



Introduction to Galaxy

Matt Gitzendanner magitz@ufl.edu
Oleksandr Moskalenko om@hpc.ufl.edu

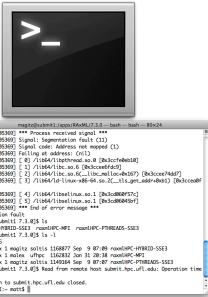



Today's research computing





Approaches









Cluster basics

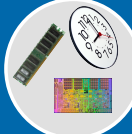
User interaction

Galaxy




Login node
(Head node)

Scheduler





Tell the scheduler
what you want to do


Compute resources




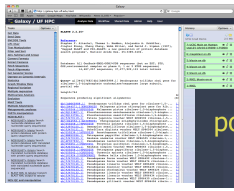
Your job runs on the cluster



What is Galaxy?



Galaxy Provides Life Support for NGS Exploration






powered by **Galaxy**

Galaxy: Data intensive biology for everyone

- ▶ Accessible, reproducible, transparent computational biology
- ▶ galaxy.hpc.ufl.edu
 - Local instance of Galaxy
 - Faster access to storage, easier upload
 - Local compute resources
 - Local control

Getting Data into Galaxy

- Upload a file from your computer
 - Direct upload (<2GB)
 - For large files: 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
 - Biomart
 - InterMine / modMine
 - EuPathDB
 - EncodeDB
 - EpiGRAPH
 - FlyMine
 - GrameneMart...

UF Information Technology | www.it.ufl.edu

Data Libraries

Data Library "GMS 6001 MACS Exercise"

Name	Message	Uploaded By	Date	File Size
2010-12-14 7_16133_ahp_sorted.bam		omihpc@ufl.edu	2011-09-15	1.8 GB
2010-12-14 7_16133_ahp_sorted.bam		omihpc@ufl.edu	2011-09-15	1.4 GB
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	80.8 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	82.5 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	74.9 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	50.9 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	36.1 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	48.1 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	55.9 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	64.5 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	33.5 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	70.6 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	145.5 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	38.5 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	17.5 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	16.9 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	126.3 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	448.0 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	118.0 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	85.7 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	102.7 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	67.8 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	89.9 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	65.9 Mb
hg19.chr11.bam		omihpc@ufl.edu	2011-09-14	64.8 Mb

UF Information Technology | www.it.ufl.edu

Data Access Control

Roles associated with new group
HPC test CHIP-seq analyses

Name	Users	Roles
HPC	0	2
Taylor HPC Lab	2	1

Users associated with new group

Name	Description	Type	Groups
HPC	Role for group HPC	system	1
HPC-test-CHIP-seq-analyses	Test analyses of CHIP-seq data	admin	1

Users

Email	User Name	Groups	Roles	External	Last Login
adecision@ufl.edu	adecision	0	1	yes	Sep 15, 2011
bozwick@ufl.edu	bozwick	0	1	yes	Sep 15, 2011
oguzonur@ufl.edu	oguzonur	0	1	yes	Sep 15, 2011
cliffrey@ufl.edu	cliffrey	0	1	yes	Sep 15, 2011
coltr@ufl.edu	coltr	0	1	yes	Sep 15, 2011

UF Information Technology | www.it.ufl.edu

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

UF Information Technology | www.it.ufl.edu

A galaxy of tools

QC and manipulation ILLUMINA DATA FASTQ Converter FASTQ quality filters FASTQ splitter FASTQ joiner FASTQ Summary Statistics by column NGS QC DATA Build base quality distribution Select high quality segments Combine FASTA and QUAL into FASTQ ABI-SOLID DATA Convert SOLID output to Fastq Compute quality statistics for SOLID data Draw quality score boxplot for SOLID data GENOMIC FASTQ MANIPULATION Filter FASTQ reads by quality score and length FASTQ Trimmer by column FASTQ Quality Trimmer by sliding window	Metagenomic analysis Human Genome Variation EMBOSS NGS TOOLBOX BETA NGS QC and manipulation NGS Mapping Map with Bowtie for Illumina Map with BWA for Illumina NGS QC DATA Latex map short reads against reference sequence Mapblast compare short reads against tips, nt, and wgs databases ABI-SOLID DATA Parse Blast XML output Convert SOLID output to Fastq Map with Bowtie for SOLID NGS SAM Tools NGS: InDel Analysis NGS: Peak Calling NGS: RNA Analysis NGS:RNA-Seq SNP/WGA: Data Filters SNP/WGA: QC: LD: Plots SNP/WGA: Statistical Models	NGS TOOLBOX BETA NGS QC and manipulation NGS Mapping 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 atlas from BAM dataset Filter atlas on coverage and CIP Filter-to-interval: condense group format into ranges of bases Mapping provides simple stats on BAM files NGS: InDel Analysis NGS: Peak Calling NGS: RNA Analysis NGS:RNA-Seq 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 MGS Model-based Analysis of ChIP-Seq GeneTrack: Indexer on a BED file Peak predictor on GeneTrack index NGS: RNA Analysis NGS: RNA-Seq TopHat Find splice junctions using RNA-seq data Cuffdiff transcript assembly and FPKM/RPM estimates for RNA-Seq data Cuffdiff compare assembled transcripts to a reference annotation and track Cuffdiff transcripts across multiple experiments Cuffdiff find significance changes in transcript expression Filter Combined Transcripts using tracking file
---	---	--	--

UF Information Technology | www.it.ufl.edu

Galaxy Workflows

Extract Workflow

Unknown	25: hg19.chr9.bam	31: Dataset Security
BAM-to-SAM	26: BAM-to-SAM on data 25: converted SAM	32: Show Deleted Datasets
Convert Genomic Intervals To Strict BEDs	27: MACS peaks on hg19.chr9.bam	33: Show Hidden Datasets
MACS	27: MACS peaks on hg19.chr9.bam	34: Show Structure
Convert BED to GeneTrack Index	27: MACS peaks on hg19.chr9.bam	35: Export to File
	27: MACS peaks on hg19.chr9.bam	36: Delete
	27: MACS peaks on hg19.chr9.bam	37: Import from File
	27: MACS peaks on hg19.chr9.bam	38: Import from File
	27: MACS peaks on hg19.chr9.bam	39: Import from File
	27: MACS peaks on hg19.chr9.bam	40: Import from File
	27: MACS peaks on hg19.chr9.bam	41: Import from File
	27: MACS peaks on hg19.chr9.bam	42: Import from File
	27: MACS peaks on hg19.chr9.bam	43: Import from File
	27: MACS peaks on hg19.chr9.bam	44: Import from File
	27: MACS peaks on hg19.chr9.bam	45: Import from File
	27: MACS peaks on hg19.chr9.bam	46: Import from File
	27: MACS peaks on hg19.chr9.bam	47: Import from File
	27: MACS peaks on hg19.chr9.bam	48: Import from File
	27: MACS peaks on hg19.chr9.bam	49: Import from File
	27: MACS peaks on hg19.chr9.bam	50: Import from File
	27: MACS peaks on hg19.chr9.bam	51: Import from File
	27: MACS peaks on hg19.chr9.bam	52: Import from File
	27: MACS peaks on hg19.chr9.bam	53: Import from File
	27: MACS peaks on hg19.chr9.bam	54: Import from File
	27: MACS peaks on hg19.chr9.bam	55: Import from File
	27: MACS peaks on hg19.chr9.bam	56: Import from File
	27: MACS peaks on hg19.chr9.bam	57: Import from File
	27: MACS peaks on hg19.chr9.bam	58: Import from File
	27: MACS peaks on hg19.chr9.bam	59: Import from File
	27: MACS peaks on hg19.chr9.bam	60: Import from File
	27: MACS peaks on hg19.chr9.bam	61: Import from File
	27: MACS peaks on hg19.chr9.bam	62: Import from File
	27: MACS peaks on hg19.chr9.bam	63: Import from File
	27: MACS peaks on hg19.chr9.bam	64: Import from File
	27: MACS peaks on hg19.chr9.bam	65: Import from File
	27: MACS peaks on hg19.chr9.bam	66: Import from File
	27: MACS peaks on hg19.chr9.bam	67: Import from File
	27: MACS peaks on hg19.chr9.bam	68: Import from File
	27: MACS peaks on hg19.chr9.bam	69: Import from File
	27: MACS peaks on hg19.chr9.bam	70: Import from File
	27: MACS peaks on hg19.chr9.bam	71: Import from File
	27: MACS peaks on hg19.chr9.bam	72: Import from File
	27: MACS peaks on hg19.chr9.bam	73: Import from File
	27: MACS peaks on hg19.chr9.bam	74: Import from File
	27: MACS peaks on hg19.chr9.bam	75: Import from File
	27: MACS peaks on hg19.chr9.bam	76: Import from File
	27: MACS peaks on hg19.chr9.bam	77: Import from File
	27: MACS peaks on hg19.chr9.bam	78: Import from File
	27: MACS peaks on hg19.chr9.bam	79: Import from File
	27: MACS peaks on hg19.chr9.bam	80: Import from File
	27: MACS peaks on hg19.chr9.bam	81: Import from File
	27: MACS peaks on hg19.chr9.bam	82: Import from File
	27: MACS peaks on hg19.chr9.bam	83: Import from File
	27: MACS peaks on hg19.chr9.bam	84: Import from File
	27: MACS peaks on hg19.chr9.bam	85: Import from File
	27: MACS peaks on hg19.chr9.bam	86: Import from File
	27: MACS peaks on hg19.chr9.bam	87: Import from File
	27: MACS peaks on hg19.chr9.bam	88: Import from File
	27: MACS peaks on hg19.chr9.bam	89: Import from File
	27: MACS peaks on hg19.chr9.bam	90: Import from File
	27: MACS peaks on hg19.chr9.bam	91: Import from File
	27: MACS peaks on hg19.chr9.bam	92: Import from File
	27: MACS peaks on hg19.chr9.bam	93: Import from File
	27: MACS peaks on hg19.chr9.bam	94: Import from File
	27: MACS peaks on hg19.chr9.bam	95: Import from File
	27: MACS peaks on hg19.chr9.bam	96: Import from File
	27: MACS peaks on hg19.chr9.bam	97: Import from File
	27: MACS peaks on hg19.chr9.bam	98: Import from File
	27: MACS peaks on hg19.chr9.bam	99: Import from File
	27: MACS peaks on hg19.chr9.bam	100: Import from File

UF Information Technology | www.it.ufl.edu

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 fastq or readmulti or bed)

Input file: Data input 'tfile' (interval or sam or bam or fastq or readmulti or bed)

Format:

Effective Genome Size: (Human hg19)

Tag size (Optional):

Summary Statistics

Summary statistics on:

Details

Edit Workflow Attributes

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

Tags:

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

UF Information Technology | www.it.ufl.edu

Galaxy Workflows

UF Information Technology | www.it.ufl.edu

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](#)

UF Information Technology | www.it.ufl.edu

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/~moskalenko/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

Share with another user

UF Information Technology | www.it.ufl.edu

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!

UF Information Technology | www.it.ufl.edu

Demo

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

Tools

- Get Data
- Send Data
- ENCODE Tools
- Life-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
- Model Tools
- Multiple Alignments
- Metagenomic analyses
- FASTA manipulation
- NGS Bases
- NGS QC and manipulation

UF HPC Galaxy News:

- 2011-08-09: 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, NH&I, and the Huck Institutes of the Life Sciences.

History

- MACS hg19
- 0 bytes

Your history is empty. Click 'Get Data' on the left pane to start

UF Information Technology | www.it.ufl.edu

Galaxy demo

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

UF Research Computing

- ▶ Help and Support
 - Help Request Tickets
 - <https://support.hpc.ufl.edu>
 - For any kind of question or help requests
 - Searchable database of solutions
 - We are here to help!
 - support@hpc.ufl.edu



Training Schedule

- ✓ Aug 28: Intro to UFHPC, getting started
- ✓ Sept 10: Modules, RHEL6 Transition, User Q&A
- ✓ Sept 17: The Linux/Unix Shell - An Introduction
- ✓ Sept 24: Running Jobs, Submission Scripts, Modules
- ✓ Oct 1: Galaxy Overview, The Basics
- ▶ Oct 8: NGS Data Techniques: General Methods and Tools
- ▶ Oct 15: NGS Data Techniques: Reference Based Mapping and de Novo Assembly
- ▶ Oct 22: Phylogenetic Analyses
- ◆ Oct 29: Research Computing Day: Moving Big Data
- ▶ Nov 5: Multiprocessing at the HPC Center
- ▶ Nov 12: Using Git and CMake to Organize and Drive Data Analysis Pipelines
- ▶ Nov 19: Introduction to GPU Nodes
- ▶ Nov 29: NGS Data Techniques: RNA-Seq
- ▶ Dec 3: NGS Data Techniques: Alternative Splicing Analysis

UF Research Computing

- ▶ Help and Support (Continued)
 - <http://wiki.hpc.ufl.edu>
 - Documents on hardware and software resources
 - Various user guides
 - Many sample submission scripts
 - <http://hpc.ufl.edu/support>
 - Frequently Asked Questions
 - Account set up and maintenance

