

# Chip – Seq Peak Calling in Galaxy

Lisa Stubbs

# Introduction

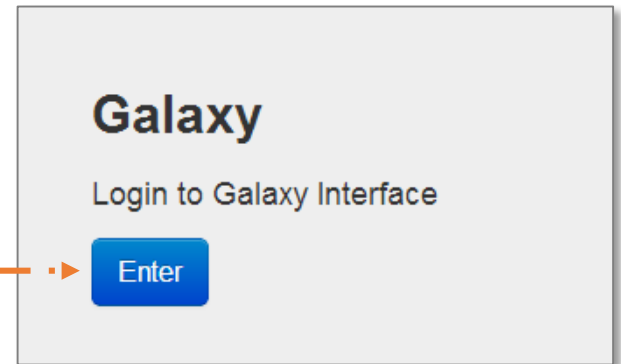
The goals of the lab are as follows:

1. Gain experience using Galaxy.
2. Teach how to map next generation (NSG) reads to a reference genome using **Bowtie**.
3. Demonstrate how to call peaks from Chip-Seq data.

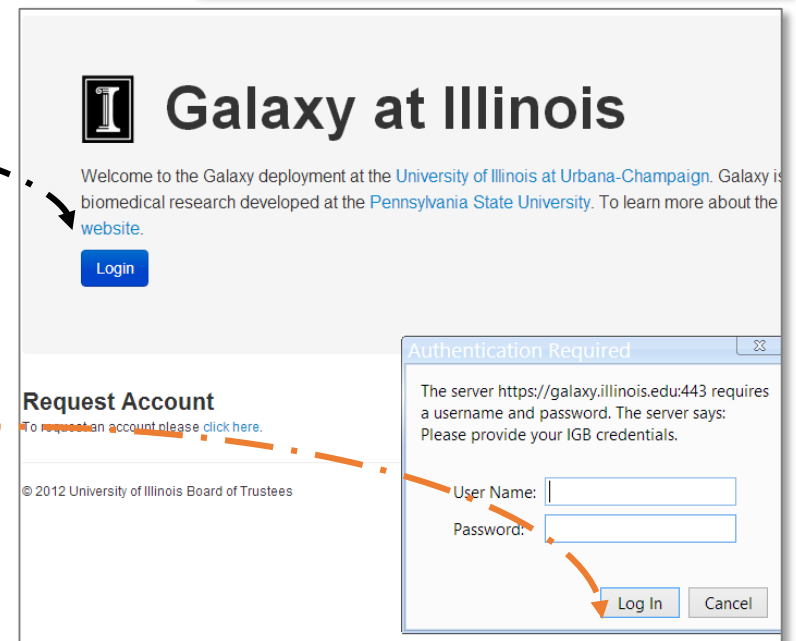
# Step 1A: Logging into Galaxy

Go to: **biocluster.igb.illinois.edu**

Click Enter



Click Login

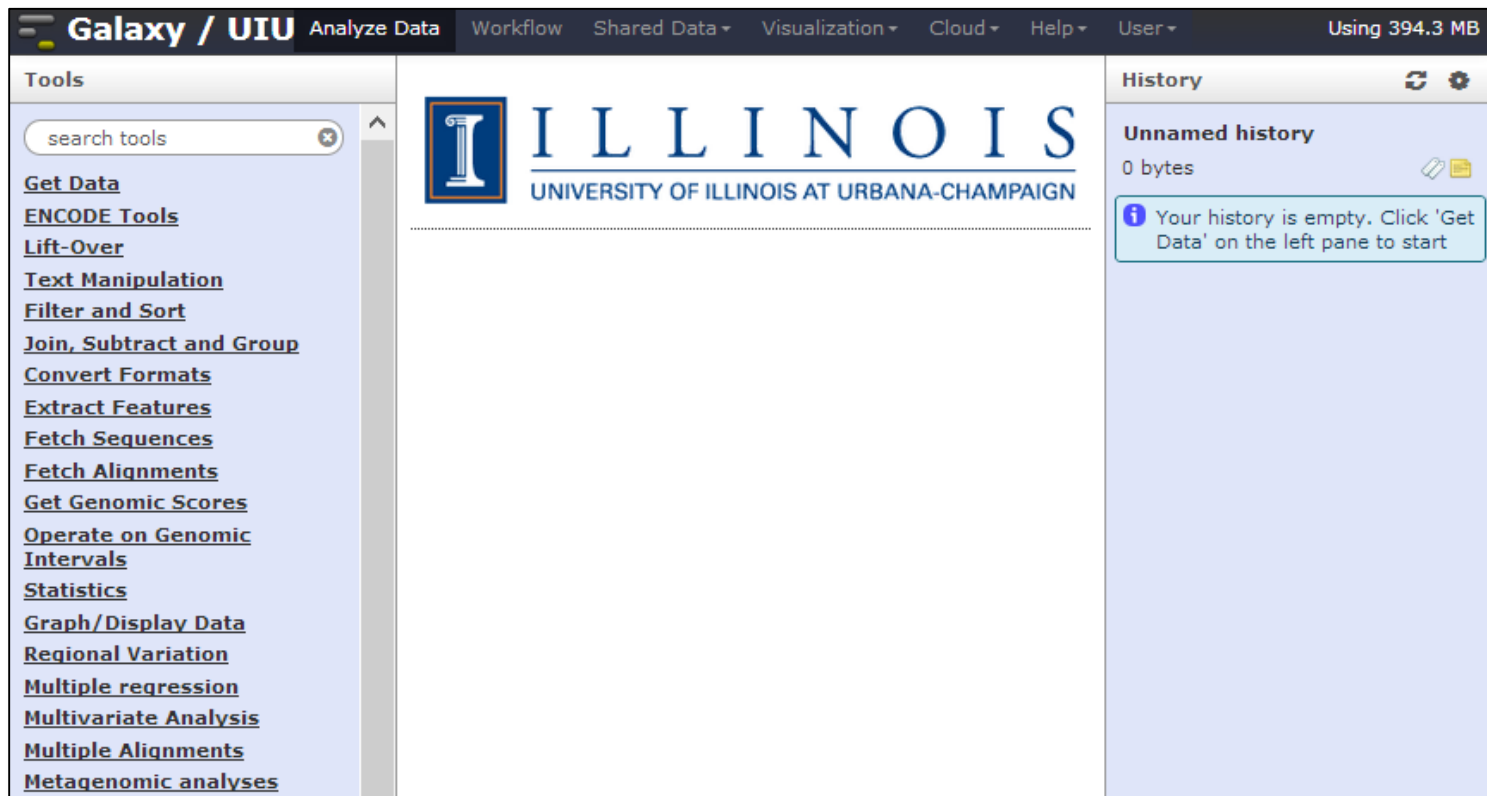


Input your login credentials.

Click Login.

# Step 1B: Galaxy Start Screen

The resulting screen should look like the **figure** below:



# Step 2A: Importing the Data

In this step, we will import the following data files:

Filename	Meaning
G1E_ER4_CTCF_(chr19).fastqsanger	A sample ChIP-seq dataset on CTCF in G1E_ER4 cells, reads have been reduced to those mapping to chr19 for demonstration use.
G1E_ER4_input_(chr19).fastqsanger	Control DNA taken from chr19.
G1E_CTCF.fastqsanger	CTCF Chip for G1E line.
G1E_input.fastqsanger	Control for G1E line.

**Note:** G1E cell lines are erythroid, red blood cell, cell lines missing the GATA-1 gene.

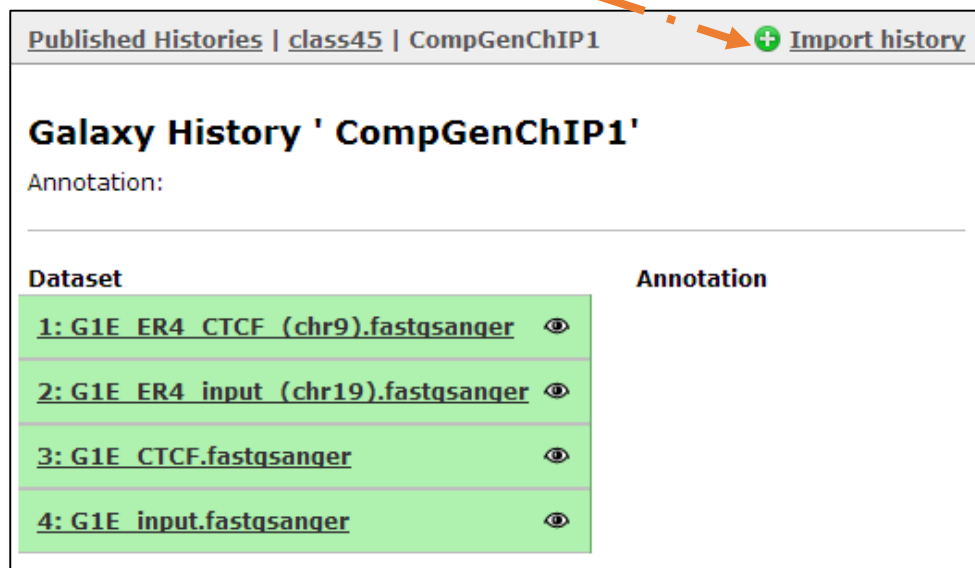
GATA-1 is crucial for the maturation of erythroid cells.

G1E\_E4R cell lines conditionally express GATA-1 in the presence of estradiol, enabling erythroid maturation.

## Step 2B: Import Data in Galaxy

Go to <https://galaxy.illinois.edu/galaxy/u/instr04/h/compgenchip1>

Click **Import History** and on the next page click **Start Using History**.




Published Histories | [class45](#) | CompGenChIP1 [+ Import history](#)

### Galaxy History ' CompGenChIP1 '

Annotation:

Dataset	Annotation
<a href="#">1: G1E ER4 CTCF (chr9).fastqsanger</a>	
<a href="#">2: G1E ER4 input (chr19).fastqsanger</a>	
<a href="#">3: G1E CTCF.fastqsanger</a>	
<a href="#">4: G1E input.fastqsanger</a>	

 History "imported: chip\_seq\_peak\_call" has been imported.  
You can [start using this history](#) or [return to the previous page](#).

# Step 2C: Import Data into Galaxy

Your Galaxy page should look like the following now:

The screenshot displays the Galaxy web interface with a dark blue header bar. The header contains the 'Galaxy / U' logo, followed by 'Analyze Data' and several menu items: 'Workflow', 'Shared Data', 'Visualization', 'Cloud', 'Help', and 'User'. The interface is divided into three main sections. On the left is the 'Tools' pane, which includes a search bar and a list of tool categories: 'Get Data', 'ENCODE Tools', 'Lift-Over', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Formats', 'Extract Features', 'Fetch Sequences', 'Fetch Alignments', 'Get Genomic Scores', 'Operate on Genomic Intervals', and 'Statistics'. The center is the 'Main Pane', which features the University of Illinois at Urbana-Champaign logo and a large empty workspace. On the right is the 'History' pane, which lists recent jobs. The top job is 'imported: chip\_seq\_peak\_call' with a size of 103.9 MB. Below it are four other jobs, each with a number, a description, and icons for viewing, editing, and deleting. The jobs are: '4: G1E input.fastqsanger', '3: G1E CTCF.fastqsanger', '2: G1E ER4 input (chr19).fastqsanger', and '1: G1E ER4 CTCF (chr9).fastqsanger'.

Tools	Main Pane	History
search tools	ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN	imported: chip_seq_peak_call 103.9 MB
Get Data		4: G1E input.fastqsanger
ENCODE Tools		3: G1E CTCF.fastqsanger
Lift-Over		2: G1E ER4 input (chr19).fastqsanger
Text Manipulation		1: G1E ER4 CTCF (chr9).fastqsanger
Filter and Sort		
Join, Subtract and Group		
Convert Formats		
Extract Features		
Fetch Sequences		
Fetch Alignments		
Get Genomic Scores		
Operate on Genomic Intervals		
Statistics		

Tools Pane

Main Pane

History Pane

# Read Mapping and Peak Calling

In this exercise, we will map ChIP Reads to a reference genome and call peaks among the mapped reads using **MACs**.



# Step 3A: Summary Statistics

In this step, we will gather summary statistics of ChIP data for quality control.

Click **NGS: QC and manipulation** from the **Tools** pane.

Then click **FASTQ Summary Statistics**.

## NGS: QC and manipulation

### **FASTQC: FASTQ/SAM/BAM**

- [FastQC:Read QC](#) reports using FastQC

### **ILLUMINA FASTQ**

- [FASTQ Groomer](#) convert between various FASTQ quality formats
- [FASTQ splitter](#) on joined paired end reads
- [FASTQ joiner](#) on paired end reads
- [FASTQ Summary Statistics](#) by column

## Statistics

### Graph/Display Data

### Regional Variation

### Multiple regression

### Multivariate Analysis

### Multiple Alignments

### Metagenomic analyses

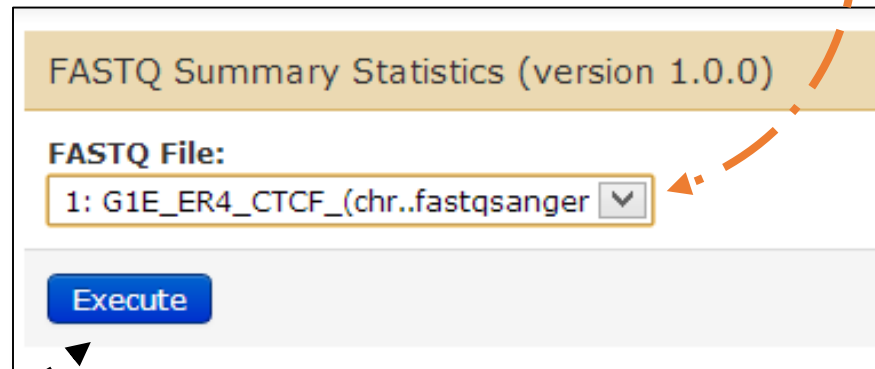
### FASTA manipulation

### NGS: QC and manipulation

## Step 3B: FASTQ Summary Statistics

On the next page, make sure **1: G1E\_ER4\_CTCF\_(chr..fastqsanger)** is selected.

Press **Execute**.



FASTQ Summary Statistics (version 1.0.0)

**FASTQ File:**

1: G1E\_ER4\_CTCF\_(chr..fastqsanger ▼

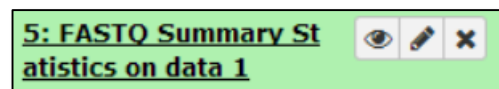
Execute

The screenshot shows the 'FASTQ Summary Statistics' tool interface. A yellow header bar contains the title. Below it, the 'FASTQ File:' label is followed by a dropdown menu showing '1: G1E\_ER4\_CTCF\_(chr..fastqsanger' with a downward arrow. At the bottom is a blue 'Execute' button. An orange dashed arrow points from the text '1: G1E\_ER4\_CTCF\_(chr..fastqsanger)' in the paragraph above to the dropdown menu. A black dashed arrow points from the text 'Press Execute.' to the 'Execute' button.

# Step 3C: FASTQ Summary Statistics

The summary file will be the 5<sup>th</sup> file in the History pane.

Click  to display the file in the **Main** pane.



#column	count	min	max	sum	mean	Q1	med	Q3	IQR	IW	rW
1	270631	2	33	8498504	31.4025518141	32.0	33.0	33.0	1.0	31	33
2	270631	2	34	8324960	30.7612948997	30.0	33.0	33.0	3.0	26	34
3	270631	2	34	8181664	30.231806408	29.0	32.0	33.0	4.0	23	34
4	270631	2	34	8184981	30.2440629492	29.0	32.0	33.0	4.0	23	34
5	270631	2	34	8161333	30.1566819766	29.0	32.0	33.0	4.0	23	34
36	270631	2	34	7244057	26.7672846052	25.0	30.0	32.0	7.0	17	34

## Discussion

How long are these reads?

What is the median quality at the last position?

## Step 4A: Map ChIP-Seq Reads to MM9 Genome

Next, we will map the reads in `G1E_E4R_CTCF_(chr9).fastqsanger` to the mouse genome.

Select NGS: Mapping — . — . — . — . — . — . →

Then select **Map with Bowtie for Illumina**

- Lastz paired reads map short paired reads against reference sequence
- Lastz map short reads against reference sequence
- Map with Bowtie for SOLiD
- Map with Bowtie for Illumina
- Map with BWA for Illumina
- Map with BWA for SOLiD

# Step 4B: Map ChIP-Seq Reads to MM9 Genome

Make sure to select **mm9** as the reference genome.

Make sure **1: G1E\_ER4\_CTCF\_(chr..fastqsanger)** is selected.

Hit **Execute**.

It will take a few moments to complete.

When done, click the view icon.



**7: Map with Bowtie for  
Illumina on data 1: mapped reads**

Map with Bowtie for Illumina (version 1.1.2)

Will you select a reference genome from your history or use a built-in index?:  
☐ Use a built-in index ☒ Use a reference genome from your history or use a built-in index  
Built-ins were indexed using default options

Select a reference genome:  
**mm9 (UCSC,full)**  
if your genome of interest is not listed - contact Galaxy team

Is this library mate-paired?:  
☐ Single-end ☒ Paired-end

FASTQ file:  
**1: G1E\_ER4\_CTCF\_(chr..fastqsanger)**  
Must have ASCII encoded quality scores

Bowtie settings to use:  
☒ Commonly used ☐ Full parameter list  
For most mapping needs use Commonly used settings. If you want full control use Full parameter list

Suppress the header in the output SAM file:  
☐  
Bowtie produces SAM with several lines of header information by default

**Execute**

# Step 4C: Map ChIP-Seq Reads to MM9 Genome

Your **Main** pane and **History** pane should look like the following:

The screenshot displays the Galaxy web interface. The main pane on the left shows a list of sequencing reads in SAM format, including header lines for each chromosome (chr8 to chrM) and a table of mapped reads for chromosome 19. The right pane shows the '6: Map with Bowtie f or Illumina on data 1' tool, which has successfully aligned the sequence file. Below this, a 'FASTQ Summary' section provides statistics on the data.

**Main Pane (Left):**

```
@SQ SN:chr8 LN:131738871
@SQ SN:chr9 LN:124076172
@SQ SN:chr10 LN:129993255
@SQ SN:chr11 LN:121843856
@SQ SN:chr12 LN:121257530
@SQ SN:chr13 LN:120284312
@SQ SN:chr14 LN:125194864
@SQ SN:chr15 LN:103494974
@SQ SN:chr16 LN:98319150
@SQ SN:chr17 LN:95272651
@SQ SN:chr18 LN:90772031
@SQ SN:chr19 LN:61342430
@SQ SN:chrX LN:166650296
@SQ SN:chrY LN:15902555
@SQ SN:chrM LN:16299
@PG ID:Bowtie VN:0.12.7 CL:"bowtie -q -p 1 -S --phred33-quals /home/a-m/galaxy/bioindices
HWUSI-EAS610_0001:3:1:4:1405#0/1    16   chr19   60874227   255   36M   *
HWUSI-EAS610_0001:3:1:5:1490#0/1    0    chr19   32960373   255   36M   *
HWUSI-EAS610_0001:3:1:6:388#0/1    16   chr19   18177553   255   36M   *
```

**History Pane (Right):**

**6: Map with Bowtie f or Illumina on data 1**  
: mapped reads  
~270,000 lines, 24 comments  
format: sam, database: mm9  
Sequence file aligned.

1.QNAME 2.FLAG 3.RNAME 4.POS 5.MAPQ 6.CIGAR

QNAME	FLAG	RNAME	POS	MAPQ	CIGAR
@HD	VN:1.0 SO:unsorted				
@SQ	SN:chr1 LN:197195432				
@SQ	SN:chr2 LN:181748887				
@SQ	SN:chr3 LN:159599783				
@SQ	SN:chr4 LN:155630120				
@SQ	SN:chr5 LN:152537259				

**5: FASTQ Summary**  
Statistics on data 1

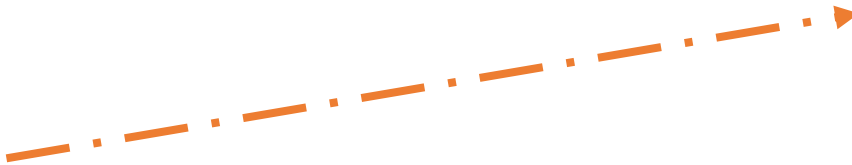
# Step 5A: Calling Peaks with MACs

With our mapped ChiP-Seq reads, we now want to call peaks.

Select NGS: Peak Calling



Select MACS



NGS: SAM Tools  
NGS: GATK Tools (beta)  
**NGS: Peak Calling**

- MACS14 Model-based Analysis of ChIP-Seq (1.4.1)
- MACS Model-based Analysis of ChIP-Seq

**NGS: Simulation**

# Step 5B: Calling Peaks with MACs

Run MACs with the default parameters.

MACS (version 1.0.1)

**Experiment Name:**  
MACS in Galaxy

**Paired End Sequencing:**  
Single End

**ChIP-Seq Tag File:**  
6: Map with Bowtie for Illumina on data 1: mapped reads

**ChIP-Seq Control File:**  
Selection is Optional

**Effective genome size:**  
270000000  
default: 2.7e+9

**Tag size:**  
25

**Band width:**  
300

**Pvalue cutoff for peak detection:**  
1e-05  
default: 1e-5

**Select the regions with MFOLD high-confidence enrichment:**  
32

**Parse xls files into into distinct interval files:**  
☐

**Save shifted raw tag count at every bp into a wiggle file:**  
Do not create wig file (faster)

**Use fixed background lambda as local lambda for every region:**  
☐  
up to 9X more time consuming

**3 levels of regions around the peak region to calculate the background:**  
1000,5000,10000

**Build Model:**  
Build the shifting model

**Diagnosis report:**  
Do not produce report (faster)  
up to 9X more time consuming

**Perform the new peak detection method (futurefdr):**  
☐  
The default method only consider the peak location, 1k, 5k, and 10k

**Execute**



# Step 5C: Calling Peaks with MACs

When done, MACs will create two files.

8: Macs on data 6 (html report) is an html document with information on the peak calling process.

7: MACs on data 6 (peaks:bed) is a BED file with coordinates and scores of ChIP-Seq peaks in chr19.

track name="MACS peaks for MACS_in_Galaxy"				
chr19	3100129	3100373	MACS_peak_1	54.24
chr19	3204496	3204775	MACS_peak_2	77.98
chr19	3208309	3208631	MACS_peak_3	127.67
chr19	3291585	3292811	MACS_peak_4	124.30
chr19	3347631	3348358	MACS_peak_5	76.32
chr19	3355095	3355575	MACS_peak_6	111.02
chr19	3450738	3451989	MACS_peak_7	177.92

**8: MACS on data 6 (html report)**

**7: MACS on data 6 (peaks: bed)**

728 regions, 1 comments  
format: **bed**, database: **mm9**

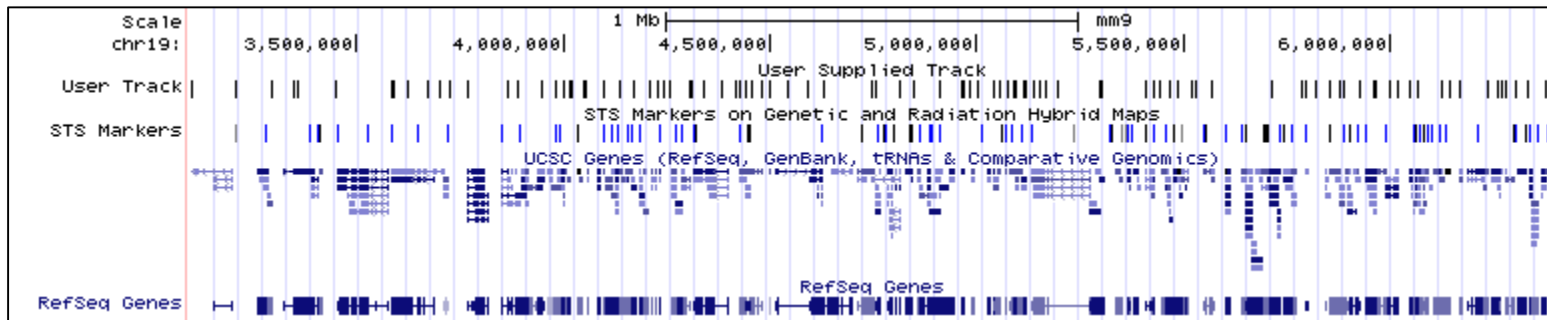
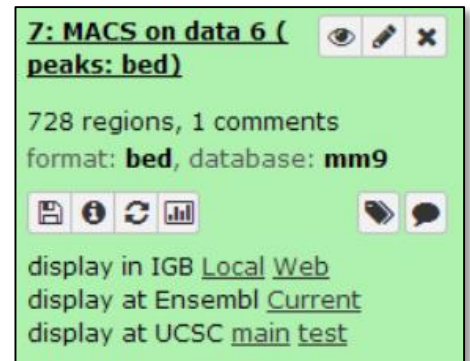
display in IGB [Local](#) [Web](#)  
display at Ensembl [Current](#)  
display at UCSC [main](#) [test](#)

1.Chrom	2.Start	3.End	4.Name	5
track name="MACS peaks for MACS_in_Galaxy"				
chr19	3100129	3100373	MACS_peak_1	54.24
chr19	3204496	3204775	MACS_peak_2	77.98
chr19	3208309	3208631	MACS_peak_3	127.67
chr19	3291585	3292811	MACS_peak_4	124.30
chr19	3347631	3348358	MACS_peak_5	76.32
chr19	3355095	3355575	MACS_peak_6	111.02
chr19	3450738	3451989	MACS_peak_7	177.92

# Step 5D: Calling Peaks with MACs

In the **7: MACs on data 6 (peaks:bed)** section in the **History** pane click [main](#) next to display at UCSC browser.

The result should look similar to below:



## Discussion

1. Look at the **BED** file. How many peaks were found?

# Call Chip-Seq Peaks with a Control Sample

We will perform the same procedure we did in the previous exercise. This time though, we will work with a control sample instead of an experimental one.

# Step 6A: Map Control ChIP-Seq Reads to MM9 Genome

Let's map the reads in `G1E_E4R_input_(chr19).fastqsanger` to the mouse genome.

Select NGS: Mapping



Then select Map with Bowtie for Illumina



[FASTA manipulation](#)  
[NGS: QC and manipulation](#)  
[Picard Tools](#)  
[NGS: Mapping](#)

- [Lastz paired reads map short paired reads against reference sequence](#)
- [Lastz map short reads against reference sequence](#)
- [Map with Bowtie for SOLiD](#)
- [Map with Bowtie for Illumina](#)
- [Map with BWA for Illumina](#)
- [Map with BWA for SOLiD](#)

## Step 6B: Map Control ChIP-Seq Reads to MM9 Genome

Make sure to select **mm9** as the reference genome.

Make sure **2: G1E\_ER4\_input\_(chr..fastqsanger)** is selected.

Click **Execute**.

It will take a few moments to complete.

When done, click the view icon.



Map with Bowtie for Illumina (version 1.1.2)

Will you select a reference genome from your built-in index?:  
 Built-ins were indexed using default options

Select a reference genome:  
 if your genome of interest is not listed - contact Galaxy

Is this library mate-paired?:

FASTQ file:  
 Must have ASCII encoded quality scores

Bowtie settings to use:  
 For most mapping needs use Commonly used settings. control use Full parameter list

Suppress the header in the output SAM file:  
☐ Bowtie produces SAM with several lines of header infor

# Step 6C: Map Control ChIP-Seq Reads to MM9 Genome

Your **Main** pane and **History** pane should look like the following:

The screenshot displays the Galaxy web interface. The **Main** pane on the left shows a SAM file with the following content:

```
@HD VN:1.0 SO:unsorted
@SQ SN:chr1 LN:197195432
@SQ SN:chr2 LN:181748087
@SQ SN:chr3 LN:159599783
@SQ SN:chr4 LN:155630120
@SQ SN:chr5 LN:152537259
@SQ SN:chr6 LN:149517037
@SQ SN:chr7 LN:152524553
@SQ SN:chr8 LN:131738871
@SQ SN:chr9 LN:124076172
@SQ SN:chr10 LN:129993255
@SQ SN:chr11 LN:121843856
@SQ SN:chr12 LN:121257530
@SQ SN:chr13 LN:120284312
@SQ SN:chr14 LN:125194864
@SQ SN:chr15 LN:103494974
@SQ SN:chr16 LN:98319150
@SQ SN:chr17 LN:95272651
@SQ SN:chr18 LN:90772031
@SQ SN:chr19 LN:61342430
@SQ SN:chrX LN:166650296
@SQ SN:chrY LN:15902555
@SQ SN:chrM LN:16299
@PG ID:Bowtie VN:0.12.7 CL:"bowtie -q -p 1 -S --phred33-quals /home/a-m/galaxy
HWUSI-EAS610:1:1:3:1131#0/1 16 chr19 44344158 255 36M *
HWUSI-EAS610:1:1:3:888#0/1 0 chr19 56973957 255 36M *
HWUSI-EAS610:1:1:4:941#0/1 16 chr19 10937148 255 36M *
```

The **History** pane on the right shows a list of jobs. The top job is **CompGenChIP3** (171.5 MB). Below it are three jobs related to mapping reads to the MM9 genome:

- 9: Map with Bowtie f or Illumina on data 2 : mapped reads** (~160,000 lines, 24 comments, format: sam, database: mm9)
- 8: MACS on data 6 (html report)**
- 7: MACS on data 6 (peaks: bed)**
- 6: Map with Bowtie f or Illumina on data 1 : mapped reads**

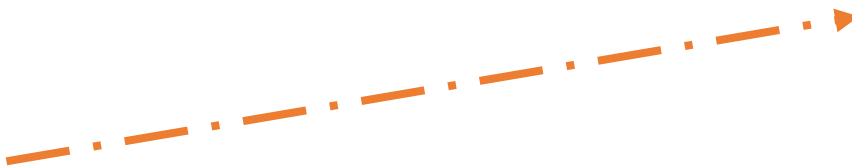
# Step 7A: Calling Peaks with MACs on Control Chip-Seq Reads

Like before, we want to call peaks in our mapped Control ChiP-Seq reads.

Select NGS: Peak Calling



Select MACS



**NGS: SAM Tools**  
**NGS: GATK Tools (beta)**  
**NGS: Peak Calling**

- MACS14 Model-based Analysis of ChIP-Seq (1.4.1)
- MACS Model-based Analysis of ChIP-Seq

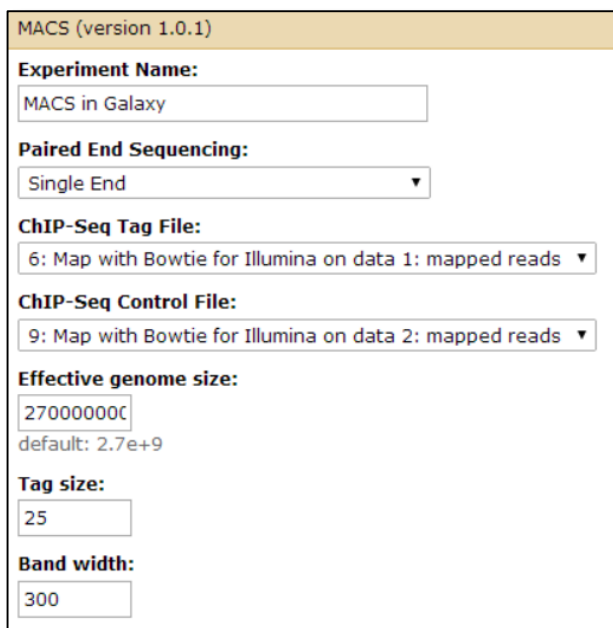
**NGS: Simulation**

# Step 7B: Calling Peaks with MACs on Control Chip-Seq Reads

Select **Map with Bowtie for Illumina** on data 1 (the experimental aligned reads) for the Chip-Seq Tag File.

Select **Map with Bowtie for Illumina** on data 2 (the control aligned reads) for the Chip-Seq Control File.

Click Execute



MACS (version 1.0.1)

**Experiment Name:**  
MACS in Galaxy

**Paired End Sequencing:**  
Single End

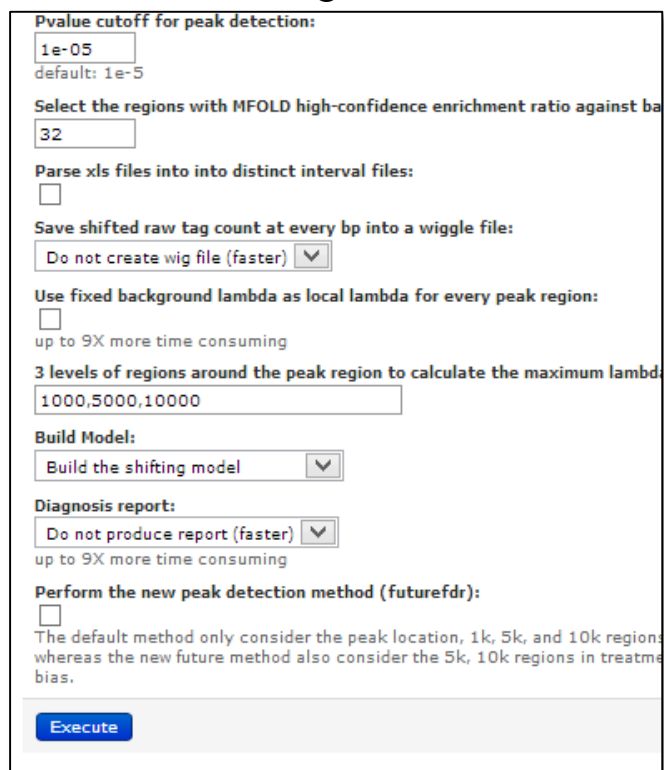
**ChIP-Seq Tag File:**  
6: Map with Bowtie for Illumina on data 1: mapped reads

**ChIP-Seq Control File:**  
9: Map with Bowtie for Illumina on data 2: mapped reads

**Effective genome size:**  
270000000  
default: 2.7e+9

**Tag size:**  
25

**Band width:**  
300



**Pvalue cutoff for peak detection:**  
1e-05  
default: 1e-5

**Select the regions with MFOLD high-confidence enrichment ratio against background:**  
32

**Parse xls files into into distinct interval files:**  
☐

**Save shifted raw tag count at every bp into a wiggle file:**  
Do not create wig file (faster)

**Use fixed background lambda as local lambda for every peak region:**  
☐  
up to 9X more time consuming

**3 levels of regions around the peak region to calculate the maximum lambda:**  
1000,5000,10000

**Build Model:**  
Build the shifting model

**Diagnosis report:**  
Do not produce report (faster)  
up to 9X more time consuming

**Perform the new peak detection method (futurefdr):**  
☐  
The default method only consider the peak location, 1k, 5k, and 10k regions whereas the new future method also consider the 5k, 10k regions in treatment bias.

**Execute**

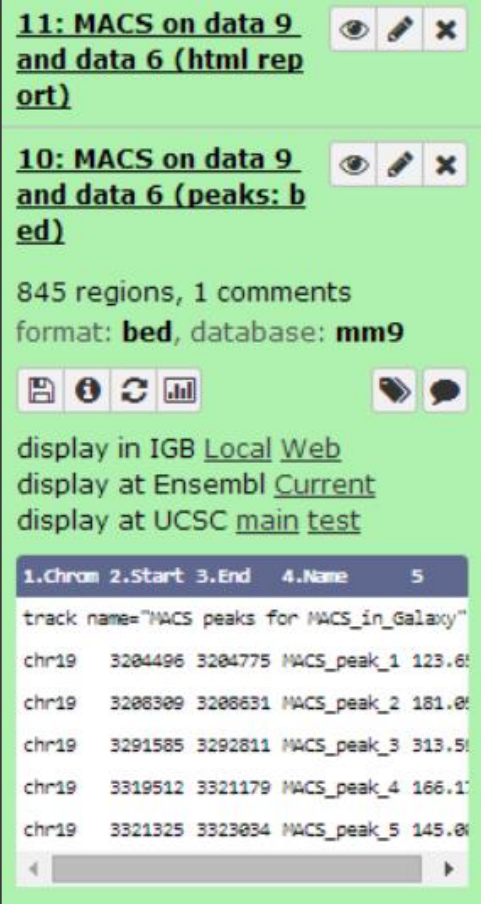


# Step 7C: Calling Peaks with MACs on Control Chip-Seq Reads

Once again, MACs creates a **BED** file containing the peak coordinates and an **HTML** file containing information on the peak calling process.

## Discussion

1. Examine the **BED** track.
2. How many peaks are called when using a control sample?
3. How does this compare to the previous situation where we only had experimental Chip-Seq reads?



**11: MACs on data 9 and data 6 (html report)**

**10: MACs on data 9 and data 6 (peaks: bed)**

845 regions, 1 comments  
format: **bed**, database: **mm9**

display in IGB [Local Web](#)  
display at Ensembl [Current](#)  
display at UCSC [main test](#)

1.Chrom	2.Start	3.End	4.Name	5
track name="MACs peaks for MACs_in_Galaxy"				
chr19	3204496	3204775	MACs_peak_1	123.6
chr19	3208309	3208631	MACs_peak_2	181.0
chr19	3291585	3292811	MACs_peak_3	313.5
chr19	3319512	3321179	MACs_peak_4	166.1
chr19	3321325	3323034	MACs_peak_5	145.0

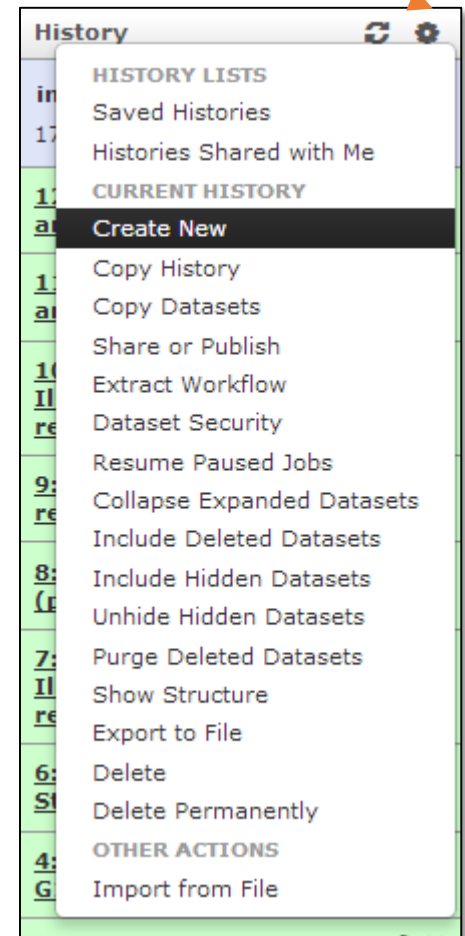
# Workflow Extraction

In this exercise, we will automate the Bowtie runs of our other 2 datasets.

# Step 1A: Workflow Extraction

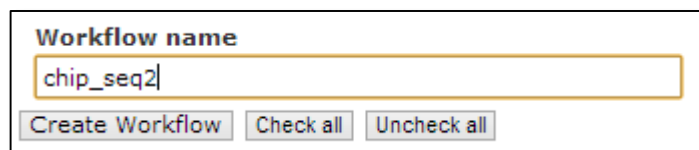
Click the  located in the **History** pane.

From the drop down menu, select **Extract Workflow**.



# Step 1B: Workflow Extraction

In the resulting window, give the workflow a new name.



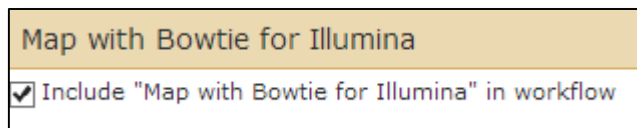
Workflow name

chip\_seq2

Create Workflow Check all Uncheck all

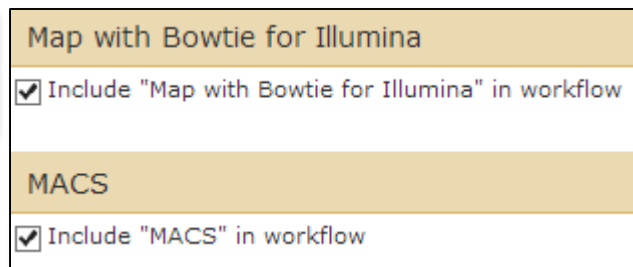
Click **Uncheck All**.

Ensure the following are checked and press **Create Workflow**.



Map with Bowtie for Illumina

☒ Include "Map with Bowtie for Illumina" in workflow

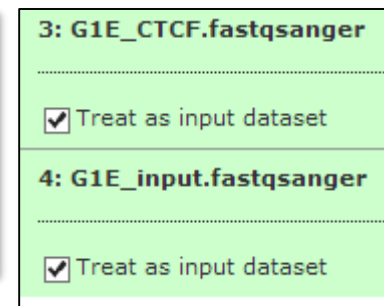


Map with Bowtie for Illumina

☒ Include "Map with Bowtie for Illumina" in workflow

MACS

☒ Include "MACS" in workflow



3: G1E\_CTCF.fastqsanger

☒ Treat as input dataset

4: G1E\_input.fastqsanger

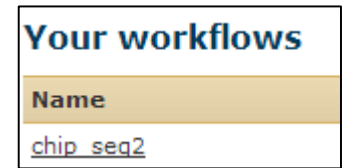
☒ Treat as input dataset

# Step 1C: Workflow Extraction

Select **All Workflows** at the bottom of the **Tool** pane.





Select **chipseq2**





On the next page, ensure the following settings and click **Run**.

# Step 1D: Workflow Extraction

**Step 1: Input dataset**

**Input Dataset**   
3: G1E\_CTCF.fastqsanger 

**Step 2: Input dataset**


**Input Dataset**   
4: G1E\_input.fastqsanger 

**Step 3: Map with Bowtie for Illumina**

**Will you select a reference genome from the list?**  
Use a built-in index

**Select a reference genome**  
mm9 (value not yet validated)

**Is this library mate-paired?**  
Single-end

**FASTQ file**  
3: G1E\_CTCF.fastqsanger 

**Bowtie settings to use**  
Commonly used


**Suppress the header in the output SAM**  
False

**Step 4: Map with Bowtie for Illumina**

**Will you select a reference genome from the list?**  
Use a built-in index

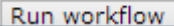
**Select a reference genome**  
mm9 (value not yet validated)

**Is this library mate-paired?**  
Single-end


**FASTQ file**  
4: G1E\_input.fastqsanger 

**Step 5: MACS (version 1.0.1)**

☐ Send results to a new history

**Run workflow** 

Click when done



# Identifying Differential Binding Sites

In this exercise, we will identify binding sites exclusive to undifferentiated and differentiated cell lines as well as those common to both.

# Step 1: Subtract Peaks Between Cell Lines

Select **Operate on Genomic Intervals** and **Subtract**.

## Operate on Genomic Intervals

- Intersect the intervals of two datasets
- Subtract the intervals of two datasets

Choose your G1E MACS peaks as your 2<sup>nd</sup> dataset and your G1E-ER4 peaks as your 1<sup>st</sup> dataset.

Click **Execute**.

**Subtract (version 1.0.0)**

**Subtract:**  
14: MACS on data 13 and data 12 (peaks: bed) ▼  
Second dataset

**from:**  
10: MACS on data 9 and data 6 (peaks: bed) ▼  
First dataset

**Return:**  
Intervals with no overlap ▼  
of the first dataset (see figure below)

**where minimal overlap is:**  
1  
(bp)

**Execute**



## Step 2: Subtract Peaks Between Cell Lines.

The resulting **BED** file contains peaks exclusive to the **differentiated** cell line (G1E-ER4).

### Discussion

1. How many peaks are exclusive to G1E-ER4?

-----

Redo Step1 only **SWITCH** the input order.

Choose your G1E MACS peaks as your 1<sup>st</sup> dataset and your G1E-ER4 peaks as your 2<sup>nd</sup> dataset.

# Step 3: Intersect Peaks Between Cell Lines

Select **Operate on Genomic Intervals** and **Intersect**.

Choose your G1E-ER4 MACS peaks as your 1<sup>st</sup> dataset and your G1E peaks as your 2<sup>nd</sup> dataset.

Click **Execute**.

**Operate on Genomic Intervals**

- Intersect the intervals of two datasets
- Subtract the intervals of two datasets

Intersect (version 1.0.0)

**Return:**  
Overlapping Intervals   
(see figure below)

**of:**  
10: MACS on data 9 and data 6 (peaks: bed)   
First dataset

**that intersect:**  
14: MACS on data 13 and data 12 (peaks: bed)   
Second dataset

**for at least:**  
1  
(bp)

**Execute**