

UF Research Computing

Introduction to Galaxy at UF HPC

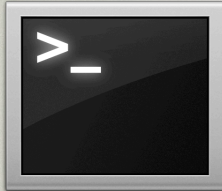
Matt Gitzendanner

Oleksandr Moskalenko

Today's research computing




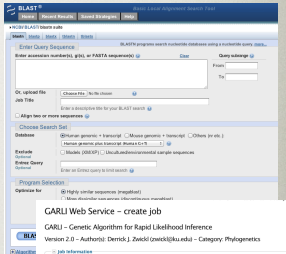
Approaches

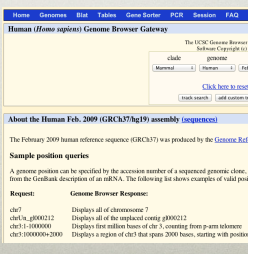


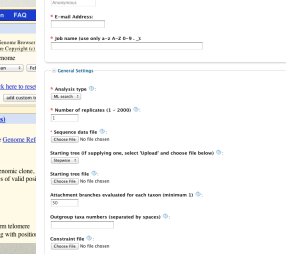
```

magitz@submit1:/apps/RAxML/7.3.0 --- bash -- 80x24
[Submit1:05369] *** Process received signal ***
[Submit1:05369] Signal: Segmentation fault (11)
[Submit1:05369] Signal code: Address not mapped (1)
[Submit1:05369] Failing at address: (nil)
[Submit1:05369] [ 0] /lib64/libpthread.so.0 [0x3ccfe0eb10]
[Submit1:05369] [ 1] /lib64/libc.so.6 [0x3ccee7fd5]
[Submit1:05369] [ 2] /lib64/libc.so.6[_libc_mlock+0x1367] [0x3ccee74d7]
[Submit1:05369] [ 3] /lib64/ld-linux-x86-64.so.2[_tls_get_addr+0xb1] [0x3ccee0f491]
[Submit1:05369] [ 4] /lib64/libselinux.so.1 [0x3cd060f57c]
[Submit1:05369] [ 5] /lib64/libselinux.so.1 [0x3cd060450f]
[Submit1:05369] *** End of error message ***
Segmentation fault
[magitz@submit1 7.3.0]# ls
roomHPC-HYBRID-SSE3 roomHPC-MPI roxmHPC-PTHREADS-SSE3
[magitz@submit1 7.3.0]# ls -l
total 3416
-rwxr-xr-x 1 magitz soltis 1168877 Sep  9 07:09 roomHPC-HYBRID-SSE3
-rwxr-xr-x 1 molax uffpc 1162832 Jan 31 20:38 roomHPC-MPI
-rwxr-xr-x 1 magitz soltis 1149164 Sep  9 07:07 roxmHPC-PTHREADS-SSE3
[magitz@submit1 7.3.0]# Read from remote host submit.hpc.ufl.edu: Operation time
d out
Connection to submit.hpc.ufl.edu closed.
Voyager-II:- matt$
                    
```

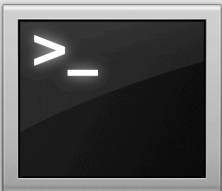









Command Line Environment




Head node

Login to head node



Scheduler

Interactive session or batch submission

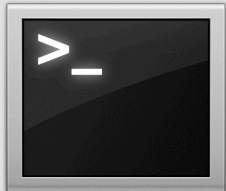


Computing resources

Your job runs on the cluster

Different approaches, same result

Head node



Scheduler



Interactive session or batch submission

Computing resources

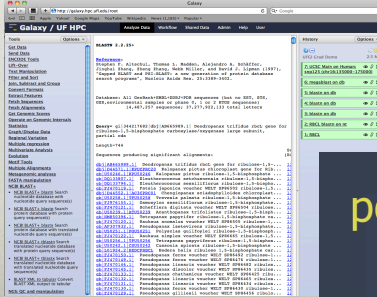


Your job runs on the cluster

What is Galaxy?



Galaxy Provides Life Support for NGS Exploration



powered by Galaxy



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: 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 (chr1:156690-165921)
- 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

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:

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 Analysis Workspace

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:

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

chr9	Start	End	Peak Name	Score
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	3899982	3899983	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



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 or copy files to HPC
 - Load from within Galaxy
 - http://wiki.hpc.ufl.edu/index.php/Galaxy_Data_Import
- ◆ External data
 - UCSC table browser
 - EuPathDB
 - Biomart
 - EncodeDB
 - interMine / modMine
 - 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

search

Advanced Search

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

For 0 selected groups:

Roles

search

Advanced Search

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

Users

search

Advanced Search

<input type="checkbox"/> Email ↓	User Name	Groups	Roles	External	Last Login
<input type="checkbox"/> aedison@ufl.edu	aedison	0	1	yes	Sep 15, 2011
<input type="checkbox"/> bostwick@ufl.edu	bostwick	0	1	yes	Sep 15, 2011
<input type="checkbox"/> cgraves3@ufl.edu	cgraves3	0	1	yes	Sep 15, 2011
<input type="checkbox"/> cjeffrey@ufl.edu	cjeffrey	0	1	yes	Sep 15, 2011
<input type="checkbox"/> colltd3@ufl.edu	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
- ◆ RNA-Seq
- ◆ ChIP-Seq
- ◆ GATK
- ◆ Phylogenetics ???



A galaxy of tools

GS: 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

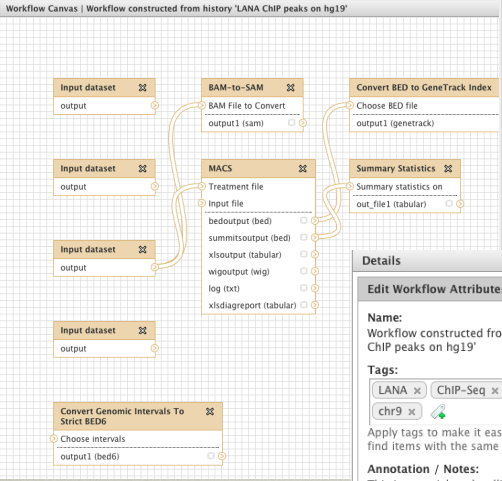
21: hg19.chr5.bam

20: hg19.chr4.bam



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

Edit Workflow Attributes

Name: Workflow constructed from history 'LANA CHIP peaks on hg19'

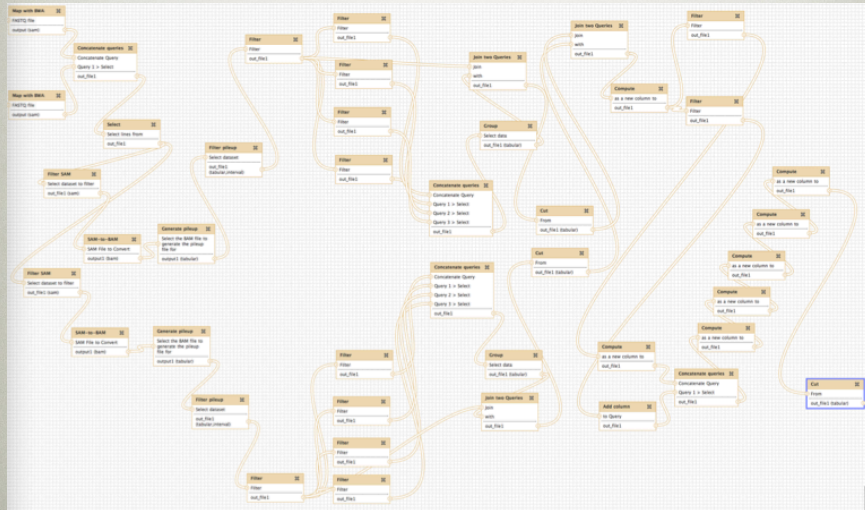
Tags: LANA, CHIP-Seq, hg19, chr9

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) p14.013 p21.1

Scale chr10: 50 kb | 0000000000 1000000000

move start 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

Base Position Chromosome Band STS Markers FISH Clones Recomb Rate Map Contigs

hide 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.uff.edu/#!/moskalenko/w/lan-a-chip-peaks-on-hg19>

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

You can:

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

Published Workflows

search name, annotation, owner, and tag

[Advanced Search](#)


Name	Annotation	Owner	Community Rating	Community Tags	Last Updated ↓
LANA ChIP peaks on hg19		moskalenko	★★★★★		2 minutes ago

Published Histories

search name, annotation, owner, and tag

[Advanced Search](#)

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



Galaxy pages

Published Pages | aun1 | Windshield Splatter

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

SERGEI KOSAKOVSKIY POND^{1,2}, SAMIR WADHWANI^{3*}, FRANCESCA CHIAROMONTE⁴, GURUPRASAD ANANDA^{1,3}, WEN-YU CHENG^{1,3,7}, JAMES TAYLOR^{1,3}, ANTON NEKRITENKO^{1,3} and THE GALAXY TEAM^{1*}

Correspondence should be addressed to S.W., E. = SW.

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 Item 1, 2 and so on using bold typeface)

About this Page

Author: aun1

Related Pages: [All published pages](#), [Published pages by aun1](#)

Rating: Community (8 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

The screenshot shows the Galaxy web interface for UF HPC. The top navigation bar includes 'Galaxy / UF HPC / Analyze Data Workflow Shared Data Help User'. The left sidebar lists various tools under the 'Tools' section, 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', 'Metagenomic analyses', 'FASTA manipulation', 'NCBI BLAST+', 'NGS: QC and manipulation', and 'NCS: Picard (beta)'. The main content area features the University of Florida logo and a news item titled 'UFL HPC Galaxy News' with a bullet point: '2001-08-09: Prototype Galaxy Instance'. Below the news item, it states: '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.' At the bottom, it mentions: 'The Galaxy project is supported in part by NSF, NHGRI, and the Huck Institutes of the Life Sciences.' The right sidebar shows a 'History' section with a message: '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:

[Request an account](#)

[Reset my password](#)

History/Shared Data

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
- Meta
- FAST
- NCBI
- NGS

UF
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

History Options

Unnamed history 0 bytes

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 User

Data Libraries

search dataset name, info, message, dbke

Advanced Search

Data library name ↓	Data library description
GMS 6001 MACS Exercise	HPC Intro and MACS exercise on 9/15/11
QM Testing	Test data for Galaxy development

MACS (NGS: Peak Calling)

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

UNIVERSITY of FLORIDA

UF 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, NHGRJ, and the Huck Institutes of the Life Sciences.

History Options

MACS hg19 80.8 Mb

1: hg19.chr10.bam

NGS: Peak Calling

- MACS Model-based Analysis for ChIP-Seq

Submission form

Tools Options

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 Mb

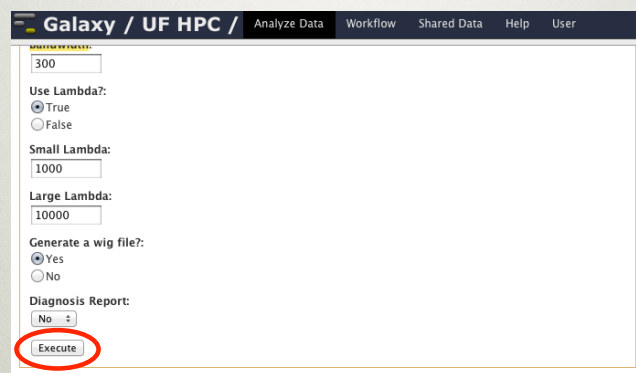
2: BAM-to-SAM on data 1: converted SAM

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



The screenshot shows the Galaxy / UF HPC interface with the following options:

- Bandwidth: 300
- Use Lambda?: True False
- Small Lambda: 1000
- Large Lambda: 10000
- Generate a wig file?: Yes No
- Diagnosis Report: No
- Execute** (button circled in red)

Cluster job run

The screenshot shows the Galaxy / UF HPC interface. The main panel displays a green notification box with a checkmark, indicating that a job has been successfully added to the queue. The job details are as follows:

- 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

Below the job list, a message states: "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."

The right-hand pane shows the job's progress, with a status of "MACS 400.5 Mb hg19". The job history pane on the right lists the following jobs:

- 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

The screenshot shows the Galaxy / UF HPC interface. The main panel displays a green notification box with a checkmark, indicating that a job has been successfully added to the queue. The job details are as follows:

- 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

Below the job list, a message states: "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."

The right-hand pane shows the job's progress, with a status of "MACS 498.3 Mb hg19". The job history pane on the right lists the following jobs:

- 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

The screenshot shows the UCSC Genome Browser interface for building a custom track. On the left, the 'Tools' sidebar lists various utilities, with 'Build custom track for UCSC genome browser' circled in red. The central 'Build custom track' panel features a 'Tracks' section with an 'Add new Track' button (circled in red) and an 'Execute' button. Below this is an information box explaining that all input datasets must share the same genome build. On the right, the 'History' panel shows a list of tracks, with '1: hg19.chr10.bam' at the bottom.

Submit a track build job

This screenshot provides a detailed view of the 'Build custom track' form. Under the 'Tracks' heading, 'Track 1' is configured with the following settings:

- Dataset:** 4: MACS summits on hg19.chr10.bam
- name:** Chr10LANA
- description:** (empty field)
- Color:** Black
- Visibility:** Dense

 The form includes 'Remove Track 1', 'Add new Track', and 'Execute' buttons. An information box at the bottom reiterates the requirement for all tracks to use the same genome build.

Open the track

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

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

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
chr10	309835
chr10	374946
chr10	382566
chr10	439141
chr10	1030693

Genome Browser

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

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 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: genome: assembly:

group: track: [manage custom tracks](#)

track hubs: [describe table schema](#)

region: genome position [lookup](#) [define regions](#)

identifiers (names/accessions): [paste list](#) [upload list](#)

filter: [create](#)

intersection: [create](#)

correlation: [create](#)

output format: 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 shows the UCSC Genome Browser interface for Human Feb. 2009 (GRCh37/hg19) Assembly. The main navigation bar includes Home, Genomes, Blat, Tables, Gene Sorter, PCR, DNA, Convert, and Ensembl. The search bar shows the position chr10:1-135,534,747. Below the search bar, there are navigation controls for zooming and moving. The 'add custom tracks' button is highlighted with a red box. Below the navigation controls, there are instructions for using the browser and a 'refresh' button. The 'Mapping and Sequencing Tracks' section is visible at the bottom, with various tracks like Base Position, Chromosome Band, STS Markers, FISH Clones, Recomb Rate, and Map Contigs.

Paste track data

The screenshot shows the 'Add Custom Tracks' form in the UCSC Genome Browser. The form includes dropdown menus for 'clade' (Mammal), 'genome' (Human), and 'assembly' (Feb. 2009 (GRCh37/hg19)). Below the dropdowns, there is a text area for pasting URLs or data. The text area contains the following track data:

```
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
```

The 'Submit' button is highlighted with a red circle. Below the text area, there is an 'Optional track documentation' section with a 'Choose File' button and 'no file selected' text.

View track

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

position/search chr10:310,802-310,803 gene jump clear size 2 bp. configure

Scale: 1 bases

Chr10LANA My Custom Track

move start < 2.0 > move end < 2.0 >

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

collapse all expand all

My Custom Track drag to reorder Custom Tracks refresh

Chr10LANA

dense

Zoom in, pan, etc.

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

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

Scale: 5 Mb

Chr10LANA My Custom Track

move start < 2.0 > move end < 2.0 >

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

collapse all expand all

Custom Tracks refresh

Chr10LANA

dense

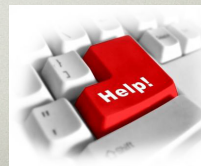
Mapping and Sequencing Tracks refresh

Help!!!

How to get help

◆ Asking for help

- Support Request Tickets
 - <http://support.hpc.ufl.edu>
 - Use for everything - not just software bugs but for any questions or help requests
 - Searchable database of solutions
- When you don't have access to web
 - support@hpc.ufl.edu
 - om@hpc.ufl.edu (Biological Support)
 - magitz@ufl.edu (Bio training and Q/A)



Documentation

◆ UF HPC Encyclopedia

- <http://wiki.hpc.ufl.edu>
 - Documents on hardware and software resources
 - User guides
 - Sample submission scripts
 - Research-specific sections
- <http://hpc.ufl.edu/support>
 - Frequently Asked Questions
 - Account set up and maintenance



Training Schedule

- ✓ Jan 26: Intro to UFHPC, getting started
- ✓ Feb 2: Galaxy overview and basic usage
- ◆ Feb 9: Intro to Linux shell
- ◆ Feb 16: No session
- ◆ Feb 23: Modules and basic submission scripts
- ◆ Mar 1: Working with NGS data
- ◆ Mar 8: No session – Spring break
- ◆ Mar 15: Perl intro
- ◆ Mar 22: Perl in practice with Gordon Burleigh
- ◆ Mar 29: Statistical analyses at HPC intro
- ◆ Apr 5: R and SAS in practice with TBA
- ◆ Apr 12: TBA
- ◆ Apr 19: TBA

Thank
you!