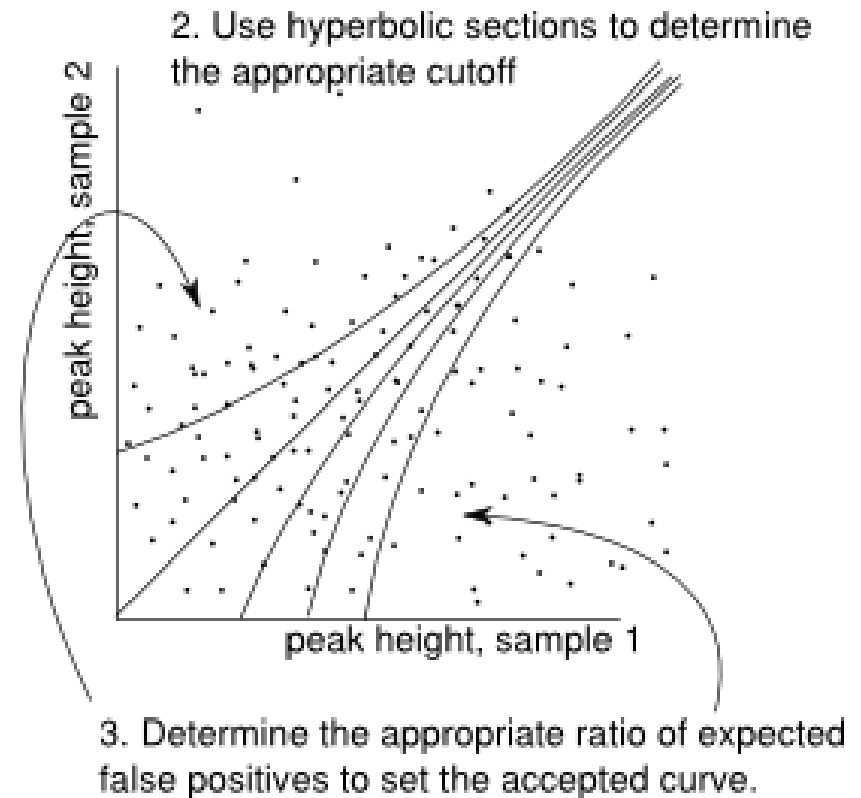
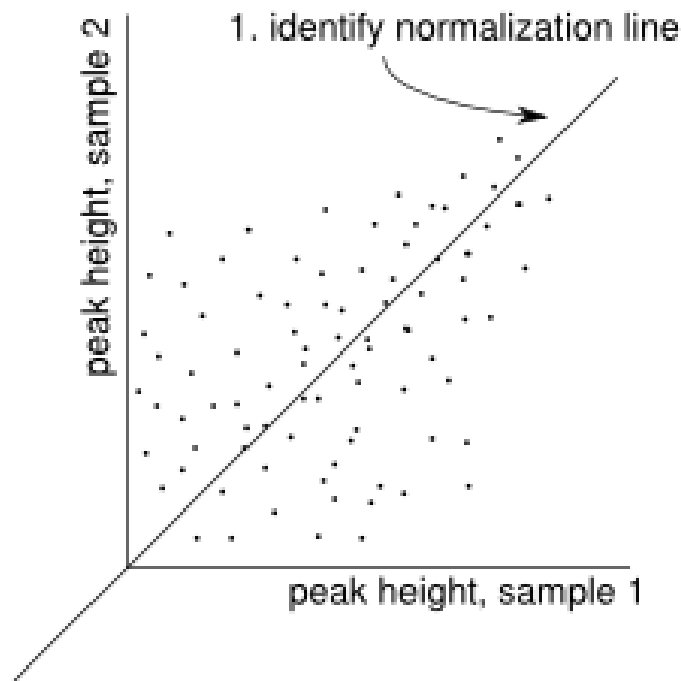
- Scan statistic gauging enrichment
 - z-score based on the binomial distribution.
- Binomial distribution
 - probability of n tags in small window given N tags large window.
 - adjust for local background fluctuations

FindPeaks

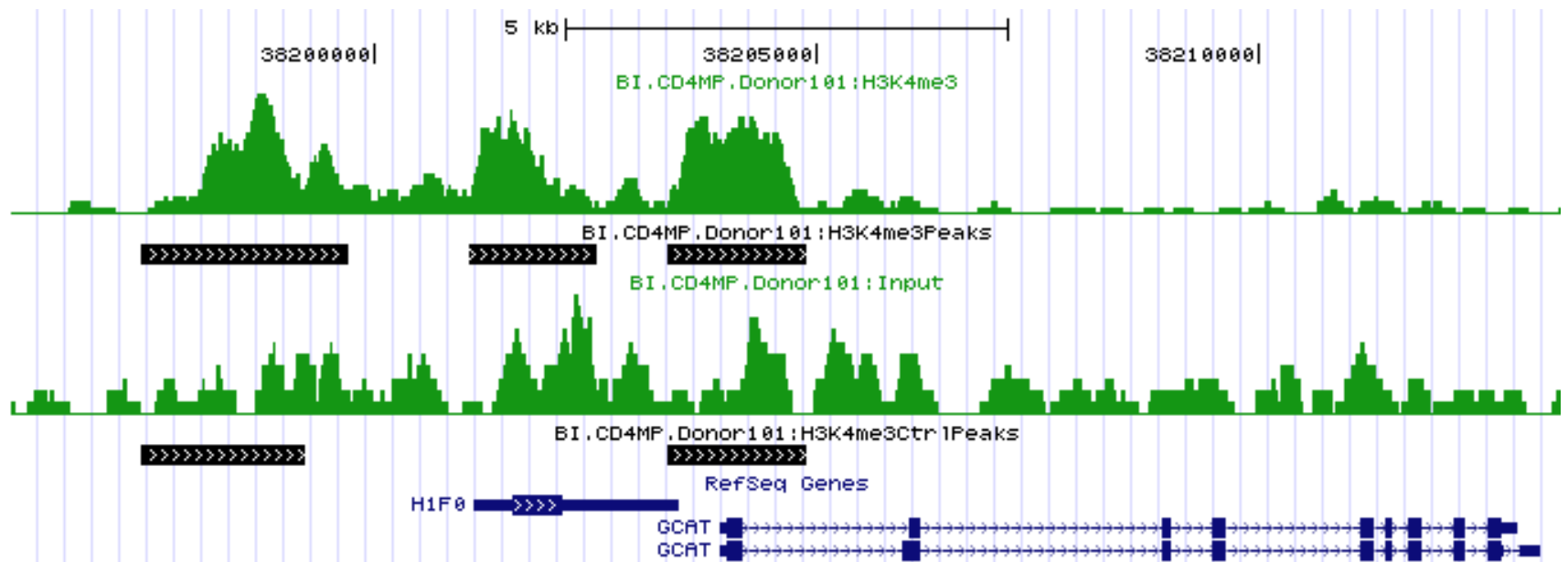
- In the absence of input
 - Determine areas of enrichment
 - Simulate coverage of genome to determine FDR
 - Subset Peaks at $\text{FDR}=0.01$

FindPeaks

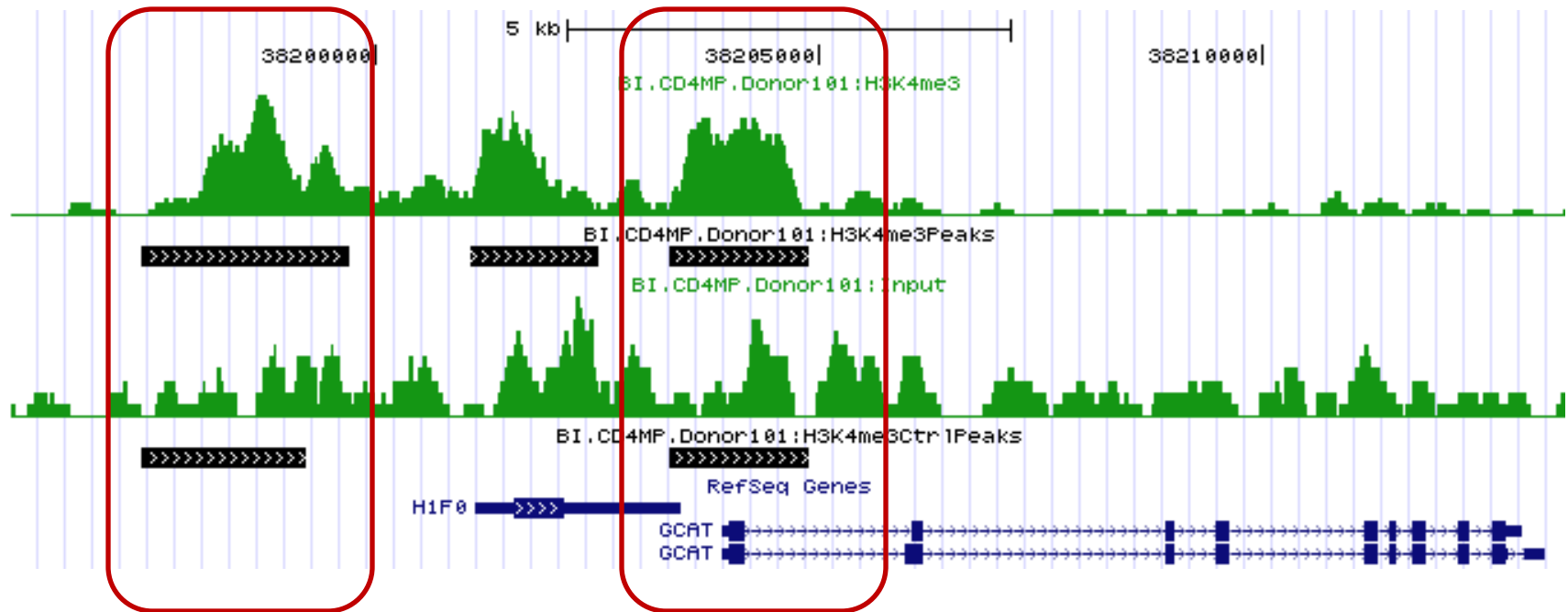
In the presence of input



FindPeaks



FindPeaks

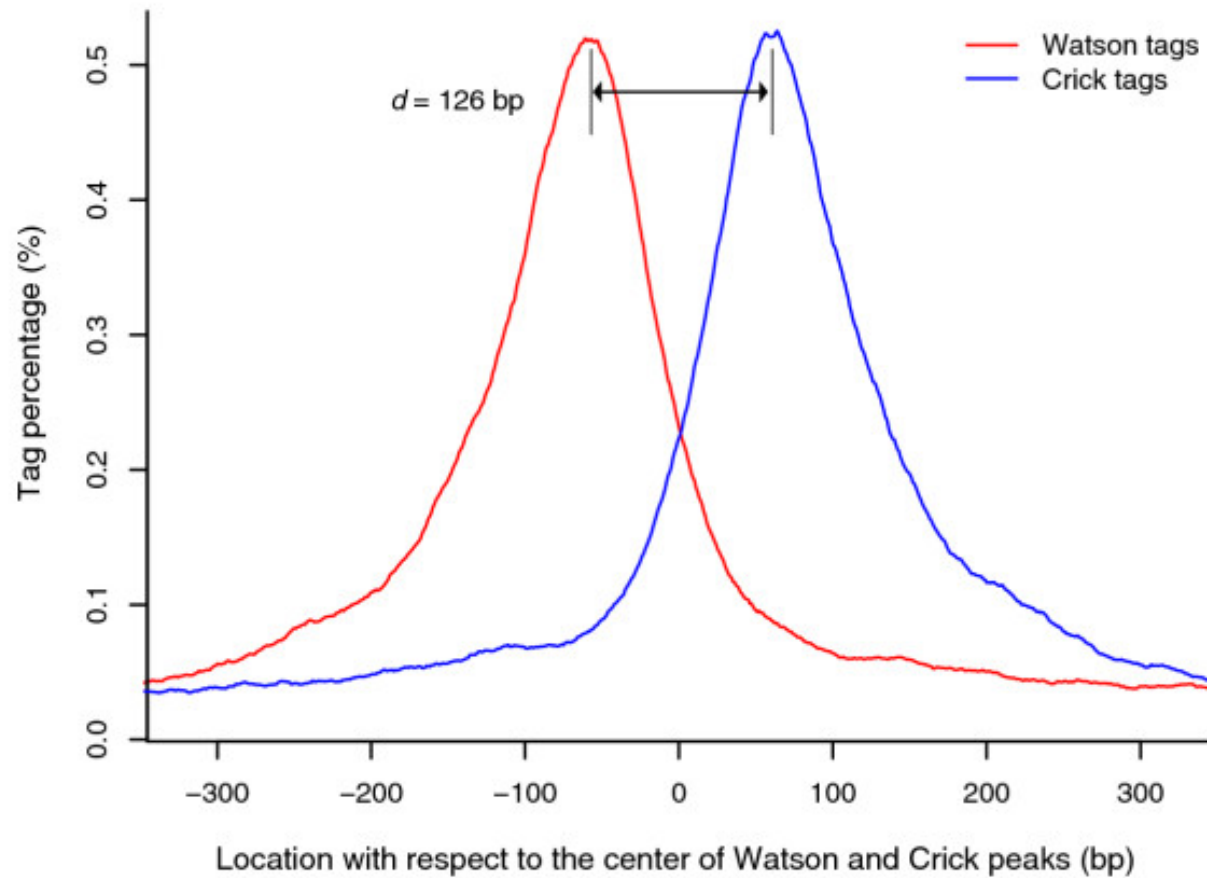


Only peaks above background are reported

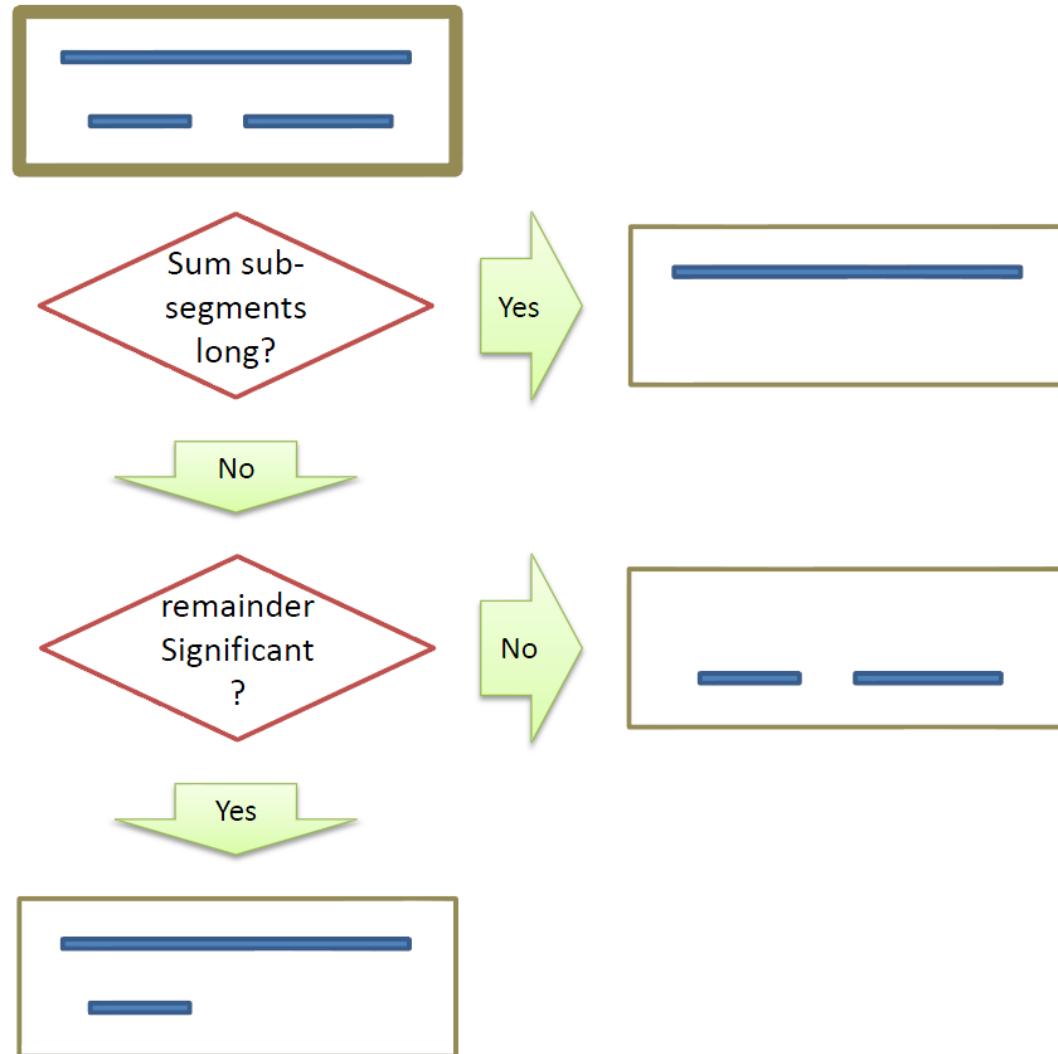
MACS

- Model Based Analysis of ChIP-Seq
- Call peaks on both strands
 - Infer bandwidth length
- Infer local distribution parameters
- Account for local biases in sequencing

MACS



Post Scripture Multi Scale Segmentation

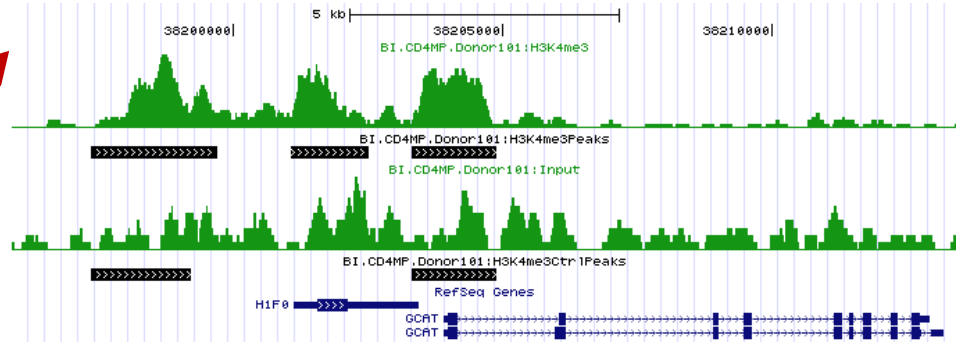


Deployed Tools

- MACS, FindPeaks
 - Histone modifications, transcription factors
- HotSpot
 - DNASE1/DGF
- Peaks called on all the experiments
 - Release 1 through Release 5
 - ChIP-Seq/DNase1/Digital Genomic Footprinting
 - ~900 experiments (out of ~1500 total)

Exposing NIH Epigenomic Consortium Peak Calls

UCSC Visualization



Human Epigenome Atlas Release 3 (ng19)

View Selections In Clear Selections Save Selections

Assay Type

Sample Type

Filter:

	Blue-Seq	MeDIP-Seq	MF-Seq	RRBS	DNAse Hypersensitivity	Digital Genomic Footprinting	mRNA-Seq	snRNA-Seq	ChIP-Seq Input	ChIP-Seq H3K27me3	ChIP-Seq H3K36me3	ChIP-Seq H3K9me3	Histone H2Ame1	Histone H3me1	Histone H3me3
Adipose Derived Mesenchymal Stem Cells									3	3	3	3			
Adipose Nuclei									5	4	5	5			
Adult Kidney				1					2	2					
Adult Liver									3	2	3	3			
Bone Marrow Derived Mesenchymal Stem Cells									4	4	4	4			
Brain Angular Gyrus									2	1	2	2			
Brain Anterior Caudate									2	2	2	2			

HTTP/FTP Download

**Index of /EdaccData/Release-5/study-sample-experiment
/BI/CD4_Memory_Primary_Cells/Histone_H3K4me3/**

BI.CD4 Memory Primary Cells.H3K4me3.20.bed.gz	28-Dec-2011	18:27	84M
BI.CD4 Memory Primary Cells.H3K4me3.20.wig.gz	21-Dec-2011	21:14	18M
BI.CD4 Memory Primary Cells.H3K4me3.Donor 100 7..>	28-Dec-2011	18:27	418M
BI.CD4 Memory Primary Cells.H3K4me3.Donor 100 7..>	21-Dec-2011	21:14	45M
BI.CD4 Memory Primary Cells.H3K4me3.Donor 101 8..>	28-Dec-2011	18:27	312M
BI.CD4 Memory Primary Cells.H3K4me3.Donor 101 8..>	21-Dec-2011	21:14	37M

MACS in the Genboree Workbench

The screenshot displays the Genboree Workbench interface. At the top, a navigation bar includes links for Home, Workbench, Browser, Profile, Groups, Projects, Databases, Tools, Log Out, and Help. The Genboree logo and the BCM Baylor College of Medicine logo are visible in the header. On the left, a 'Data Selector' tree shows a hierarchy starting from 'genboree.org' down to 'EDACC_Workshop_Test'. A 'Welcome to the Genboree' message is displayed below the selector. A 'Tool Settings' dialog box is open in the center, titled 'Additional details about the output files are provided in the Addendum at the bottom of the Help dialog.' The dialog has two main sections: 'Output Folder:' and 'Database File Settings'. The 'Output Folder:' section contains fields for 'Group:' (cristian_coarfa_group), 'Database:' (EDACC_Workshop_Test), and 'Sub-Folder:'. The 'Database File Settings' section includes a 'Select File' field with the path 'C:\Users\coarfa\Docum...' and a 'Browse...' button, an 'Unpack Multi-File Archive' checkbox, a 'Create in Sub-Folder' field with the value 'BedFiles', and a 'File Description' text area. At the bottom of the dialog are 'Submit' and 'Cancel' buttons. In the background, a list of items is visible, including 'an_coarfa_group', 'ate for Human Genome, UCSC', 'Hg19', and 'C_Workshop_Test'. A large red text overlay at the bottom right reads 'Upload BED file'.

Home Workbench Browser Profile Groups Projects Databases Tools Log Out Help

GENBOREE

BCM[®]
Baylor College of Medicine

Data Analysis Visualization

Welcome to the Genboree

- The **Data Selector** tree on the left
- Drag items to be used as tool *inputs*
- Drag items to be used as *output destinations*
- Tools which can be run on your selected items
- Unsure about what kinds of items are available?
 - Just click the tool button when it appears

Data Selector

Refresh

- genboree.org
 - 1. Scratch-Man
 - cristian_coarfa_group
 - Databases
 - EDACC Workshop , \$
 - EDACC_Workshop_Test
 - Projects
 - Encode Repository
 - Epigenome Informatics Workshop (M
 - Epigenome ToolSet Demo
 - Epigenomics Roadmap Repository
 - Public
 - ROI Repository
 - Targeted Atlases

Tool Settings

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

Output Folder:

Group: cristian_coarfa_group
Database: EDACC_Workshop_Test
Sub-Folder:

Database File Settings

Select File C:\Users\coarfa\Docum Browse...

Unpack Multi-File Archive ☐

Create in Sub-Folder BedFiles

File Description

Submit Cancel

EDACC_Workshop_Test

an_coarfa_group
ate for Human Genome, UCSC
Hg19
C_Workshop_Test

Upload BED file

Genboree is built & maintained by the Bioinformatics Research Laboratory at Baylor College of Medicine.

B
R

HGSC

MACS in the Genboree Workbench

The screenshot displays the Genboree Workbench interface. At the top, there are tabs for 'Data', 'Analysis', and 'Visualization'. The 'Analysis' tab is active, and a dropdown menu is open, showing various analysis tools. The 'MACS' tool is highlighted under the 'Epigenome Atlas' category. In the background, the 'Data Selector' panel on the left shows a tree view of data sources, with 'h1.h3k4me3.bed.gz' selected under the 'EDACC_Workshop_Test' database. On the right, the 'Details' panel shows metadata for the selected file, and the 'Input Data' and 'Output Targets' panels are visible. The 'Output Targets' panel shows 'EDACC_Workshop_Test' as the target.

Analysis Tools:

- QC Tool:
 - Epigenomic Percentile-Based QC
- Analysis & Search Tools:
 - Spark Analysis
 - Signal Search
 - Signal Comparison
 - LIMMA
 - LIMMA Signal Comparison
 - Epigenomic Heatmap
 - Sample Centric LIMMA
 - Epigenomic Slicer
 - Gene Element Data Slice
- MACS**
- Epigenome Atlas:
 - Epigenome Atlas-Aware Tools

Data Selector:

- genboree.org
 - 1. Scratch-Man
 - cristian_coarfa_group
 - Databases
 - EDACC Workshop , \$
 - EDACC_Workshop_Test
 - All Annotations in Database
 - Tracks
 - Lists & Selections
 - Sample Sets
 - Samples
 - Files
 - BedFiles
 - h1.h3k4me3.bed.gz**
 - Queries
 - Projects
 - Encode Repository

Details:

Attribute	Value
Download	Click to Download File
Group	cristian_coarfa_group
Database	EDACC_Workshop_Test
Description	

Input Data:

- h1.h3k4me3.bed.gz

Output Targets:

- EDACC_Workshop_Test

Invoke MACS

MACS in the Genboree Workbench

The screenshot displays the Genboree Workbench interface. At the top, the 'GENBOREE' logo is visible on the left, and the 'BCM Baylor College of Medicine' logo is on the right. Below the logo, there are tabs for 'Data', 'Analysis', and 'Visualization'. The main content area on the left is titled 'Welcome to the Genboree' and contains a list of instructions. Below this is a 'Data Selector' panel with a 'Refresh' button and a tree view showing the project structure. The tree view includes 'genboree.org', '1. Scratch-Man', 'cristian_coarfa_group', 'Databases', 'EDACC Workshop, \$', 'EDACC_Workshop_Tes', 'All Annotations in D', 'Tracks', 'Lists & Selections', 'Sample Sets', 'Samples', 'Files', 'BedFiles', 'h1.h3k4me3.b', 'Queries', 'Projects', and 'Encode Repository'. A 'Tool Settings' dialog box is open in the center, showing the 'MACS Settings' tab. The dialog box has a 'Database' field set to 'EDACC_Workshop_Test' and a 'Group' field set to 'cristian_coarfa_group'. The 'MACS Settings' section includes fields for 'Analysis Name' (MACS-2012-3-6-9:13:59), 'Results Name' (MACS-2012-3-6-9:13:59), 'P Value' (1e-5), 'M Fold' (10.30), 'Upload Results ?' (checked), 'Track Name' (H1 H3K4me3 : Peaks), 'No Lambda' (unchecked), 'Small Local' (1000), 'Large Local' (10000), 'Off Auto' (unchecked), and 'No Model' (unchecked). At the bottom of the dialog box are 'Submit' and 'Cancel' buttons. The background interface is partially obscured by the dialog box.

MACS settings

MACS in the Genboree Workbench

The screenshot displays the Genboree Workbench interface. At the top, the 'GENBOREE' logo is on the left, and the 'BCM Baylor College of Medicine' logo is on the right. Below the logos is a navigation bar with 'Data', 'Analysis', and 'Visualization' tabs. A 'Welcome to the Genboree Workbench!' message is visible, followed by a list of instructions for using the Data Selector and Analysis tools. A 'Data Selector' panel on the left shows a tree view of data sources, including 'genboree.org', '1. Scratch-Man', 'cristian_coarfa_group', 'Databases', 'EDACC Workshop', 'EDACC_Workshop_Test', 'All Annotations in Database', 'Tracks', 'Lists & Selections', 'Sample Sets', 'Samples', 'Files', 'BedFiles', 'h1.h3k4me3.bed.gz', 'Queries', and 'Projects'. A 'Job Submission Status' dialog box is centered on the screen, titled 'MACS'. It contains a green checkmark icon and the text: 'Job Id: wbJob-macs-1331046996_001544', 'Your job has been successfully submitted. You will be notified by email when your job has completed.', and 'If you have questions, please contact genboree_admin@genboree.org for assistance.' An 'OK' button is at the bottom of the dialog. In the background, the 'Input Data' and 'Output Targets' panels are visible. The 'Input Data' panel shows 'h1.h3k4me3.bed.gz' and the 'Output Targets' panel shows 'EDACC_Workshop_Test'.

GENBOREE

BCM
Baylor College of Medicine

Data Analysis Visualization

Welcome to the Genboree Workbench!

- The **Data Selector** tree on the left contains all the data sources available in the workbench.
- Drag items to be used as **Input Data** for your analysis.
- Drag items to be used as **Output Targets** for your analysis.
- Tools which can be run on your selected data are listed in the **Analysis** tab.
- Unsure about what kinds of jobs you can run? Just click the tool button in the **Analysis** tab.

Data Selector

Refresh

genboree.org

- 1. Scratch-Man
- cristian_coarfa_group
 - Databases
 - EDACC Workshop, \$
 - EDACC_Workshop_Test
 - All Annotations in Database
 - Tracks
 - Lists & Selections
 - Sample Sets
 - Samples
 - Files
 - BedFiles
 - h1.h3k4me3.bed.gz
 - Queries
 - Projects

Job Submission Status

MACS

Job Id: wbJob-macs-1331046996_001544

Your job has been successfully submitted. You will be notified by email when your job has completed.

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

OK

Input Data

h1.h3k4me3.bed.gz

Output Targets

EDACC_Workshop_Test

MACS in the Genboree Workbench

Welcome to the Genboree Workbench!

- The **Data Selector** tree on the left shows the data entities to which you have access.
- Drag items to be used as tool *inputs* over to the **Input Data** area.
- Drag items to be used as *output destinations* for tool results over to the **Output Targets** area.
- Tools which can be run on your selections will be highlighted in **green**.
- Unsure about what kinds of items a particular tool needs in the **Input Data** and **Output Targets** ??
 - Just click the tool button when it is *not highlighted* to see help information.

Data Selector

Refresh Data Filter: Select a filter...

- Databases
 - EDACC Workshop, \$
 - EDACC_Workshop_Test
 - All Annotations in Database
 - Tracks
 - Lists & Selections
 - SampleSets
 - Samples
 - Files
 - MACS
 - MACS-2012-3-6-10:00:49
 - H1 H3K4me3:Peaks.lff.zip
 - MACS-2012-3-6-10:00:49_peaks.xls
 - raw
 - jobFile.json
 - MACS-2012-3-6-9:48:34
 - MACS-2012-3-6-9:49:56
 - MACS-2012-3-6-9:43:56

Details

Attribute	Value
Download	Click to Download File
Group	cristian_coarfa_group
Database	EDACC_Workshop_Test
Description	

Input Data

⬆ ⬇ ✖

Output Targets

⬆ ⬇ ✖

MACS results

MACS in the Genboree Workbench

R20										
	A	B	C	D	E	F	G	H	I	J
1	# This file is generated by MACS version 1.4.1 20110627									
2	# ARGUMENTS LIST:									
3	# name = MACS-2012-3-6-10%3A00%3A49									
4	# format = AUTO									
5	# ChIP-seq file = ./9b73a7c6c9b64f701f6028f5be1da83c3f353421/file_0_BedFiles%2Fh1.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 190 bps									
15	# total tags in treatment: 2780585									
16	# tags after filtering in treatment: 2780585									
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	#NAME?	fold_enrichment		
21	chr1	2	600	599	208	20	181.37	35.71		
22	chr1	703123	704011	889	706	16	71.02	13.51		
23	chr1	704346	705325	980	260	18	77.65	11.84		
24	chr1	713667	715462	1796	1291	34	63.98	7.53		
25	chr1	715635	716998	1364	1181	33	87.14	8.22		
26	chr1	751824	753341	1518	367	33	153.52	14.86		
27	chr1	833597	834000	404	176	8	67.08	17.5		
28	chr1	883526	884756	1231	495	58	286.31	17.14		
29	chr1	938155	938782	628	431	11	73.41	13.04		
30	chr1	944708	945626	919	440	24	141.22	20.59		

MACS

MACS-2012-3-6-10_00_49_peaks

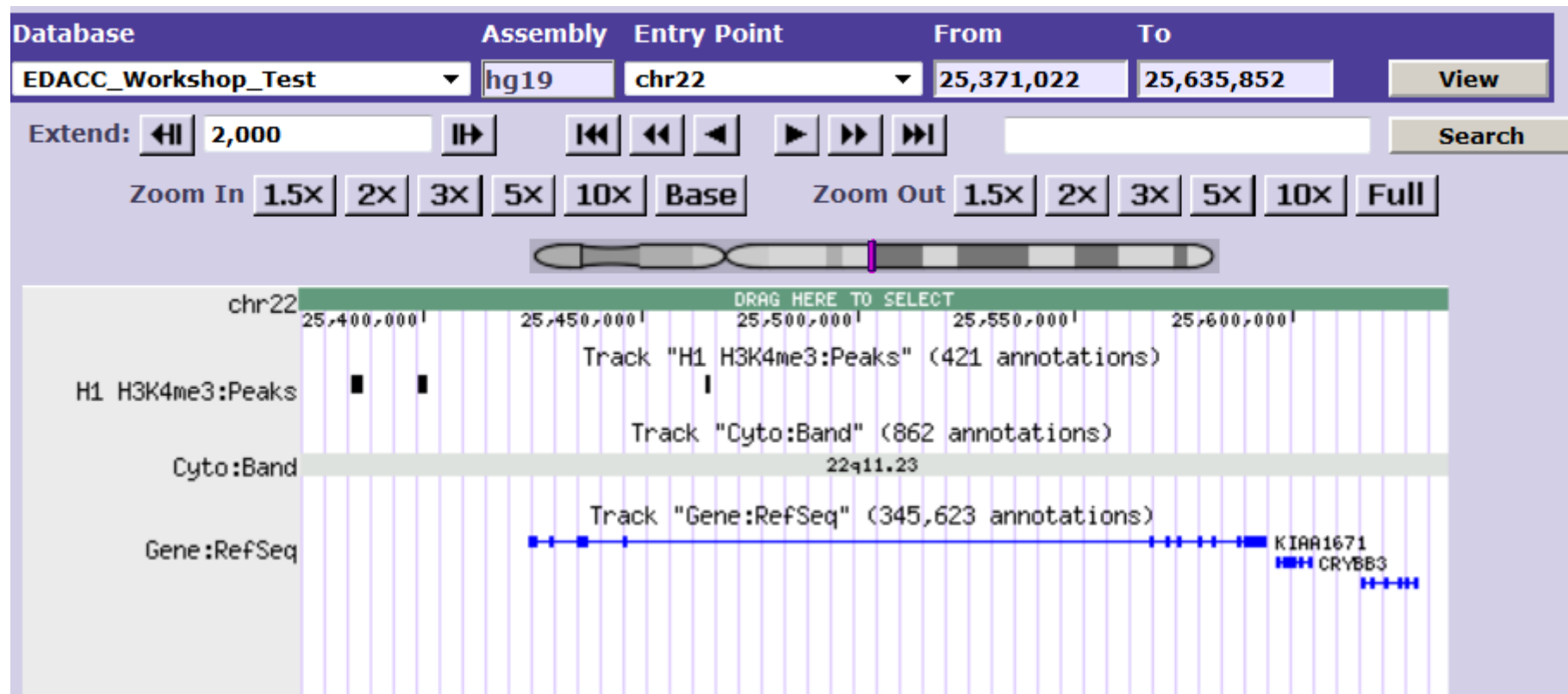
Ready

MACS results

MACS-2012-3-6-10_00_49_peaks

Ready

Results uploaded to Genboree



Coming attractions

- Peak interpretation
 - Gene promoters, miRNA promoters
 - Gene elements: introns, exons, 3'UTR, 5'UTR
 - Repeats
 - User defined

mRNA-Seq

- Deployed
 - TopHat
 - Cufflinks
 - Cuffdiff
 - FastQC
- Results usable in resequencing pipeline

TopHat in the Genboree Workbench

The screenshot displays the Genboree Workbench interface. At the top, there are tabs for 'Data', 'Analysis', and 'Visualization'. The 'Analysis' tab is active, showing a sidebar with categories like Epigenomics, Track Tools, Small RNA, Microbiome Workbench, SNPs, Cancer Analysis Workbench, and RNA-Seq. The 'RNA-Seq' category is expanded, revealing a list of tools: TopHat, Cufflinks, Cuffdiff, and FastQC. The 'TopHat' tool is selected, and a tooltip provides details: 'TopHat is a fast splice junction mapper for RNA-Seq reads. It aligns RNA-Seq reads to mammalian-sized genomes using the ultra high-throughput short read aligner Bowtie, and then analyzes the mapping results to identify splice junctions between exons.'

On the left, the 'Data Selector' panel shows a tree view of data entities. Under 'Files', the 'TopHat' folder is expanded, listing several fastq.gz files. The file 'BT474.1.fastq.gz' is highlighted.

On the right, the 'Details' panel shows a table with columns 'Value' and 'Description'. The table contains three rows: 'Click to Download File', 'BCM-MDA SOLID Matepairs', and 'JoeGray_BC_RNA-Seq'. Below this, the 'Input Data' section shows a list of input files: 'BT474.1.fastq.gz' and 'BT474.2.fastq.gz'. The 'Output Targets' section shows a list of output targets: 'JoeGray_BC_RNA-Seq'.

TopHat

TopHat is a fast splice junction mapper for RNA-Seq reads. It aligns RNA-Seq reads to mammalian-sized genomes using the ultra high-throughput short read aligner Bowtie, and then analyzes the mapping results to identify splice junctions between exons.

Value	Description
Click to Download File	
BCM-MDA SOLID Matepairs	
JoeGray_BC_RNA-Seq	

Input Data

- BT474.1.fastq.gz
- BT474.2.fastq.gz

Output Targets

- JoeGray_BC_RNA-Seq

TopHat in the Genboree Workbench

Tool Settings

Output Location:

Database: JoeGray_BC_RNA-Seq Group: BCM-MDA SOLID Matepairs

MACS Settings

Analysis Name: TopHat-2012-3-6-10:01:34

Upload Results ? ☒

Track Name: MyRNA-Seq : Tags

Auto-Determine Mate Inner Dist ? ☒

Mate Stdev: 20

Min Anchor Length: 8

Splice Mismatches: 0

Min Intron Length: 70

Max Intron Length: 70

Max Insertion Length: 3

Max Deletion Length: 3

Closure search ? ☐

Coverage Search ? ☐

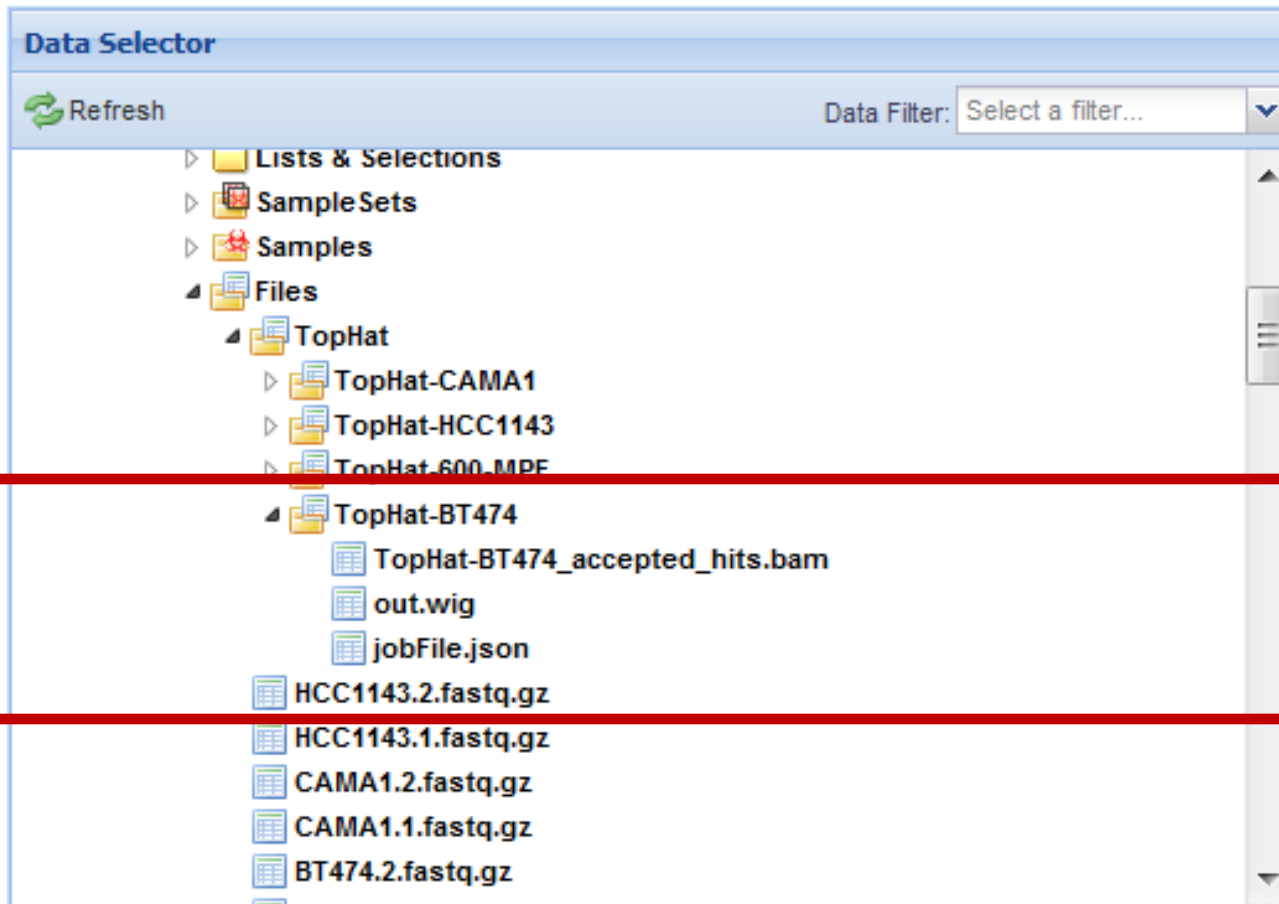
Butterfly search ? ☐

+ Advanced Settings:

Submit Cancel

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

TopHat in the Genboree Workbench



TopHat in the Genboree Workbench



Coming attractions

- Alternative splicing
- Intron retention
- Gene fusions
- Current research collaborations
 - Ovarian cancer
 - Novel fusion with 20% recurrence in tumors

Resequencing

- “Genome in a day”
 - HiSeq 2500/1500
 - 40x coverage
- Whole genome: affordable
 - Less than \$5000
- Whole exome: very attractive
 - Heterogeneous “tumor” samples
- SNP/Indel caller: Atlas-SNP2

Atlas-SNP2



IMPACT
FACTOR
3.03

Search for

Home

Articles

Authors

Reviewers

About this journal

My BMC Bioinformatics

Top

Abstract

Background

Implementation

Results

Conclusions

Software

Highly accessed

Open access

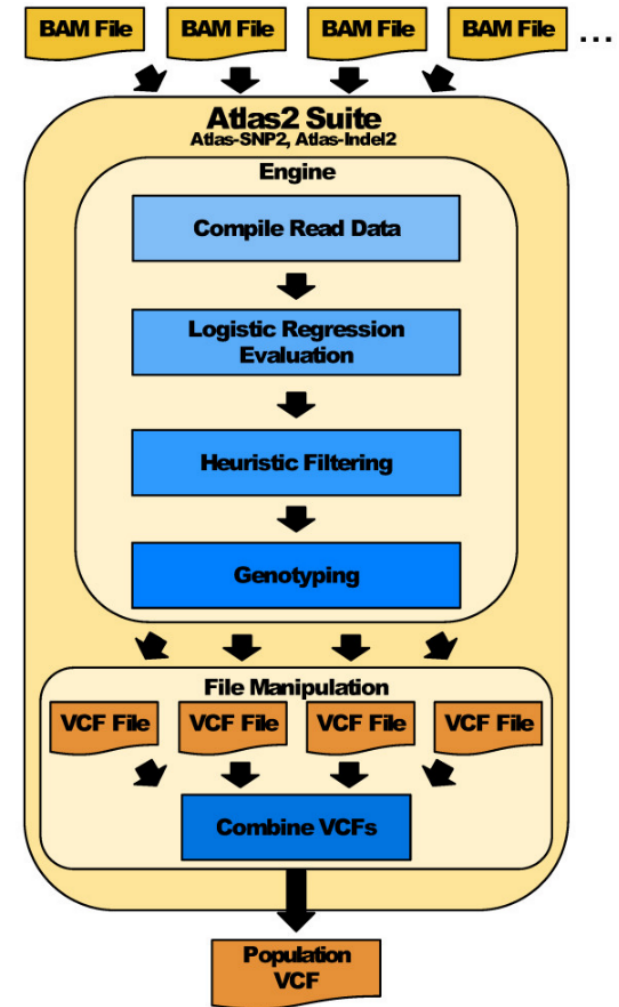
An integrative variant analysis suite for whole exome next-generation sequencing data

Danny Challis^{1†}, Jin Yu^{1†}, Uday S Evani¹, Andrew R Jackson², Sameer Paithankar², Cristian Coarfa², Aleksandar Milosavljevic^{2*}, Richard A Gibbs^{1, 2*} and Fuli Yu^{1, 2*}

Over 4000 accesses since publication

Atlas-SNP2

- Linear regression model accounting for sources of sequencing biases and errors
- Trained for individual sequencing platforms and chemistry
- Models for 454, Illumina, SOLID
- Constantly updated




$$\text{logit}(p) = -19.18 - 0.4779 \times \text{ref_var_ratio} + 0.2368 \times \text{mean_NBQ} + 4.520 \times \text{strand_dir}$$

$$+ 0.5436 \times \text{mean_dist3'} + 0.08243 \times \text{mean_VBQ} - 0.00796 \times \text{mean_NBQ} \times \text{mean_dist_3'}$$

$$- 0.1122 \times \text{strand_dir} \times \text{mean_dist_3'}$$

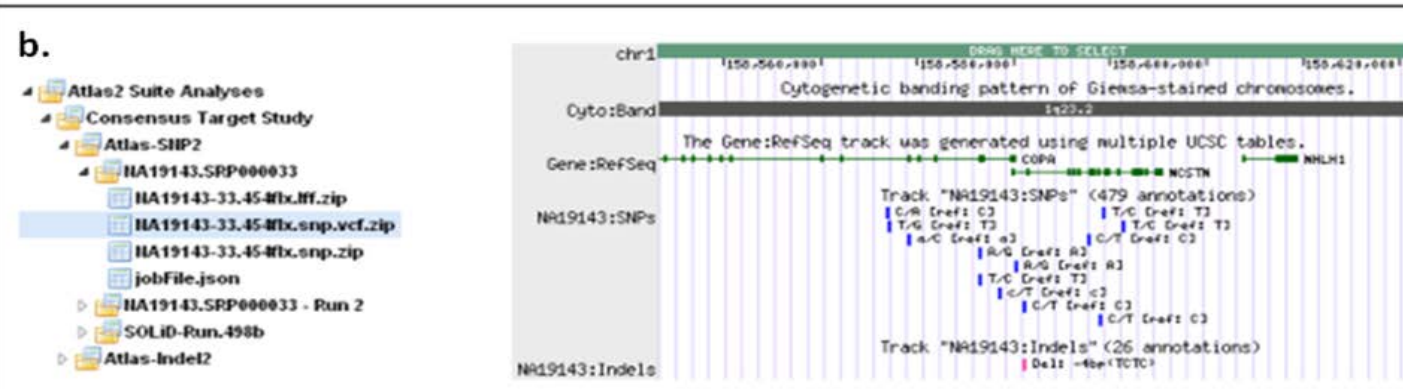
Atlas-SNP2 and Genboree

a.



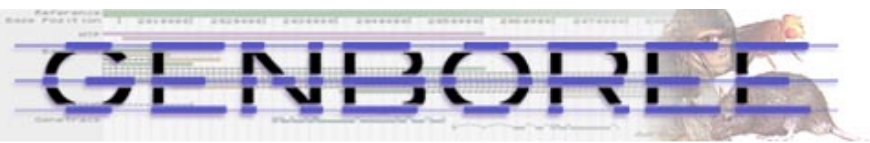
The screenshot shows the Atlas-SNP2 web interface. On the left, the 'Data Selector' panel displays a tree view of data sources: '1000-Genomes', 'Analyses DB', 'Tracks', 'Files', 'BAM & SAM Files', and 'Atlas2 Suite Analyses'. The 'BAM & SAM Files' section is expanded, showing files like 'NA19143.SLX.SRP000033.bam'. The 'Tool Settings' panel on the right is titled 'Atlas-SNP2 Settings' and contains fields for 'Study Name' (Consensus Target Study), 'Job Name' (NA19143.SRP000033), 'Platform Type' (454 (GS FLX)), 'Sample Name' (NA19143), and 'Upload as a Track?' (checked). The 'Advanced Settings' section includes sliders for 'Min. Coverage' (8), 'Max. Pile-Up' (200), 'Max. % Substitutions' (5.0), 'Max. % Indels' (5.0), 'Posterior Prob. Cutoff' (0.95), 'Prior Prob. for Coverage > 2' (0.1), 'Prior Prob. for Coverage <= 2' (0.9), and 'Insert Size' (0). 'Submit' and 'Cancel' buttons are at the bottom.

b.



The screenshot shows the Genboree web interface. On the left, the 'Atlas2 Suite Analyses' panel is expanded, showing 'Consensus Target Study' and 'Atlas-SNP2'. The 'Atlas-SNP2' section is expanded, showing files like 'NA19143.SRP000033', 'NA19143-33.45-flx.bff.zip', 'NA19143-33.45-flx.snp.vcf.zip', 'NA19143-33.45-flx.snp.zip', 'jobFile.json', 'NA19143.SRP000033 - Run 2', 'SOLID-Run.498b', and 'Atlas-Indel2'. The main panel displays a genomic track view for chromosome 1 (chr1). The tracks include 'Cyto:Band' (Cytogenetic banding pattern), 'Gene:RefSeq' (Gene/RefSeq track), 'NA19143:SNPs' (Track 'NA19143:SNPs' (479 annotations)), and 'NA19143:Indels' (Track 'NA19143:Indels' (26 annotations)). The 'NA19143:SNPs' track shows various SNPs with their positions and annotations. The 'NA19143:Indels' track shows indels with their positions and annotations.

Atlas-SNP2 in the Genboree Workbench



Data ▾

Analysis ▾

Visualization ▾

Welcom

- The
- Drag
- Drag
- Tools
- Unsu
 - Ju

Epigenomics

Track Tools

Small RNA

Microbiome Workbench

SNPs

Cancer Analysis Workbench

RNA-Seq

Atlas2 Suite Tools - SNP Calling:

Atlas-SNP2

Atlas-Indel2

Atlas2 Suite Tools - Re-Genotyping:

Atlas-SNP2 Re-Genotyper

Workbench!

shows the data entities to which you have access.
over to the **Input Data** area.
tinations for tool results over to the **Output Targets** area.
times will be highlighted in **green**.
Input Data and **Output Targets** ??
information.

Data Selector

Refresh

All Annotations in Database

- Tracks
- Lists & Selections
- Sample Sets
- Samples
- Files
 - TopHat
 - TopHat-CAMA1
 - TopHat-HCC1143
 - TopHat-600-MPE
 - TopHat-BT474
 - TopHat-BT474_accepted_hits.bam
 - out.wig
 - jobFile.json
 - HCC1143.2.fastq.gz
 - HCC1143.1.fastq.gz
 - CAMA1.2.fastq.gz

Details

Attribute	Value
Download	Click to Download File
Group	BCM-MDA SOLID Matepairs
Database	JoeGray_BC_RNA-Seq
Description	

Input Data

⬆ ⬇ ✖

TopHat-BT474_accepted_hits.bam

Output Targets

⬆ ⬇ ✖

TestAtlasSNP2

Atlas-SNP2 in the Genboree Workbench

will appear as a column header in the VCF output file.

The Advanced Settings section contains parameters you can use to tweak prior probabilities and alignment filters. Currently, the defaults are set for all these options according to the selected platform.

Additional details are available in the Help dialog.

[Atlas-SNP2 is described in [this journal article.](#)]

Input Data:

Data File: TopHat/TopHat-BT474/TopHat-BT474_accepted_hits.bam
Group: BCM-MDA SOLID Matepairs,
Database: JoeGray_BC_RNA-Seq

Output Location:

Database: TestAtlasSNP2
Group: ccristi_group

Atlas-SNP2 Settings

Study Name AtlasTools-Study-2012-3-6

Job Name AtlasSNP2-Job-2012-3-6-1

Platform Illumina

Sample Name 454 (GS FLX)
454 (Titanium)
Illumina
SOLiD

Upload as a Track ?

SNPs Track Name SNPs

Advanced Settings:

Submit **Cancel**

Atlas-SNP2 in the Genboree Workbench

The screenshot displays the Genboree Workbench interface. At the top, there are tabs for 'Data', 'Analysis', and 'Visualization'. Below these is a 'Welcome to the Genboree Workbench!' message with instructions on how to use the interface. On the left, the 'Data Selector' panel shows a tree of data entities, including 'All Annotations in Data', 'Tracks', 'Lists & Selections', 'SampleSets', 'Samples', and 'Files'. Under 'Files', there are sub-folders for 'TopHat' and 'TopHat-BT474', with files like 'TopHat-BT474_accepted_hits.bam', 'out.wig', 'jobFile.json', 'HCC1143.2.fastq.gz', 'HCC1143.1.fastq.gz', and 'CAMA1.2.fastq.gz'. In the center, a 'Job Submission Status' dialog box for 'Atlas-SNP2' is open, showing a green checkmark and the message: 'Job Id: vbJob-atlassnp2-1331056634_943736. Your job has been successfully submitted. You will be notified by email when your job has completed. If you have questions, please contact genboree_admin@genboree.org for assistance.' The dialog has an 'OK' button. On the right, the 'Output Targets' panel shows a list of targets, including 'TestAtlas SNP2'. At the bottom, there is a footer with the Genboree logo, copyright information (© 2001-2012 Bioinformatics Research Laboratory), and the HGSC logo (Human Genome Sequencing Center).

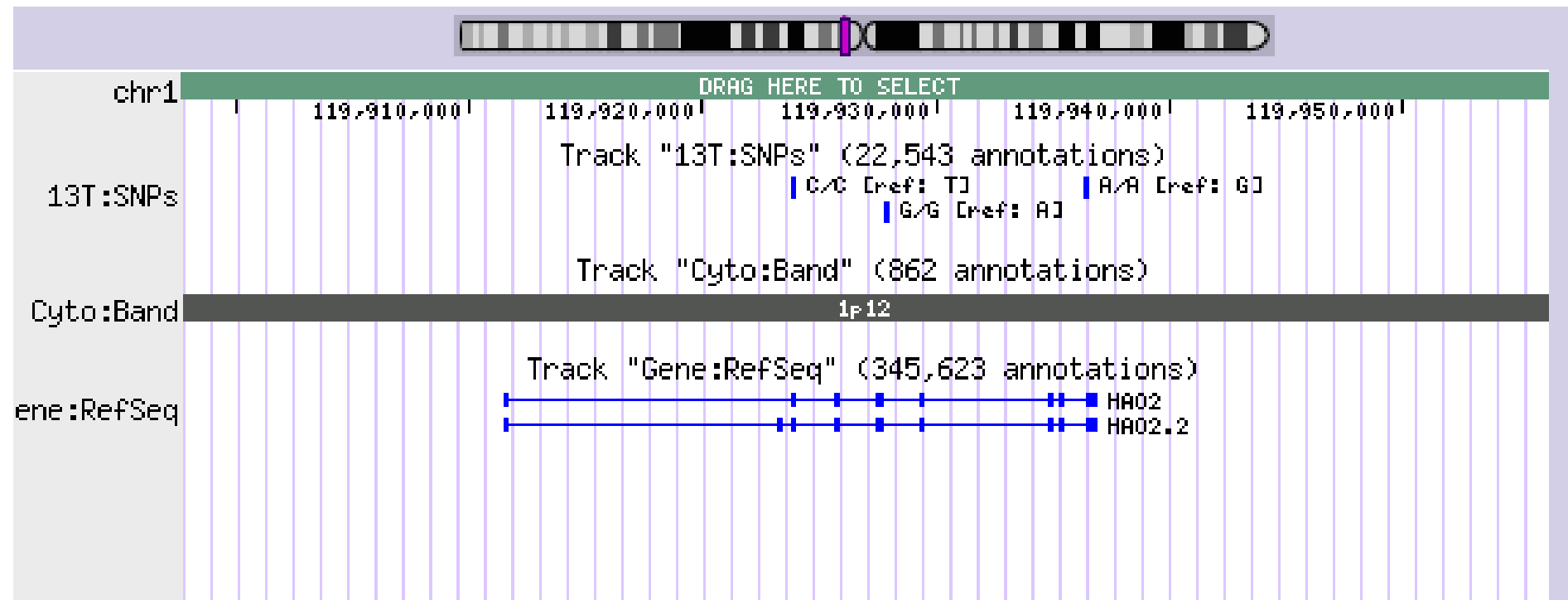
© 2001-2012
Bioinformatics Research Laboratory
4000 South West Ave. MS-BM335

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.**

HGSC
HUMAN GENOME SEQUENCING CENTER

Atlas-SNP2 in the Genboree Workbench



Atlas-SNP2 in the Genboree Workbench

G1		f _x		FILTER							
	A	B	C	D	E	F	G	H	I	J	
1	#CHROI	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMA	My RNA-Seq sample	
26	chr1	14653	.	C	T	60	PASS	.	GT:VR:RR:	0/1:37:71:108:.	
31	chr1	14677	.	G	A	60	PASS	.	GT:VR:RR:	0/1:37:97:134:.	
45	chr1	14907	.	A	G	60	PASS	.	GT:VR:RR:	0/1:4:6:10:.	
46	chr1	14930	.	A	G	60	PASS	.	GT:VR:RR:	0/1:5:1:6:.	
59	chr1	15211	.	T	G	60	PASS	.	GT:VR:RR:	0/1:8:3:11:.	
60	chr1	15274	.	a	T	60	PASS	.	GT:VR:RR:	1/1:9:0:9:.	
64	chr1	15696	.	A	T	60	PASS	.	GT:VR:RR:	0/1:3:12:15:.	
68	chr1	16378	.	T	C	60	PASS	.	GT:VR:RR:	0/1:4:2:6:.	
71	chr1	16606	.	C	A	60	PASS	.	GT:VR:RR:	0/1:3:5:8:.	
85	chr1	16963	.	G	A	60	PASS	.	GT:VR:RR:	0/1:34:131:165:.	
89	chr1	16996	.	T	C	60	PASS	.	GT:VR:RR:	0/1:67:110:177:.	
119	chr1	18064	.	G	T	60	PASS	.	GT:VR:RR:	0/1:6:44:50:.	
122	chr1	18200	.	A	G	60	PASS	.	GT:VR:RR:	0/1:3:7:10:.	
130	chr1	18643	.	G	A	60	PASS	.	GT:VR:RR:	0/1:3:13:16:.	
153	chr1	20250	.	t	C	60	PASS	.	GT:VR:RR:	0/1:4:3:7:.	
163	chr1	20813	.	A	G	60	PASS	.	GT:VR:RR:	0/1:3:25:28:.	
199	chr1	23647	.	T	G	60	PASS	.	GT:VR:RR:	0/1:4:8:12:.	
237	chr1	136781	.	T	C	60	PASS	.	GT:VR:RR:	1/1:4:0:4:.	
238	chr1	136830	.	T	C	60	PASS	.	GT:VR:RR:	0/1:4:3:7:.	

Coming Attraction

- SNP annotation
 - Annovar
 - Known SNPs : dbSNP, 1000 genomes
 - Genic/intergenic
 - Coding/non-coding variants
 - GWAS
 - Regulatory regions

Acknowledgments

- NIH Epigenomic Roadmap
 - Lisa Chadwick, Fred Tyson, Joni Rutter, John Satterlee, Astrid Haugen, Kim McAllister
- Peak Calling
 - Anshul Kundaje, Bob Thurman (UW), Noam Shores (BI), Martin Hirst (BCGSC), Lee Daniels (NIH), Wei Li (BCM)
- RNA-Seq
 - Adrian Lee, Joe Gray, Laising Yen, Kalpana Kannan
- SNP calling, resequencing
 - Fuli Yu, Richard Gibbs, Danny Challis, Jin Yu, Uday Evani
- BRL
 - Cristian Coarfa, Alan R Harris, Chia-Chin Wu, Aleksandar Milosavljevic
- BRL core
 - Matt Roth, Kevin Riehle
- Genboree programming team
 - Andrew Jackson, Sameer Paithankar, Sriram Raghuram,

Questions ?

Comments ?

Suggestions ?

amilosav@bcm.edu

coarfa@bcm.edu