

# UF Research Computing

---

## Introduction to Galaxy at UF HPC

Oleksandr Moskalenko  
Assoc. Sci., UF HPC Center  
Biological Applications Support  
Matt Gitzendanner  
Assoc Sci., Biology / HPC Training

UF Research Computing Day 2011



# Today's research computing

---





# Approaches

---





# Approaches

---





# Approaches

---







## ▶ NCBI/ BLAST/ blastn suite

blastn

[blastp](#)[blastx](#)[tblastn](#)[tblastx](#)BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

## Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) ?

[Clear](#)

Query subrange ?

From

To

Or, upload file

Choose File

No file chosen ?

Job Title

Enter a descriptive title for your BLAST search ?

☐ Align two or more sequences ?

## Choose Search Set

Database

☒ Human genomic + transcript ☐ Mouse genomic + transcript ☐ Others (nr etc.):

Human genomic plus transcript (Human G+T) ?

Exclude

Optional

Entrez Query

Optional

☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Enter an Entrez query to limit search ?

## Program Selection

Optimize for

- ☒ Highly similar sequences (megablast)
- ☐ More dissimilar sequences (discontiguous megablast)
- ☐ Somewhat similar sequences (blastn)

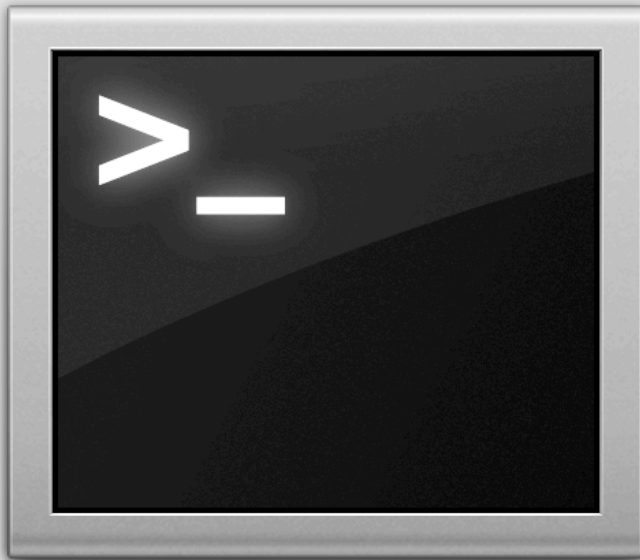
Choose a BLAST algorithm ?

**BLAST**Search **database Human G+T** using **Megablast (Optimize for highly similar sequences)**☐ Show results in a new window



# Command Line Environment

Head node



Login to  
head  
node

Scheduler



Interactive  
session or batch  
submission

Computing  
resources



Your job  
runs on the  
cluster



**Bio·IT World.com**

Cambridge Healthtech Institute

[CHI Home](#) | [Conferences](#) | [Reports](#) |

September 28, 2011 | Bio-IT World &gt; Galaxy Provides Life Support for NGS Exploration


 Abstract

**September 27, 2011** | Enter the term "galaxy" in a Web search engine, Penn State's Anton Nekrutenko muses, and the top hits are likely to be an astrophysical entity or "a very bad soccer team." But making fast strides up the web charts is the Galaxy open-source tool, which is coming into its own as more and more researchers seek ways to easily handle and manipulate next-generation sequencing (NGS) and other large

all or download  
you can analyze  
roles and much,  
for Comparative  
Taylor, v







# What is Galaxy?

---

## ◆ Computational biology platform

- Open and Web-based
- Accessible
- Reproducible
- Transparent





# Galaxy Analysis Workspace

**Galaxy / UF HPC** Analyze Data Workflow Shared Data Admin Help User

**Tools** Options ▾  
[Get Data](#)  
[Send 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](#)  
[Statistics](#)  
[Graph/Display Data](#)  
[Regional Variation](#)  
[Multiple regression](#)  
[Multivariate Analysis](#)  
[Evolution](#)  
[Motif Tools](#)  
[Multiple Alignments](#)  
[Metagenomic analyses](#)  
[FASTA manipulation](#)  
[NCBI BLAST+](#)  
[NGS: QC and manipulation](#)  
[NGS: Picard \(beta\)](#)  
[NGS: Mapping](#)  
[NGS: Indel Analysis](#)  
[NGS: RNA Analysis](#)  
[NGS: SAM Tools](#)

**MACS**  
**Treatment file:**  
3: hg19.chr10.bam  
**Input file:**  
Selection is Optional  
**Format:**  
Auto  
**Effective Genome Size:**  
Human (hg18)  
**Tag size (Optional):**  
25  
**P-Value:**  
1e-05  
**Keep duplicate tags at the exact same location?:**  
☐ Keep ALL  
☐ Auto by Binomial  
☒ Keep Single  
**Use Model?:**  
True  
**small fold enrichment for model building:**  
10  
**large fold:**  
30  
**Advanced Options:**

**History** Options ▾  
0915 Macs Exercise 5.3 Gb  
35: Summary Statistics on data 28  
33: UCSC Main on Human: ct UserTrack 3545 (chr1:156690-165971)  
31: MACS job log on hg19.chr9.bam  
30: MACS wiggle on hg19.chr9.bam  
29: MACS xls on hg19.chr9.bam  
28: MACS summits on hg19.chr9.bam  
27: MACS peaks on hg19.chr9.bam  
26: BAM-to-SAM on data 25: converted SAM  
25: hg19.chr9.bam  
24: hg19.chr8.bam  
23: hg19.chr7.bam





# Galaxy Analysis Workspace

**Galaxy / UF HPC** Analyze Data Workflow Shared Data Admin Help User

**Tools** Options ▾  
[Get Data](#)  
[Send 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](#)  
[Statistics](#)  
[Graph/Display Data](#)  
[Regional Variation](#)  
[Multiple regression](#)  
[Multivariate Analysis](#)  
[Evolution](#)  
[Motif Tools](#)  
**NGS: Peak Calling**

- [MACS](#) Model-based Analysis for ChIP-Seq
- [CCAT](#) Control-based ChIP-seq Analysis Tool
- [GeneTrack](#) indexer on a BED file
- [Peak predictor](#) on GeneTrack index

**NGS: Simulation**  
**SNP/WGA: Data; Filters**

**MACS**  
**Treatment file:**  
3: hg19.chr10.bam  
**Input file:**  
Selection is Optional  
**Format:**  
Auto  
**Effective Genome Size:**  
Human (hg18)  
**Tag size (Optional):**  
25  
**P-Value:**  
1e-05  
**Keep duplicate tags at the exact same location?:**  
☐ Keep ALL  
☐ Auto by Binomial  
☒ Keep Single  
**Use Model?:**  
True  
**small fold enrichment for model building:**  
10  
**large fold:**  
30  
**Advanced Options:**

**History** Options ▾  
0915 Macs Exercise 5.3 Gb  
[35: Summary Statistics on data 28](#) 👁 ✂ ✕  
[33: UCSC Main on Human: ct UserTrack 3545 \(chr1:156690-165971\)](#) 👁 ✂ ✕  
[31: MACS job log on hg19.chr9.bam](#) 👁 ✂ ✕  
[30: MACS wiggle on hg19.chr9.bam](#) 👁 ✂ ✕  
[29: MACS xls on hg19.chr9.bam](#) 👁 ✂ ✕  
[28: MACS summits on hg19.chr9.bam](#) 👁 ✂ ✕  
[27: MACS peaks on hg19.chr9.bam](#) 👁 ✂ ✕  
[26: BAM-to-SAM on data 25: converted SAM](#) 👁 ✂ ✕  
[25: hg19.chr9.bam](#) 👁 ✂ ✕  
[24: hg19.chr8.bam](#) 👁 ✂ ✕  
[23: hg19.chr7.bam](#) 👁 ✂ ✕





# Galaxy

# workspace

Galaxy / UF HP

Tools

Options

[Get Data](#)

[Send 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](#)

[Statistics](#)

[Graph/Display Data](#)

[Regional Variation](#)

[Multiple regression](#)

[Multivariate Analysis](#)

[Evolution](#)

[Motif Tools](#)

## NGS: Peak Calling

- [MACS Model-based Analysis for ChIP-Seq](#)
- [CCAT Control-based ChIP-seq Analysis Tool](#)
- [GeneTrack indexer on a BED file](#)
- [Peak predictor on GeneTrack index](#)

## NGS: Simulation

## SNP/WGA: Data; Filters

### MACS

Treatment file:

3: hg19.chr10.bam

Input file:

Selection is Optional

Format:

Auto

Effective Genome Size:

Human (hg18)

Tag size (Optional):

25

P-Value:

1e-05

Keep duplicate tags at the exact same location?:

☐ Keep ALL

☐ Auto by Binomial

☒ Keep Single

Use Model?:

True

small fold enrichment for model building:

10

large fold:

30

Advanced Options:

No

Diagnosis Report:

No

Execute

History

Options



0915 Macs Exercise



5.3 Gb

[35: Summary Statistics on data 28](#)

[33: UCSC Main on Human: ct UserTrack 3545 \(chr1:156690-165971\)](#)

[31: MACS job log on hg19.chr9.bam](#)

[30: MACS wiggle on hg19.chr9.bam](#)

[29: MACS xls on hg19.chr9.bam](#)

[28: MACS summits on hg19.chr9.bam](#)

[27: MACS peaks on hg19.chr9.bam](#)

[26: BAM-to-SAM on data 25: converted SAM](#)

[25: hg19.chr9.bam](#)

[24: hg19.chr8.bam](#)

[23: hg19.chr7.bam](#)





# Galaxy

Galaxy / UF HP

Tools Options

[Get Data](#)  
[Send 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](#)  
[Statistics](#)  
[Graph/Display Data](#)  
[Regional Variation](#)  
[Multiple regression](#)  
[Multivariate Analysis](#)  
[Evolution](#)  
[Motif Tools](#)

## NGS: Peak Calling

- [MACS](#) Model-based Analysis for ChIP-Seq
- [CCAT](#) Control-based ChIP-seq Analysis Tool
- [GeneTrack indexer](#) on a BED file
- [Peak predictor](#) on GeneTrack index

## NGS: Simulation

## SNP/WGA: Data; Filters

### MACS

Treatment file:

3: hg19.chr10.bam

Input file:

Selection is Optional

Format:

Auto

Effective Genome Size:

Human (hg18)

Tag size (Optional):

25

P-Value:

1e-05

Keep duplicate tags at the exact same location?:

- ☐ Keep ALL  
☐ Auto by Binomial  
☒ Keep Single

Use Model?:

True

small fold enrichment for model building:

10

large fold:

30

Advanced Options:

No

Diagnosis Report:

No

Execute

1

### History

Options



0915 Macs Exercise



5.3 Gb

[35: Summary Statistics on data 28](#)

[33: UCSC Main on Human: ct UserTrack 3545 \(chr1:156690-165971\)](#)

[31: MACS job log on hg19.chr9.bam](#)

[30: MACS wiggle on hg19.chr9.bam](#)

[29: MACS xls on hg19.chr9.bam](#)

[28: MACS summits on hg19.chr9.bam](#)

[27: MACS peaks on hg19.chr9.bam](#)

[26: BAM-to-SAM on data 25: converted SAM](#)

[25: hg19.chr9.bam](#)





## Galaxy

### Tools

[Get Data](#)

[Send Data](#)

[ENCODE Tools](#)

[Lift-Over](#)

[Text Manipulation](#)

[Filter and Sort](#)

[Join, Subtract and G](#)

[Convert Formats](#)

[Extract Features](#)

[Fetch Sequences](#)

[Fetch Alignments](#)

[Get Genomic Scores](#)

[Operate on Genomic](#)

[Statistics](#)

[Graph/Display Data](#)

[Regional Variation](#)

[Multiple regression](#)

[Multivariate Analysis](#)

[Evolution](#)

[Motif Tools](#)

### NGS: Peak Calling

▪ [MACS Model-based A](#)  
[ChIP-Seq](#)

▪ [CCAT Control-based](#)  
[Analysis Tool](#)

▪ [GeneTrack indexer](#) or

▪ [Peak predictor](#) on Gen  
[index](#)

### NGS: Simulation

[SNP/WGA: Data; Filters](#)

### MACS

chr9	179077	179078	MACS_peak_1	14.00	
chr9	503365	503366	MACS_peak_2	17.00	
chr9	764211	764212	MACS_peak_3	20.00	
chr9	2241905	2241906	MACS_peak_4	15.00	
chr9	3161806	3161807	MACS_peak_5	10.00	
chr9	3467733	3467734	MACS_peak_6	14.00	
chr9	3526275	3526276	MACS_peak_7	19.00	
chr9	3809982	3809983	MACS_peak_8	17.00	
chr9	3907058	3907059	MACS_peak_9	15.00	
chr9	4315804	4315805	MACS_peak_10	17.00	
chr9	4887865	4887866	MACS_peak_11	11.00	
chr9	5186618	5186619	MACS_peak_12	13.00	
chr9	5439013	5439014	MACS_peak_13	14.00	
chr9	5510340	5510341	MACS_peak_14	13.00	
chr9	5566231	5566232	MACS_peak_15	11.00	
chr9	5609455	5609456	MACS_peak_16	9.00	
chr9	5832438	5832439	MACS_peak_17	12.00	
chr9	6015764	6015765	MACS_peak_18	17.00	
chr9	6038019	6038020	MACS_peak_19	16.00	
chr9	6681231	6681232	MACS_peak_20	29.00	
chr9	6757871	6757872	MACS_peak_21	12.00	
chr9	7028374	7028375	MACS_peak_22	11.00	
chr9	9428809	9428810	MACS_peak_23	8.00	
chr9	9442235	9442236	MACS_peak_24	5.00	
chr9	9487422	9487423	MACS_peak_25	3.00	
chr9	9524985	9524986	MACS_peak_26	5.00	
chr9	9677411	9677412	MACS_peak_27	7.00	
chr9	12776446	12776447	MACS_peak_28	14.00	
chr9	13034378	13034379	MACS_peak_29	12.00	
chr9	14201262	14201263	MACS_peak_30	12.00	
chr9	15038466	15038467	MACS_peak_31	7.00	
chr9	16371450	16371451	MACS_peak_32	12.00	
chr9	16704876	16704877	MACS_peak_33	10.00	
chr9	16964119	16964120	MACS_peak_34	11.00	
chr9	17005070	17005071	MACS_peak_35	11.00	
chr9	17063745	17063746	MACS_peak_36	10.00	
chr9	18168582	18168583	MACS_peak_37	9.00	
chr9	19050354	19050355	MACS_peak_38	13.00	
chr9	21085741	21085742	MACS_peak_39	47.00	
chr9	21591829	21591830	MACS_peak_40	16.00	
chr9	22016338	22016339	MACS_peak_41	7.00	

Execute

Options ▾



5.3 Gb

cs on

man:

chr1:156690-



data





25: ng19:chr9.bam







# Metadata

**History** **Options** ▼

LANA ChIP peaks on hg19 5.3 Gb

**Tags:**


LANA ×

chip ×




hg19 ×

peaks ×




chr9 ×





**Annotation / Notes:**  
Peak calling on LANA ChIP-Seq data using Human chromosome 9 from hg19 build

**27: MACS peaks on hg19.chr9.bam**   

236 regions  
format: bed, database: ?

**Tags:**


LANA ×

chip ×

hg19 ×

chr9 ×

MACS ×



view in [GeneTrack](#)

1.Chrom	2.Start	3.End	4.Name
chr9	176690	179457	MACS_pea
chr9	502364	506252	MACS_pea
chr9	763181	765291	MACS_pea
chr9	2241428	2243431	MACS_pea
chr9	3161298	3162300	MACS_pea
chr9	3467312	3468066	MACS_pea





# Getting Data into Galaxy

---

## ◆ Upload a file from your computer

- scp to HPC and load from within Galaxy
- Copy files to HPC using Samba

## ◆ External data

- UCSC table browser
- Biomart
- interMine / modMine
- EuPathDB
- EncodeDB
- EpiGRAPH
- FlyMine
- GrameneMart...





# Data libraries

## Data Library “GMS 6001 MACS Exercise”

MACS test data

<input type="checkbox"/> Name	Message	Uploaded By	Date	File Size
<input type="checkbox"/> 2010-12-14 7 hg19 aln sorted.bam ▼		om@hpc.ufl.edu	2011-09-13	1.6 Gb
<input type="checkbox"/> 2010-12-14 7 hhv8 aln sorted.bam ▼		om@hpc.ufl.edu	2011-09-13	1.4 Gb
<input type="checkbox"/> hg19.chr10.bam ▼		om@hpc.ufl.edu	2011-09-14	80.8 Mb
<input type="checkbox"/> hg19.chr11.bam ▼		om@hpc.ufl.edu	2011-09-14	82.5 Mb
<input type="checkbox"/> hg19.chr12.bam ▼		om@hpc.ufl.edu	2011-09-14	74.9 Mb
<input type="checkbox"/> hg19.chr13.bam ▼		om@hpc.ufl.edu	2011-09-14	50.9 Mb
<input type="checkbox"/> hg19.chr14.bam ▼		om@hpc.ufl.edu	2011-09-14	36.1 Mb
<input type="checkbox"/> hg19.chr15.bam ▼		om@hpc.ufl.edu	2011-09-14	48.1 Mb
<input type="checkbox"/> hg19.chr16.bam ▼		om@hpc.ufl.edu	2011-09-14	55.9 Mb
<input type="checkbox"/> hg19.chr17.bam ▼		om@hpc.ufl.edu	2011-09-14	64.5 Mb
<input type="checkbox"/> hg19.chr18.bam ▼		om@hpc.ufl.edu	2011-09-14	33.5 Mb
<input type="checkbox"/> hg19.chr19.bam ▼		om@hpc.ufl.edu	2011-09-14	39.6 Mb
<input type="checkbox"/> hg19.chr1.bam ▼		om@hpc.ufl.edu	2011-09-14	148.5 Mb
<input type="checkbox"/> hg19.chr20.bam ▼		om@hpc.ufl.edu	2011-09-14	38.5 Mb
<input type="checkbox"/> hg19.chr21.bam ▼		om@hpc.ufl.edu	2011-09-14	17.5 Mb
<input type="checkbox"/> hg19.chr22.bam ▼		om@hpc.ufl.edu	2011-09-14	16.9 Mb
<input type="checkbox"/> hg19.chr2.bam ▼		om@hpc.ufl.edu	2011-09-14	126.3 Mb
<input type="checkbox"/> hg19.chr2.sam ▼		om@hpc.ufl.edu	2011-09-14	488.0 Mb
<input type="checkbox"/> hg19.chr3.bam ▼		om@hpc.ufl.edu	2011-09-14	118.0 Mb
<input type="checkbox"/> hg19.chr4.bam ▼		om@hpc.ufl.edu	2011-09-14	85.7 Mb
<input type="checkbox"/> hg19.chr5.bam ▼		om@hpc.ufl.edu	2011-09-14	102.7 Mb
<input type="checkbox"/> hg19.chr6.bam ▼		om@hpc.ufl.edu	2011-09-14	65.7 Mb
<input type="checkbox"/> hg19.chr7.bam ▼		om@hpc.ufl.edu	2011-09-14	89.9 Mb
<input type="checkbox"/> hg19.chr8.bam ▼		om@hpc.ufl.edu	2011-09-14	85.9 Mb
<input type="checkbox"/> hg19.chr9.bam ▼		om@hpc.ufl.edu	2011-09-14	64.8 Mb

For selected datasets:





# Data Access Control

## Roles associated with new group

HPC test ChIP-seq analyses

>>

## Users associated with new group

om@hpc.ufl.edu  
magitz@ufl.edu

## Groups

[Advanced Search](#)

<input type="checkbox"/> <u>Name</u> ↓	Users	Roles
<input type="checkbox"/> <u>HPC</u> ▼	0	2
<input type="checkbox"/> <u>Taylor HPC Lab</u> ▼	2	1

For 0 selected groups: [Delete](#) [Undelete](#) [Purge](#)

## Roles

[Advanced Search](#)

<input type="checkbox"/> <u>Name</u> ↓	<u>Description</u>	<u>Type</u>	<u>Groups</u>
<input type="checkbox"/> <u>HPC</u> ▼	Role for group HPC	system	1
<input type="checkbox"/> <u>HPC test ChIP-seq analyses</u> ▼	Test analyses of ChIP-seq data	admin	1

## Users

[Advanced Search](#)

<input type="checkbox"/> <u>Email</u> ↓	<u>User Name</u>	<u>Groups</u>	<u>Roles</u>	<u>External</u>	<u>Last Login</u>
<input type="checkbox"/> <u>aedison@ufl.edu</u> ▼	aedison	0	1	yes	Sep 15, 2011
<input type="checkbox"/> <u>bostwick@ufl.edu</u> ▼	bostwick	0	1	yes	Sep 15, 2011
<input type="checkbox"/> <u>cgraves3@ufl.edu</u> ▼	cgraves3	0	1	yes	Sep 15, 2011
<input type="checkbox"/> <u>cjeffrey@ufl.edu</u> ▼	cjeffrey	0	1	yes	Sep 15, 2011
<input type="checkbox"/> <u>colltd3@ufl.edu</u> ▼	colltd3	0	1	yes	Sep 15, 2011





# Galaxy Tool Suites

---

- ◆ Text Manipulation
- ◆ Format Converters
- ◆ Filtering and Sorting
- ◆ Join, Subtract, Group
- ◆ Sequence Tools
- ◆ Multi-species Alignment Tools
- ◆ Genomic Interval Operation
- ◆ Summary Statistics, graphing
- ◆ Regional Variation
- ◆ EMBOSS
- ◆ Evolution / Phylogeny
- ◆ RNA-Seq
- ◆ ChIP-Seq
- ◆ GATK





# A galaxy of tools

## NGS: QC and manipulation

### ILLUMINA DATA

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

### ROCHE-454 DATA

Build base quality distribution

Select high quality segments

Combine FASTA and QUAL into FASTQ

### AB-SOLID DATA

Convert SOLID output to fastq

Compute quality statistics for SOLID data

Draw quality score boxplot for SOLID data

### GENERIC FASTQ MANIPULATION

Filter FASTQ reads by quality score and length

FASTQ Trimmer by column

FASTQ Quality Trimmer by sliding window

## Evolution

### Metagenomic analyses

### Human Genome Variation

### EMBOSS

## NGS TOOLBOX BETA

### NGS: QC and manipulation

### NGS: Mapping

#### ILLUMINA

- Map with Bowtie for Illumina
- Map with BWA for Illumina

#### ROCHE-454

- Lastz map short reads against reference sequence
- Megablast compare short reads against htgs, nt, and wgs databases

- Parse blast XML output

#### AB-SOLID

- Map with Bowtie for SOLID

### NGS: SAM Tools

### NGS: Indel Analysis

### NGS: Peak Calling

### NGS: RNA Analysis

## RGENETICS

### SNP/WGA: Data; Filters

### SNP/WGA: QC; LD; Plots

### SNP/WGA: Statistical Models

## NGS TOOLBOX BETA

### NGS: QC and manipulation

### NGS: Mapping

### NGS: SAM Tools

- Filter SAM on bitwise flag values
- Convert SAM to interval
- SAM-to-BAM converts SAM format to BAM format
- BAM-to-SAM converts BAM format to SAM format
- Merge BAM Files merges BAM files together
- Generate pileup from BAM dataset
- Filter pileup on coverage and SNPs
- Pileup-to-Interval condenses pileup format into ranges of bases
- flagstat provides simple stats on BAM files

### NGS: Indel Analysis

### NGS: Peak Calling

### NGS: RNA Analysis

## RGENETICS

### SNP/WGA: Data; Filters

### SNP/WGA: QC; LD; Plots

### SNP/WGA: Statistical Models

## NGS: SAM Tools

### NGS: Indel Analysis

- Filter Indels for SAM
- Extract indels from SAM
- Indel Analysis

### NGS: Peak Calling

- MACS Model-based Analysis of ChIP-Seq
- GeneTrack indexer on a BED file
- Peak predictor on GeneTrack index

### NGS: RNA Analysis

#### RNA-SEQ

- Tophat Find splice junctions using RNA-seq data
- Cufflinks transcript assembly and FPKM (RPKM) estimates for RNA-Seq data
- Cuffcompare compare assembled transcripts to a reference annotation and track Cufflinks transcripts across multiple experiments
- Cuffdiff find significant changes in transcript expression, splicing, and promoter use

#### FILTERING

- Filter Combined Transcripts using tracking file





# Galaxy Workflows

**Unknown**  
*This tool cannot be used in workflows*

**BAM-to-SAM**  
☒ Include "BAM-to-SAM" in workflow

**Convert Genomic Intervals To Strict BED6**  
☒ Include "Convert Genomic Intervals To Strict BED6" in workflow

**MACS**  
☒ Include "MACS" in workflow

**Convert BED to GeneTrack Index**  
☒ Include "Convert BED to GeneTrack Index" in workflow

25: hg19.chr9.bam  
☒ Treat as input dataset

26: BAM-to-SAM on data 25:  
converted SAM

27: MACS peaks on hg19.chr9.bam

27: MACS peaks on hg19.chr9.bam

28: MACS summits on hg19.chr9.bam

29: MACS xls on hg19.chr9.bam

30: MACS wiggle on hg19.chr9.bam

31: MACS job log on hg19.chr9.bam

27: MACS peaks on hg19.chr9.bam

31: M  
hg19.

30: M  
hg19.

29: M  
hg19.

28: M  
hg19.

27: MACS peaks on  
hg19.chr9.bam

26: BAM-to-SAM on data  
25: converted SAM

25: hg19.chr9.bam

24: hg19.chr8.bam

23: hg19.chr7.bam

22: hg19.chr6.bam

21: hg19.chr5.bam

20: hg19.chr4.bam

19: hg19.chr3.bam

Extract Workflow

Dataset Security

Show Deleted Datasets

Show Hidden Datasets

Show Structure

Export to File

Delete

Other Actions

Import from File





# Galaxy Workflows

Workflow Canvas | Workflow constructed from history 'LANA ChIP peaks on hg19'

**Details**

**Tool: MACS**

**Treatment file**  
Data input 'tfile' (interval or sam or bam or eland or elandmulti or bed)

**Input file**  
Data input 'cfile' (interval or sam or bam or eland or elandmulti or bed)

**Format:** ▼  
Auto

**Effective Genome Size:**  
Human (hg19)

**Tag size (Optional):** ▼  
25

**Details**

**Edit Workflow Attributes**

**Name:**  
Workflow constructed from history 'LANA ChIP peaks on hg19'

**Tags:**  
LANA x ChIP-Seq x hg19 x chr9 x

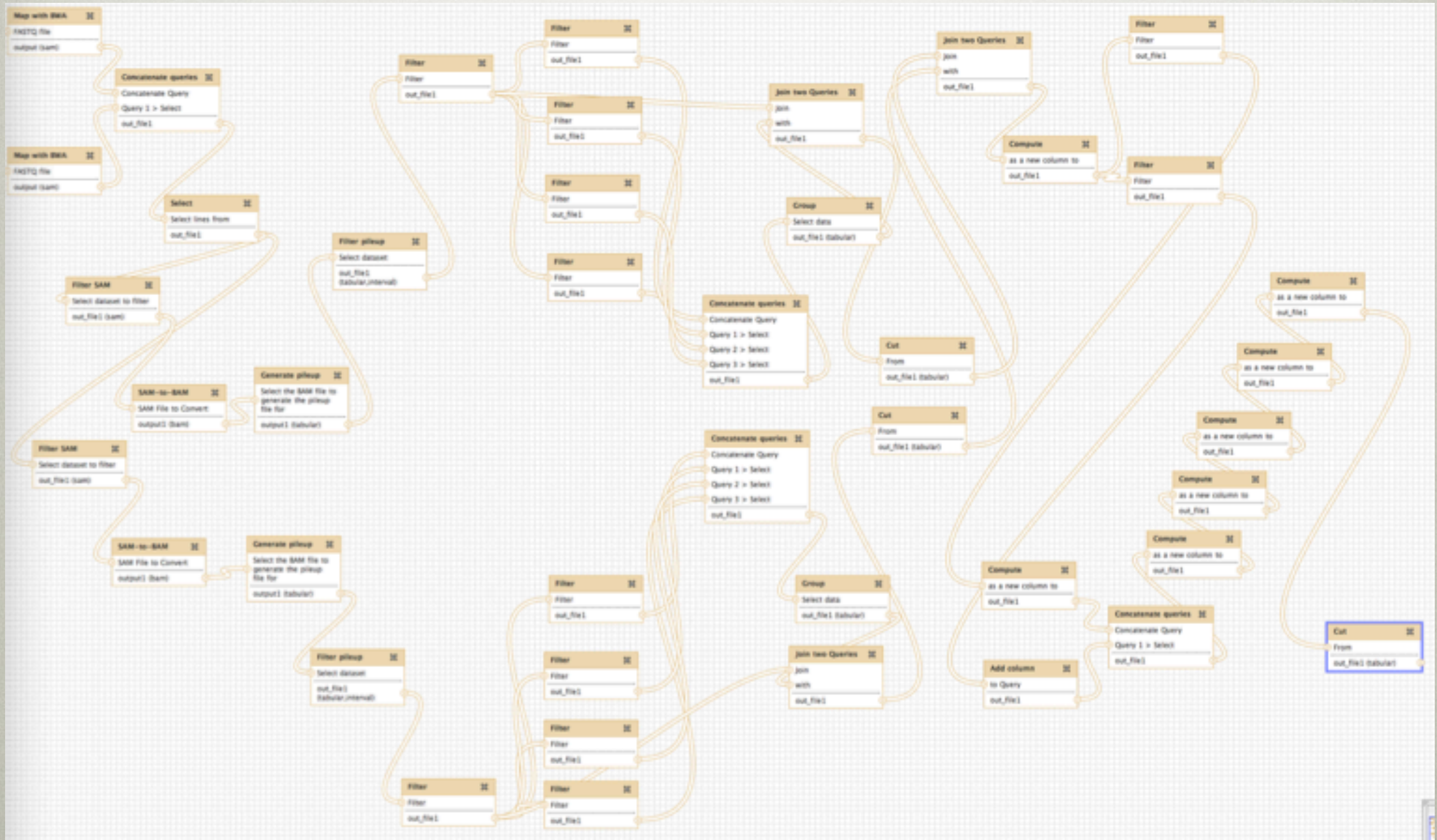
Apply tags to make it easy to search for and find items with the same tag.

**Annotation / Notes:**  
This is a partial peak calling with MACS using hg19 and chr9 data





# Galaxy Workflows







# Visualization

Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl

## UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search chr10:1-135,534,747 [gene](#) jump clear size 135,534,747 bp. configure

chr10 (p15.3-q26.3) p14 p13 q21.1

Scale chr10: 50 Mb 50000000 100000000

move start < 2.0 > Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. move end < 2.0 >

track search default tracks default order hide all add custom tracks track hubs configure reverse resize

refresh

collapse all Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes. expand all

### Mapping and Sequencing Tracks

refresh

<a href="#">Base Position</a>	<a href="#">Chromosome Band</a>	<a href="#">STS Markers</a>	<a href="#">18 FISH Clones</a>	<a href="#">Recomb Rate</a>	<a href="#">Map Contigs</a>
dense		hide	hide	hide	hide





# Sharing and publishing

---

## Share or Publish History 'LANA ChIP peaks on hg19'

### Making History Accessible via Link and Publishing It

This history is currently restricted so that only you and the users listed below can access it. You can:

**Make History Accessible via Link**

Generates a web link that you can share with other people so that they can view and import the history.

**Make History Accessible and Publish**

Makes the history accessible via link (see above) and publishes the history to Galaxy's [Published Histories](#) section, where it is publicly listed and searchable.

### Sharing History with Specific Users

You have not shared this history with any users.

**Share with a user**

[Back to Histories List](#)





# Sharing and publishing

## Share or Publish History 'LANA ChIP peaks on hg19'

### Making History Accessible via Link and Publishing It

This history is currently accessible via link and published.

Anyone can view and import this history by visiting the following URL:

<http://galaxy.hpc.ufl.edu/u/moskalenko/h/lana-chip-peaks-on-hg19> ✎

This history is publicly listed and searchable in Galaxy's Published Histories section.

You can:

#### Unpublish History

Removes this history from Galaxy's Published Histories section so that it is not publicly listed or searchable.

#### Disable Access to History via Link and Unpublish

Disables this history's link so that it is not accessible and removes history from Galaxy's Published Histories section so that it is not publicly listed or searchable.

### Sharing History with Specific Users

The following users will see this history in their history list and will be able to view, import, and run it.

Email

magitz@ufl.edu ▼

Share with another user





# Sharing and publishing

## Share or Publish Workflow 'LANA ChIP peaks on hg19'

### Making Workflow Accessible via Link and Publishing It

This workflow is currently accessible via link and published.

Anyone can view and import this workflow by visiting the following URL:

<http://galaxy.hpc.ufl.edu/u/moskalenko/w/lana-chip-peaks-on-hg19>

This workflow is publicly listed and searchable in Galaxy's [Published Workflows](#) section.

You can:

[Unpublish Workflow](#)

Removes this workflow from Galaxy's [Published Workflows](#) section so that it is not publicly listed or searchable.

[Disable Access to Workflow via Link and Unpublish](#)

## Published Workflows

[Advanced Search](#)

Name	Annotation	Owner	Community Rating	Community Tags	Last Updated ↓
<a href="#">LANA ChIP peaks on hg19</a>		moskalenko	★★★★★		2 minutes ago

## Published Histories

[Advanced Search](#)

Name	Annotation	Owner	Community Rating	Community Tags	Last Updated ↓
<a href="#">LANA ChIP peaks on hg19</a>	Peak calling on LANA ChIP-Seq data using Human chromosome 9 from hg19 build	moskalenko	★★★★★	<a href="#">chr9</a> <a href="#">hg19</a> <a href="#">peaks</a> <a href="#">lana</a> <a href="#">chip</a>	4 minutes ago





# Galaxy pages

Published Pages | [aun1](#) | Windshield Splatter

## Windshield splatter analysis with the Galaxy metagenomic pipeline: A live supplement

SERGEI KOSAKOVSKY POND<sup>1,2,\*</sup>, SAMIR WADHAWAN<sup>3,6\*</sup>, FRANCESCA CHIAROMONTE<sup>4</sup>, GURUPRASAD ANANDA<sup>1,3</sup>, WEN-YU CHUNG<sup>1,3,7</sup>, JAMES TAYLOR<sup>1,5</sup>, ANTON NEKRUTENKO<sup>1,3</sup> and THE GALAXY TEAM<sup>1\*</sup>

Correspondence should addressed to [SKP](#), [JT](#), or [AN](#).

### How to use this document

This document is a live copy of supplementary materials for [the manuscript](#). It provides access to the **exact** analyses and workflows discussed in the paper, so you can play with them by re-running, changing parameters, or even applying them to your own data. Specifically, we provide the two histories and one workflow found below. You can view these items by clicking on their name to expand them. You can also import these items into your Galaxy workspace and start using them; click on the green plus to import an item. To import workflows you must [create a Galaxy account](#) (unless you already have one) – a hassle-free procedure where you are only asked for a username and password.

This is the Galaxy history detailing the comparison of our pipeline to MEGAN:

[+](#) [Galaxy History | Galaxy vs MEGAN](#) [+](#) [↗](#)  
Comparison of Galaxy vs. MEGAN pipeline.

This is the Galaxy history showing a generic analysis of metagenomic data. (This corresponds to the "A complete metagenomic pipeline" section of the manuscript and **Figure 3A**):

[+](#) [Galaxy History | metagenomic analysis](#) [+](#) [↗](#)

This is the Galaxy workflow for generic analysis of metagenomic data. (This corresponds to the "A complete metagenomic pipeline" section of the manuscript and **Figure 3B**):

[+](#) [Galaxy Workflow | metagenomic analysis](#) [+](#) [↗](#)  
Generic workflow for performing a metagenomic analysis on NGS data.

### Accessing the Data

Windshield Splatter datasets analyzed in this manuscript can be accessed through this [Galaxy Library](#). From there they can be re-analyzed through Galaxy using the above workflows or downloaded.

### Supplemental Analysis

#### Comparison between Galaxy pipeline and Megan

(Use [this link](#) to see Galaxy history representing this analysis. Individual elements of this history are referred to as **History Item1**, **2** and so on using **bold** typeface)

#### About this Page

##### Author

aun1



##### Related Pages

[All published pages](#)

[Published pages by aun1](#)

##### Rating

Community  
(6 ratings, 5.0 average)



##### Tags

Community:

[megan](#)

[galaxy](#)

[paper](#)





# Summary

---

- ◆ Analyze data without the CLI
- ◆ Visualize the results
- ◆ Publish histories, workflows, and annotated pages
- ◆ Add new tools, get support @ HPC
- ◆ Focus on your science, not minutiae
- ◆ **UF Galaxy** – coming to a browser near you!



# Demo

 **Galaxy / UF HPC /** [Analyze Data](#) [Workflow](#) [Shared Data](#) [Help](#) [User](#)

**Tools** Options ▾

[Get Data](#)  
[Send 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](#)  
[Statistics](#)  
[Graph/Display Data](#)  
[Regional Variation](#)  
[Multiple regression](#)  
[Multivariate Analysis](#)  
[Evolution](#)  
[Motif Tools](#)  
[Multiple Alignments](#)  
[Metagenomic analyses](#)  
[FASTA manipulation](#)  
[NCBI BLAST+](#)  
[NGS: QC and manipulation](#)  
[NGS: Picard \(beta\)](#)



**UNIVERSITY of FLORIDA**

---

### UFL HPC Galaxy News:

- 2001-08-09: Prototype Galaxy Instance  
An instance of Galaxy Platform for Biological Research Computing was brought online at the University of Florida High-Performance Computing Center for testing and demonstration purposes. This instance is not available for public use, yet. However, you can email HPC or the biological applications support directly to request to be notified of its general availability.

---

The Galaxy project is supported in part by NSF, NHGRI, and the Huck Institutes of the Life Sciences.

**History** Options ▾

   
MACS hg19 0 bytes  
 Your history is empty. Click 'Get Data' on the left pane to start



# MACS demo

---

<http://galaxy.hpc.ufl.edu>



# MACS demo

---

<http://galaxy.hpc.ufl.edu>

## UF HPC Center Login

Username:

Password:

Login

[Request an account](#)

[Reset my password](#)




# History/Shared Data

The screenshot displays the Galaxy / UF HPC web interface. The top navigation bar includes 'Galaxy / UF HPC /', 'Analyze Data', 'Workflow', 'Shared Data' (highlighted with a red circle), 'Help', and 'User'. The left sidebar lists various tools under the 'Tools' header, including 'Get Data', 'Send 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', 'Statistics', 'Graph/Display Data', 'Regional Variation', 'Multiple regression', 'Multivariate Analysis', 'Evolution', 'Motif Tools', 'Multiple Alignments', 'Meta', 'FAST', 'NCBI', 'NGS', and 'NGS'. The main content area features the 'UF UNIVERSITY of FLORIDA' logo and a section titled 'UFL HPC Galaxy News:' with a bullet point for '2001-08-09: Prototype Galaxy Instance'. The right sidebar, titled 'History', shows 'Unnamed history' (highlighted with a red circle) and '0 bytes'. A message box states: 'Your history is empty. Click 'Get Data' on the left pane to start'.


<http://galaxy.hpc.ufl.edu>



# Shared Data

 **Galaxy / UF HPC /** [Analyze Data](#) [Workflow](#) **[Shared Data](#)** [Help](#) [Use](#)

## Data Libraries



[Advanced Search](#)

<a href="#">Data library name</a> ↓	<a href="#">Data library description</a>
<a href="#">GMS 6001 MACS Exercise</a>	HPC Intro and MACS exercise on 9/15/11
<a href="#">OM Testing</a>	Test data for Galaxy development



# MACS – Load data

## Data Library “GMS 6001 MACS Exercise”

MACS test data

<input type="checkbox"/> Name	Message	Uploaded By	Date	File Size
<input type="checkbox"/> 2010-12-14 7 hg19 aln sorted.bam ▾		om@hpc.ufl.edu	2011-09-13	1.6 Gb
<input type="checkbox"/> 2010-12-14 7 hhv8 aln sorted.bam ▾		om@hpc.ufl.edu	2011-09-13	1.4 Gb
<input checked="" type="checkbox"/> hg19.chr10.bam ▾		om@hpc.ufl.edu	2011-09-14	80.8 Mb
<input type="checkbox"/> hg19.chr11.bam ▾		om@hpc.ufl.edu	2011-09-14	82.5 Mb
<input type="checkbox"/> hg19.chr12.bam ▾		om@hpc.ufl.edu	2011-09-14	74.9 Mb
<input type="checkbox"/> hg19.chr13.bam ▾		om@hpc.ufl.edu	2011-09-14	50.9 Mb
<input type="checkbox"/> hg19.chr14.bam ▾		om@hpc.ufl.edu	2011-09-14	36.1 Mb
<input type="checkbox"/> hg19.chr15.bam ▾		om@hpc.ufl.edu	2011-09-14	48.1 Mb
<input type="checkbox"/> hg19.chr16.bam ▾		om@hpc.ufl.edu	2011-09-14	55.9 Mb
<input type="checkbox"/> hg19.chr17.bam ▾		om@hpc.ufl.edu	2011-09-14	64.5 Mb
<input type="checkbox"/> hg19.chr18.bam ▾		om@hpc.ufl.edu	2011-09-14	33.5 Mb
<input type="checkbox"/> hg19.chr19.bam ▾		om@hpc.ufl.edu	2011-09-14	39.6 Mb
<input type="checkbox"/> hg19.chr1.bam ▾		om@hpc.ufl.edu	2011-09-14	148.5 Mb
<input type="checkbox"/> hg19.chr20.bam ▾		om@hpc.ufl.edu	2011-09-14	38.5 Mb
<input type="checkbox"/> hg19.chr21.bam ▾		om@hpc.ufl.edu	2011-09-14	17.5 Mb
<input type="checkbox"/> hg19.chr22.bam ▾		om@hpc.ufl.edu	2011-09-14	16.9 Mb
<input type="checkbox"/> hg19.chr2.bam ▾		om@hpc.ufl.edu	2011-09-14	126.3 Mb
<input type="checkbox"/> hg19.chr2.sam ▾		om@hpc.ufl.edu	2011-09-14	488.0 Mb
<input type="checkbox"/> hg19.chr3.bam ▾		om@hpc.ufl.edu	2011-09-14	118.0 Mb
<input type="checkbox"/> hg19.chr4.bam ▾		om@hpc.ufl.edu	2011-09-14	85.7 Mb
<input type="checkbox"/> hg19.chr5.bam ▾		om@hpc.ufl.edu	2011-09-14	102.7 Mb
<input type="checkbox"/> hg19.chr6.bam ▾		om@hpc.ufl.edu	2011-09-14	65.7 Mb
<input type="checkbox"/> hg19.chr7.bam ▾		om@hpc.ufl.edu	2011-09-14	89.9 Mb
<input type="checkbox"/> hg19.chr8.bam ▾		om@hpc.ufl.edu	2011-09-14	85.9 Mb
<input type="checkbox"/> hg19.chr9.bam ▾		om@hpc.ufl.edu	2011-09-14	64.8 Mb

For selected datasets: Import to current history ▴

Go



# What's inside

Galaxy / UF HPC /

Analyze Data

Workflow

Shared Data

Help

User

Tools

Options

Get Data

Send 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

Statistics

Graph/Display Data

Regional Variation

Multiple regression

Multivariate Analysis

Evolution

Motif Tools

Multiple Alignments

Metagenomic analyses

FASTA manipulation

NCBI BLAST+

NGS: QC and manipulation

NGS: Picard (beta)

NGS: Mapping

NGS: Indel Analysis

NGS: RNA Analysis

NGS: SAM Tools

Filter SAM on bitwise flag values

Convert SAM to interval

SAM-to-BAM converts SAM format to BAM format

**BAM-to-SAM** converts BAM format to SAM format

Merge BAM Files merges BAM files

✓

The following job has been successfully added to the queue:

2: BAM-to-SAM on data 1: converted SAM

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

Options

MACS hg19

80.8 Mb

2: BAM-to-SAM on data 1: converted SAM

1: hg19.chr10.bam

@SQ SN:chrUn\_g1000248 LN:39786

@SQ SN:chrUn\_g1000249 LN:38502

@SQ SN:chrX LN:155270560

@SQ SN:chrY LN:59373566

@PG ID:Bowtie VN:0.12.5 CL:"bowtie --chunkmbs 1024 --sam --solexa1.3-quals -e 70 -M 1 -n 2 -l 28 --tryhard

HWUSI-EAS1654\_0011:7:101:15569:18265#0/1 16 chr10 60132 0 40M \* 0 0 0 TTAATTGACG

HWUSI-EAS1654\_0011:7:70:15927:14469#0/1 16 chr10 60156 0 40M \* 0 0 0 TTGAGTTCGGTTGAGTTT

HWUSI-EAS1654\_0011:7:77:2963:15680#0/1 16 chr10 60189 0 40M \* 0 0 0 TCTTCCACAAGGGATTGT

HWUSI-EAS1654\_0011:7:56:13734:7224#0/1 0 chr10 60478 0 40M \* 0 0 0 CGCCTTTGGAAGGAGCAT

HWUSI-EAS1654\_0011:7:27:17072:12387#0/1 0 chr10 60491 255 40M \* 0 0 0 AGCATTTATCCCCAGCA

HWUSI-EAS1654\_0011:7:41:17138:6404#0/1 0 chr10 60491 255 40M \* 0 0 0 AGCATATATCCCCAGCA

HWUSI-EAS1654\_0011:7:82:1846:20862#0/1 0 chr10 60491 255 40M \* 0 0 0 AGCATATATCCCCAGCA

HWUSI-EAS1654\_0011:7:96:15342:7286#0/1 16 chr10 60513 0 40M \* 0 0 0 TCCGGTTTTTTGAAGTCT

HWUSI-EAS1654\_0011:7:91:12232:19734#0/1 0 chr10 60618 0 40M \* 0 0 0 AATCTTTGTGTATACAT

HWUSI-EAS1654\_0011:7:4:3181:2553#0/1 0 chr10 60918 0 40M \* 0 0 0 GAAGTATCAATATGCCTT

HWUSI-EAS1654\_0011:7:43:15903:1917#0/1 16 chr10 61129 0 40M \* 0 0 0 TTTGTATTGGTAGGATA

HWUSI-EAS1654\_0011:7:11:8621:5776#0/1 16 chr10 61203 0 40M \* 0 0 0 ATGAGGCCCTCACTCTGT

HWUSI-EAS1654\_0011:7:44:10220:2559#0/1 0 chr10 62846 0 40M \* 0 0 0 ATACTGGGGAGGAGCTGT

HWUSI-EAS1654\_0011:7:95:15568:1582#0/1 16 chr10 63074 0 40M \* 0 0 0 TGAAAAGCCAATGGCTGG

HWUSI-EAS1654\_0011:7:32:3459:2766#0/1 0 chr10 63307 0 40M \* 0 0 0 AAAGGACATATAATCTTG

HWUSI-EAS1654\_0011:7:89:18639:10678#0/1 0 chr10 63307 0 40M \* 0 0 0 AAAGGACATATAATCTTG

HWUSI-EAS1654\_0011:7:109:19350:5114#0/1 16 chr10 63367 0 40M \* 0 0 0 TACAGCTGATGCTTTCTG

HWUSI-EAS1654\_0011:7:68:3139:9708#0/1 16 chr10 63523 0 40M \* 0 0 0 CCCAAAGATGGTTCACAT

HWUSI-EAS1654\_0011:7:27:15884:16777#0/1 0 chr10 63723 0 40M \* 0 0 0 CAGTCTTCAGCCCTAGAC

HWUSI-EAS1654\_0011:7:64:1626:5553#0/1 0 chr10 63915 0 40M \* 0 0 0 AGCTAATCAGGGAGGGGG

HWUSI-EAS1654\_0011:7:21:7450:15409#0/1 0 chr10 64133 0 40M \* 0 0 0 GACAGGGCTTTTGATTTA

HWUSI-EAS1654\_0011:7:67:16528:10957#0/1 0 chr10 64143 0 40M \* 0 0 0 TTGATTTAACCCTAATCCA

HWUSI-EAS1654\_0011:7:61:13906:4548#0/1 0 chr10 64190 0 40M \* 0 0 0 TATGAGCAAAAGTCTCCA

HWUSI-EAS1654\_0011:7:42:8178:16716#0/1 16 chr10 64272 0 40M \* 0 0 0 ATTCTAAAAGCCAGGAAP

HWUSI-EAS1654\_0011:7:64:10597:5725#0/1 16 chr10 64305 0 40M \* 0 0 0 TAATCTAGGAAAACCTCC

HWUSI-EAS1654\_0011:7:50:6555:17127#0/1 16 chr10 64375 0 40M \* 0 0 0 TGGGAAATTCATCACAAP

HWUSI-EAS1654\_0011:7:119:10260:12806#0/1 0 chr10 64633 0 40M \* 0 0 0 0 TGAGGGAAG

HWUSI-EAS1654\_0011:7:117:5789:14945#0/1 0 chr10 64648 0 40M \* 0 0 0 GTCATTTTCAGACAAAAC

HWUSI-EAS1654\_0011:7:49:10018:10089#0/1 0 chr10 64844 0 40M \* 0 0 0 AAAGAGCATGATGAAAGC

HWUSI-EAS1654\_0011:7:81:5640:15183#0/1 0 chr10 65212 0 40M \* 0 0 0 ACTGGAGCTCCCAATTTT

HWUSI-EAS1654\_0011:7:67:19633:13182#0/1 0 chr10 65483 0 40M \* 0 0 0 GCCATAAAATGAGTCTCA

HWUSI-EAS1654\_0011:7:77:17790:3617#0/1 0 chr10 65626 0 40M \* 0 0 0 CTGAATGAGCATTGGGCC

HWUSI-EAS1654\_0011:7:60:8645:3117#0/1 16 chr10 65700 0 40M \* 0 0 0 TATCGAACCCTCTGGGAT

HWUSI-EAS1654\_0011:7:60:6931:13228#0/1 0 chr10 66024 0 40M \* 0 0 0 TCACTAAGAAATGAAAC

HWUSI-EAS1654\_0011:7:76:1871:15186#0/1 0 chr10 66045 0 40M \* 0 0 0 GATATTACAACTGACACC

HWUSI-EAS1654\_0011:7:114:9163:3182#0/1 0 chr10 66273 0 40M \* 0 0 0 CGGATTTCACAGCAGAA



# MACS (NGS: Peak Calling)

The screenshot displays the Galaxy / UF HPC web interface. The top navigation bar includes links for Analyze Data, Workflow, Shared Data, Help, and User. The left sidebar, titled 'Tools', lists various bioinformatics tools. The 'NGS: Peak Calling' tool is highlighted with a red box, and a red arrow points from it to a detailed view of the tool in a separate box. The main content area features the University of Florida logo and a news section titled 'UFL HPC Galaxy News:'. The right sidebar, titled 'History', shows a list of datasets, with '1: hg19.chr10.bam' highlighted in green.

**Galaxy / UF HPC /** Analyze Data Workflow Shared Data Help User

**Tools** Options ▾

- [Get Data](#)
- [Send 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](#)
- [Statistics](#)
- [Graph/Display Data](#)
- [Regional Variation](#)
- [Multiple regression](#)
- [Multivariate Analysis](#)
- [Evolution](#)
- [Motif Tools](#)
- [Multiple Alignments](#)
- [Metagenomic analyses](#)
- [FASTA manipulation](#)
- [NCBI BLAST+](#)
- [NGS: QC and manipulation](#)
- [NGS: Picard \(beta\)](#)
- [NGS: Mapping](#)
- [NGS: Indel Analysis](#)
- [NGS: RNA Analysis](#)
- [NGS: SAM Tools](#)
- [NGS: GATK Tools](#)
- NGS: Peak Calling**
- [NGS: Simulation](#)
- [SNP/WGA: Data; Filters](#)

**Options** ▾

**History** Options ▾

MACS hg19 80.8 Mb

**1: hg19.chr10.bam**

**UFL HPC Galaxy News:**

- 2001-08-09: Prototype Galaxy Instance**  
An instance of [Galaxy Platform](#) for Biological Research Computing was brought online at the University of Florida [High-Performance Computing Center](#) for testing and demonstration purposes. This instance is not available for public use, yet. However, you can email [HPC](#) or the [biological applications support](#) directly to request to be notified of its general availability.

The Galaxy project is supported in part by [NSF](#), [NHGRI](#), and the [Huck Institutes of the Life Sciences](#).

**NGS: Peak Calling**

- [MACS](#) Model-based Analysis for ChIP-Seq



# Submission form

Tools

Options ▾

[Get Data](#)  
[Send 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](#)  
[Statistics](#)  
[Graph/Display Data](#)  
[Regional Variation](#)  
[Multiple regression](#)  
[Multivariate Analysis](#)  
[Evolution](#)  
[Motif Tools](#)  
[Multiple Alignments](#)  
[Metagenomic analyses](#)  
[FASTA manipulation](#)  
[NCBI BLAST+](#)  
[NGS: QC and manipulation](#)  
[NGS: Picard \(beta\)](#)  
[NGS: Mapping](#)  
[NGS: Indel Analysis](#)  
[NGS: RNA Analysis](#)  
[NGS: SAM Tools](#)  
[NGS: GATK Tools](#)  
[NGS: Peak Calling](#)  

- MACS Model-based Analysis for ChIP-Seq
- CCAT Control-based ChIP-seq Analysis Tool

MACS

Treatment file:  
1: hg19.chr10.bam ▾

Input file:  
Selection is Optional ▾

Format:  
Auto ▾

Effective Genome Size:  
Human (hg19) ▾

Tag size (Optional):  
25

P-Value:  
1e-05

Keep duplicate tags at the exact same location?:  
☐ Keep ALL  
☐ Auto by Binomial  
☒ Keep Single

Use Model?:  
True ▾

small fold enrichment for model building:  
10

large fold:  
30



Advanced Options:  
No ▾

Diagnosis Report:  
No ▾

Execute

History




Options ▾






MACS hg19

400.5 Mi

2: BAM-to-SAM on data 1: converted SAM



1: hg19.chr10.bam





# MACS options

---

- Basic:
  - Treatment file: **Your alignment file – choose BAM file**
  - Effective genome size: **Human (hg19) – must set once**
- Advanced:
  - Use model or shift size
  - Model - fold enrichment (small and large): 10:30
  - Bandwidth – scan bandwidth size for model or  $\frac{1}{2}$  window size without the model: default is 300



# Submit the job to cluster

**Galaxy / UF HPC /** Analyze Data Workflow Shared Data Help User

**Parameters:**

300

**Use Lambda?:**  
☒ True  
☐ False

**Small Lambda:**  
1000

**Large Lambda:**  
10000

**Generate a wig file?:**  
☒ Yes  
☐ No

**Diagnosis Report:**  
No


**Execute**



# Cluster job run

**Galaxy / UF HPC /** Analyze Data Workflow Shared Data Help User

**Tools** Options ▾  
[Get Data](#)  
[Send 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](#)  
[Statistics](#)  
[Graph/Display Data](#)  
[Regional Variation](#)  
[Multiple regression](#)  
[Multivariate Analysis](#)  
[Evolution](#)  
[Motif Tools](#)

 The following job has been successfully added to the queue:

3: MACS peaks on hg19.chr10.bam

4: MACS summits on hg19.chr10.bam

5: MACS xls on hg19.chr10.bam

6: MACS wiggle on hg19.chr10.bam

7: MACS job log on hg19.chr10.bam




8: MACS diagnosis report on hg19.chr10.bam

You can check the status of queued jobs and view the resulting data by refreshing the **History** pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.




**Options** ▾  

MACS 400.5 Mb hg19




8:

   [MACS diagnosis report on hg19.chr10.bam](#)




7:

   [MACS job log on hg19.chr10.bam](#)




6:

   [MACS wiggle on hg19.chr10.bam](#)




5:

   [MACS xls on hg19.chr10.bam](#)

4:

   [MACS summits on hg19.chr10.bam](#)

3:

   [MACS peaks on hg19.chr10.bam](#)



# Job completion

**Galaxy / UF HPC /** Analyze Data Workflow Shared Data Help User

**Tools** Options ▾  
[Get Data](#)  
[Send 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](#)  
[Statistics](#)  
[Graph/Display Data](#)  
[Regional Variation](#)  
[Multiple regression](#)  
[Multivariate Analysis](#)  
[Evolution](#)

✓

The following job has been successfully added to the queue:

- 3: MACS peaks on hg19.chr10.bam
- 4: MACS summits on hg19.chr10.bam
- 5: MACS xls on hg19.chr10.bam
- 6: MACS wiggle on hg19.chr10.bam
- 7: MACS job log on hg19.chr10.bam
- 8: MACS diagnosis report on hg19.chr10.bam

You can check the status of queued jobs and view the resulting data by refreshing the **History** pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

**Options** ▾  
MACS 498.3 Mb hg19  

8: MACS [diagnosis report on hg19.chr10.bam](#)

7: MACS [job log on hg19.chr10.bam](#)

6: MACS [wiggle on hg19.chr10.bam](#)

5: MACS [xls on hg19.chr10.bam](#)

4: MACS [summits on hg19.chr10.bam](#)

3: MACS [peaks on hg19.chr10.bam](#)



# Build a genome browser track

**Tools** Options ▾

- [Get Data](#)
- [Send 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](#)
- [Statistics](#)
- [Graph/Display Data](#)
  - [Histogram](#) of a numeric column
  - [Scatterplot](#) of two numeric columns
  - [Bar chart](#) for multiple columns
  - [Plotting tool](#) for multiple series and graph types
  - [Boxplot](#) of quality statistics
  - [GMAJ](#) Multiple Alignment Viewer
  - [LAJ](#) Pairwise Alignment Viewer
  - [Build custom track](#) for UCSC genome browser

**Build custom track**

Tracks

[Add new Track](#)

[Execute](#)

**i** This tool allows you to build custom tracks using datasets in your history for the UCSC genome browser. You can view these custom tracks on the UCSC genome browser by clicking on **display at UCSC main/test** link in the history panel of the output dataset.

**!** Please note that this tool requires **all input datasets(tracks) to have the same genome build**. The tool throws an error when this requirement is not met. You may then have to choose a valid dataset or remove invalid tracks.

**History** Options ▾

MACS hg19 498.3 Mb

- 8: [MACS diagnosis report on hg19.chr10.bam](#)
- 7: [MACS job log on hg19.chr10.bam](#)
- 6: [MACS wiggle on hg19.chr10.bam](#)
- 5: [MACS xls on hg19.chr10.bam](#)
- 4: [MACS summits on hg19.chr10.bam](#)
- 3: [MACS peaks on hg19.chr10.bam](#)
- 2: [BAM-to-SAM on data 1: converted SAM](#)
- 1: [hg19.chr10.bam](#)



# Submit a track build job

---

### Build custom track

**Tracks**

**Track 1**

**Dataset:**  
4: MACS summits on hg19.chr10.bam

**name:**  
Chr10LANA

**description:**

**Color:**  
Black

**Visibility:**  
Dense

Remove Track 1

Add new Track

Execute

**i** This tool allows you to build custom tracks using datasets in your history for the UCSC genome browser. You can view these custom tracks on the UCSC genome browser by clicking on **display at UCSC main/test** link in the history panel of the output dataset.

---

**!** Please note that this tool requires **all input datasets(tracks) to have the same genome build**. The tool throws an error when this requirement is not met. You may then have to choose a valid dataset or remove invalid tracks.



# Open the track

ToolsOptions ▾



[Get Data](#)  
[Send 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](#)  
[Statistics](#)  
[Graph/Display Data](#)

- [Histogram](#) of a numeric column
- [Scatterplot](#) of two numeric columns




track name="Chr10LANA" description="User Supplied Track (from Galaxy)" color=0,0,255 visibility=1






chr10	309835	311665	MACS_peak_1	134.30
chr10	374946	376165	MACS_peak_2	87.06
chr10	382566	385025	MACS_peak_3	54.46
chr10	439141	440977	MACS_peak_4	53.43
chr10	1030693	1036216	MACS_peak_5	68.77
chr10	1093464	1096423	MACS_peak_6	126.75
chr10	1196247	1198127	MACS_peak_7	68.34
chr10	3237793	3240452	MACS_peak_8	62.17
chr10	3268557	3270788	MACS_peak_9	56.84
chr10	3355912	3357691	MACS_peak_10	98.16
chr10	3371713	3374115	MACS_peak_11	

HistoryOptions ▾



MACS hg19498.3 Mb

9: Build custom track on data 4 and data 3

995 lines, 3 comments  
format: customtrack, database: ?  
Info: Generated a custom track containing 2 subtracks.  


1	2
track name="Chr10LANA" description="User Supplied Track (from Galaxy)"	
chr10	309835
chr10	374946
chr10	382566
chr10	439141
chr10	1030693



# Genome Browser

Tools Options ▾

Get Data

- Upload File from your computer
- UCSC Main table browser
- UCSC Archaea table browser
- BX main browser
- BioMart Central server
- GrameneMart Central server
- Flymine server
- modENCODE fly server
- modENCODE modMine server
- Ratmine server
- YeastMine server
- modENCODE worm server
- Wormbase server
- EuPathDB server
- EncodeDB at NHGRI
- EpiGRAPH server

Send Data

ENCODE Tools

Lift-Over

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Sequences

Fetch Alignments

Home Genomes **Genome Browser** Blat Tables Gene Sorter PCR Session FAQ

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichment, send the data to [GREAT](#). Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with this data.

clade: Mammal ▾

genome: Human ▾

assembly: Feb. 2009 (GRCh37/hg19) ▾

group: Custom Tracks ▾

track: Chr9 hg19 LANA MACS peaks ▾

manage custom tracks

track hubs

table: ct\_Chr9hg19LANAMACSpeaks\_3943 ▾

describe table schema

region: ☐ genome ☒ position

chr9:1-140127172

lookup

define regions

identifiers (names/accessions):

paste list

upload list

filter:

create

intersection:

create

correlation:

create

output format:

BED - browser extensible data ▾

Send output to ☒ [Galaxy](#) ☐ [GREAT](#)

output file:

(leave blank to keep output in browser)

file type returned:

☒ plain text ☐ gzip compressed

get output

summary/statistics

To reset **all** user cart settings (including custom tracks), [click here](#).



# Add a custom track

The screenshot displays the UCSC Genome Browser interface for Human Feb. 2009 (GRCh37/hg19) Assembly. The top navigation bar includes links for Home, Genomes, Blat, Tables, Gene Sorter, PCR, DNA, Convert, and Ensembl. The main title is "UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly". Below the title, there are navigation controls for moving and zooming. The position/search field shows "chr10:1-135,534,747" with a "gene" link and a "jump" button. The size is indicated as 135,534,747 bp. A chromosome ideogram for chromosome 10 is shown, with a red box highlighting the region from p14 to p13. Below the ideogram, a scale bar for chromosome 10 is displayed. The main content area contains instructions on how to use the browser, including a red box around the "add custom tracks" button. At the bottom, there is a section titled "Mapping and Sequencing Tracks" with a "refresh" button and several track categories: Base Position, Chromosome Band, STS Markers, FISH Clones (with a count of 18), Recomb Rate, and Map Contigs. Each category has a "hide" button and a small upward arrow.

Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl

**UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly**

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search chr10:1-135,534,747 [gene](#) jump clear size 135,534,747 bp. configure

chr10 (p15.3-q26.3) p14 p13 q21.1

Scale chr10: 50 Mb 50000000 100000000

move start < 2.0 > move end < 2.0 >

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

track search default tracks default order hide all **add custom tracks** track hubs configure reverse resize

refresh

collapse all Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes. expand all

**Mapping and Sequencing Tracks** refresh

[Base Position](#) [Chromosome Band](#) [STS Markers](#) **18** [FISH Clones](#) [Recomb Rate](#) [Map Contigs](#)

hide hide hide hide hide



# Paste track data

[Home](#) [Genomes](#) [Genome Browser](#) [Blat](#) [Tables](#) [Gene Sorter](#) [PCR](#) [Se](#)

## Add Custom Tracks

clade  genome  assembly

Display your own data as custom annotation tracks in the browser. Data must be formatted in [GFF](#), [GTF](#), [WIG](#), [bigWig](#), [MAF](#), [BAM](#), [BED detail](#), [Personal Genome SNP](#), or [PSL](#) formats. To [track](#) and [browser](#) line attributes as described in the [User's Guide](#). URLs for data in the bigBed and embedded in a track line in the box below. Publicly available custom tracks are listed [here](#). Example

Paste URLs or data: Or upload:  no file selected

```
track name="Chr10LANA" description="User Supplied Track (from Galaxy)"
color=0,0,255 visibility=1
chr10    309835    311665    MACS_peak_1  134.30
chr10    374946    376165    MACS_peak_2  87.06
chr10    382566    385025    MACS_peak_3  54.46
chr10    439141    440977    MACS_peak_4  53.43
chr10    1030693   1036216   MACS_peak_5  68.77
```

Optional track documentation: Or upload:  no file selected



# View track

Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/P

## UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search chr10:310,802-310,803 [gene](#)  jump clear size 2 bp. configure

chr10 (p15.3) p14 p13 q21.1 21.3 23.1 25.1

Scale chr10: 1 bases | 310802 | T

Chr10LANA My Custom Track

move start < 2.0 > move end < 2.0 >

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

track search default tracks default order hide all manage custom tracks track hubs configure reverse resize refresh

collapse all Use drop-down controls below and press refresh to alter tracks displayed. expand all

Tracks with lots of items will automatically be displayed in more compact modes.

[Chr10LANA](#) My Custom Track Custom Tracks refresh

drag to reorder

dense ⇅



# Zoom in, pan, etc.

omes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF

## UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search chr10:86,249,385-97,585,01 [gene](#) jump clear size 11,335,633 bp. configure

chr10 (q23.1-q24.1) p14 p13 q21.1

Scale chr10: 5 Mb | 88000000 89000000 90000000 91000000 92000000 93000000 94000000 95000000 96000000 97000000 |

Chr10LANA My Custom Track

move start < 2.0 > Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. move end < 2.0 >

track search default tracks default order hide all manage custom tracks track hubs configure reverse resize refresh

collapse all Use drop-down controls below and press refresh to alter tracks displayed. expand all

Custom Tracks refresh

Chr10LANA dense

Mapping and Sequencing Tracks refresh



# Thank you!

---

- <http://wiki.hpc.ufl.edu>
- <https://fisher.bioinformatics.ufl.edu>
- <http://hpc.ufl.edu/support>
  - Frequently Asked Questions
  - Account set up and maintenance
  - Problem report submission
  - [om@hpc.ufl.edu](mailto:om@hpc.ufl.edu) - Biological applications support
  - [magitz@ufl.edu](mailto:magitz@ufl.edu) - Training

