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

BLAST ® Basic Local Alignment Search Tool

Home | Recent Results | Saved Strategies | Help

NCBI/ BLAST/ blastn suite

blastn | blastp | blastx | tblastn | tblastx

Enter Query Sequence

BLASTN programs search nucleotide databases using a nucleotide query. more...

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

Entrez Query Optional

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

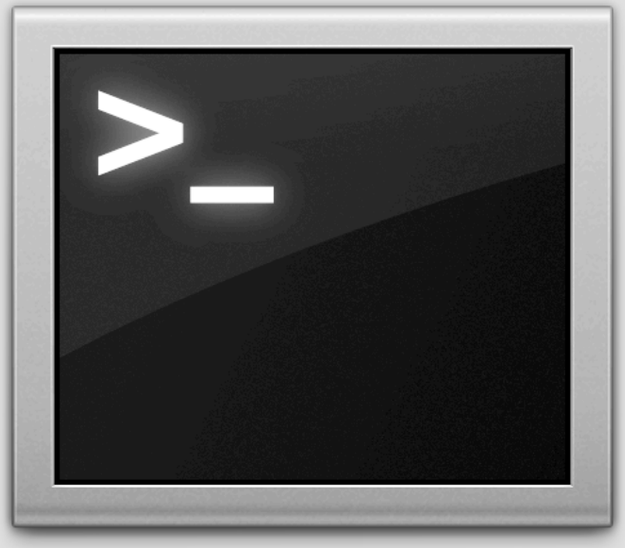
Search database Human G+T using Megablast (Optimize for highly similar sequences)

Show results in a new window

Algorithm parameters

Command Line Environment

Head node



Login to head node

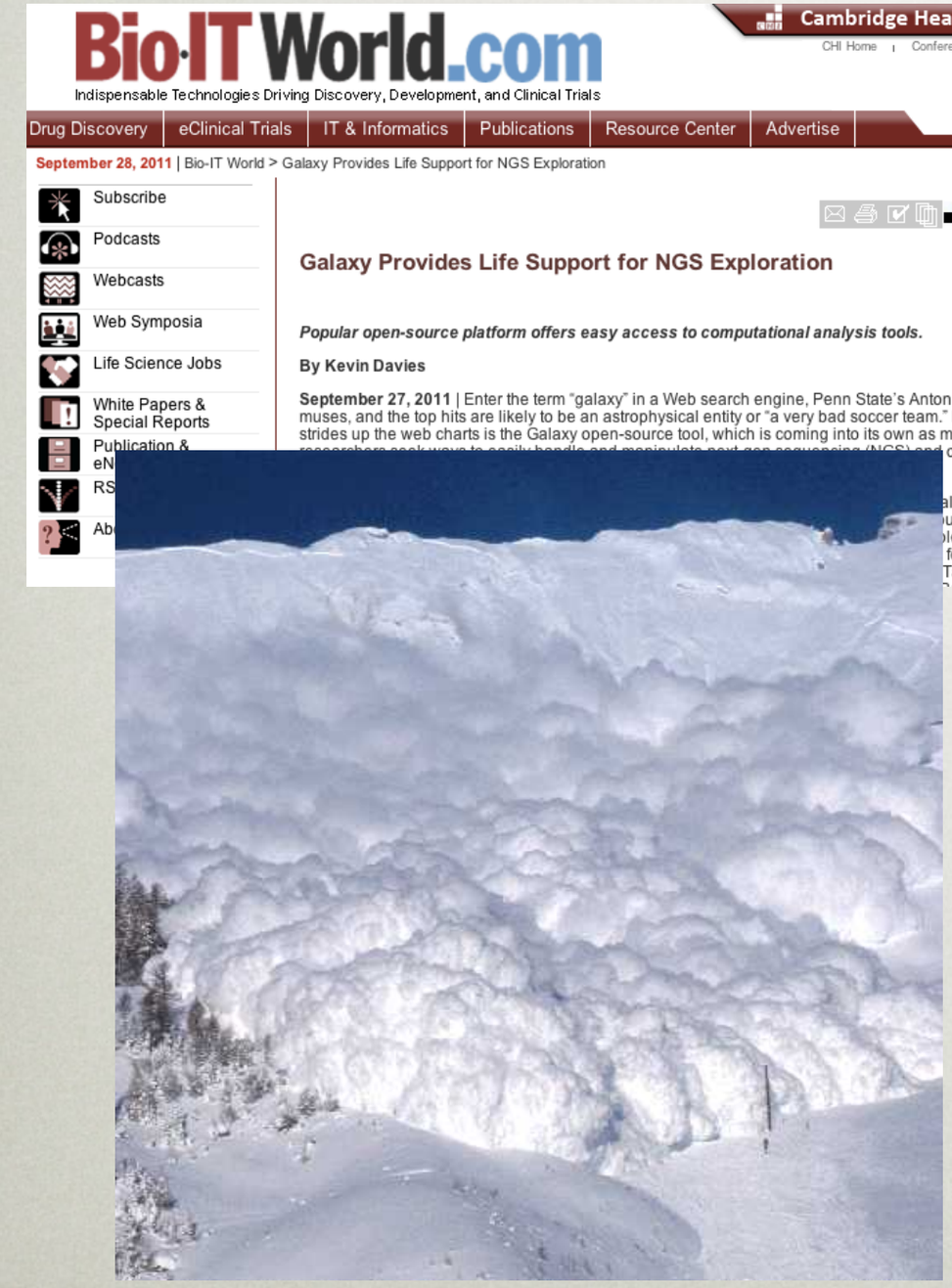
Scheduler

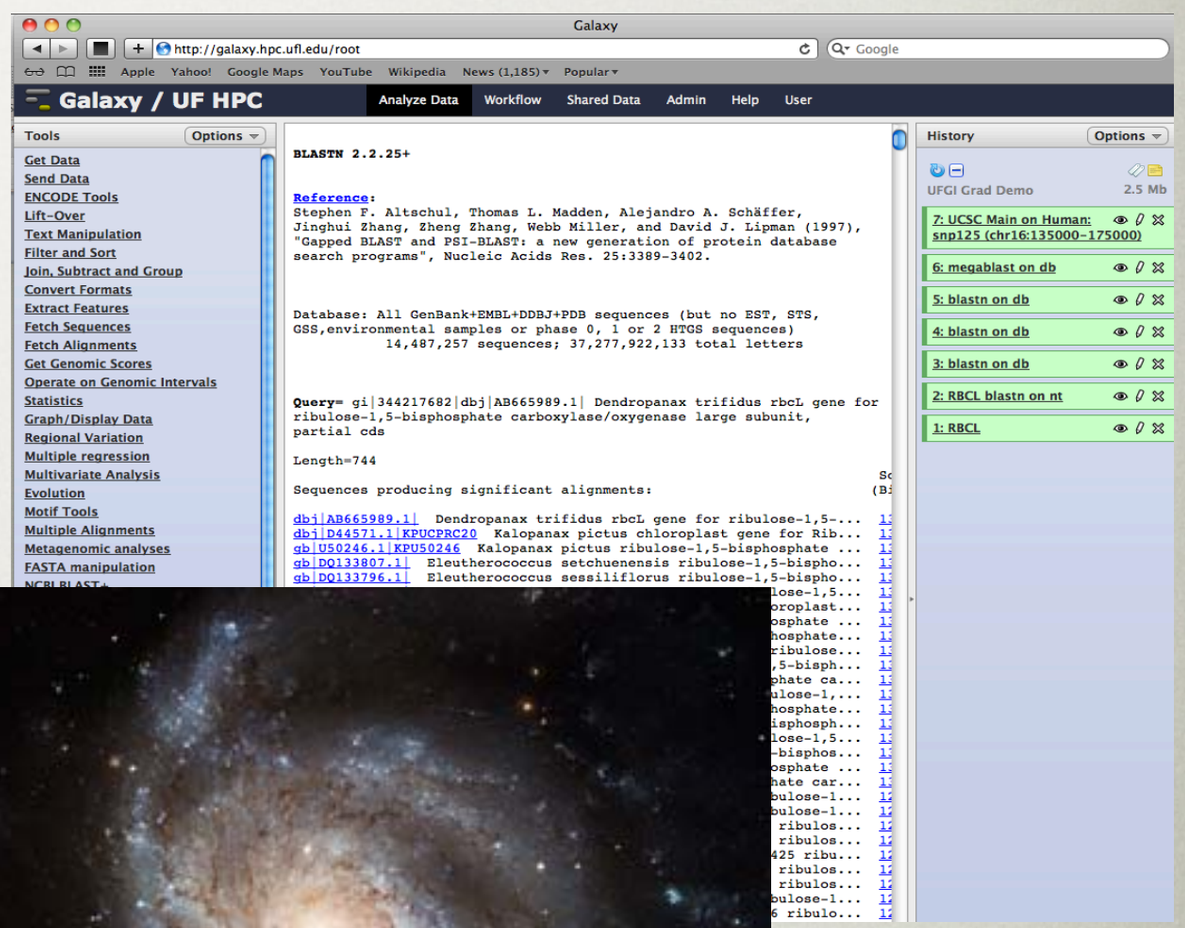
Interactive session or batch submission

Computing resources

Your job runs on the cluster

What is 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: 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

Gala orkspace

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

Gala

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

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 Genomi
- Statistics
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysi
- Evolution
- Motif Tools

NGS: Peak Calling

- MACS Model-based A ChIP-Seq
- CCAT Control-based Analysis Tool
- GeneTrack indexer or
- Peak predictor on Ge 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: hg19.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

Name	Message	Uploaded By	Date	File Size
2010-12-14_7_hg19_aln_sorted.bam		om@hpc.ufl.edu	2011-09-13	1.6 Gb
2010-12-14_7_hhv8_aln_sorted.bam		om@hpc.ufl.edu	2011-09-13	1.4 Gb
hg19.chr10.bam		om@hpc.ufl.edu	2011-09-14	80.8 Mb
hg19.chr11.bam		om@hpc.ufl.edu	2011-09-14	82.5 Mb
hg19.chr12.bam		om@hpc.ufl.edu	2011-09-14	74.9 Mb
hg19.chr13.bam		om@hpc.ufl.edu	2011-09-14	50.9 Mb
hg19.chr14.bam		om@hpc.ufl.edu	2011-09-14	36.1 Mb
hg19.chr15.bam		om@hpc.ufl.edu	2011-09-14	48.1 Mb
hg19.chr16.bam		om@hpc.ufl.edu	2011-09-14	55.9 Mb
hg19.chr17.bam		om@hpc.ufl.edu	2011-09-14	64.5 Mb
hg19.chr18.bam		om@hpc.ufl.edu	2011-09-14	33.5 Mb
hg19.chr19.bam		om@hpc.ufl.edu	2011-09-14	39.6 Mb
hg19.chr1.bam		om@hpc.ufl.edu	2011-09-14	148.5 Mb
hg19.chr20.bam		om@hpc.ufl.edu	2011-09-14	38.5 Mb
hg19.chr21.bam		om@hpc.ufl.edu	2011-09-14	17.5 Mb
hg19.chr22.bam		om@hpc.ufl.edu	2011-09-14	16.9 Mb
hg19.chr2.bam		om@hpc.ufl.edu	2011-09-14	126.3 Mb
hg19.chr2.sam		om@hpc.ufl.edu	2011-09-14	488.0 Mb
hg19.chr3.bam		om@hpc.ufl.edu	2011-09-14	118.0 Mb
hg19.chr4.bam		om@hpc.ufl.edu	2011-09-14	85.7 Mb
hg19.chr5.bam		om@hpc.ufl.edu	2011-09-14	102.7 Mb
hg19.chr6.bam		om@hpc.ufl.edu	2011-09-14	65.7 Mb
hg19.chr7.bam		om@hpc.ufl.edu	2011-09-14	89.9 Mb
hg19.chr8.bam		om@hpc.ufl.edu	2011-09-14	85.9 Mb
hg19.chr9.bam		om@hpc.ufl.edu	2011-09-14	64.8 Mb

For selected datasets: Import to current history Go

Data Access Control

Roles associated with new group

HPC test ChIP-seq analyses

>>

Users associated with new grou

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

Groups

search

Advanced Search

Name ↓	Users	Roles
HPC	0	2
Taylor HPC Lab	2	1

For 0 selected groups: Delete Undelete Purge

Roles

search

Advanced Search

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

search

Advanced Search

Email ↓	User Name	Groups	Roles	External	Last Login
aedison@ufl.edu	aedison	0	1	yes	Sep 15, 2011
bostwick@ufl.edu	bostwick	0	1	yes	Sep 15, 2011
cgraves3@ufl.edu	cgraves3	0	1	yes	Sep 15, 2011
cjeffrey@ufl.edu	cjeffrey	0	1	yes	Sep 15, 2011
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/Phylogeny
- RNA-Seq
- ChIP-Seq
- GATK

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

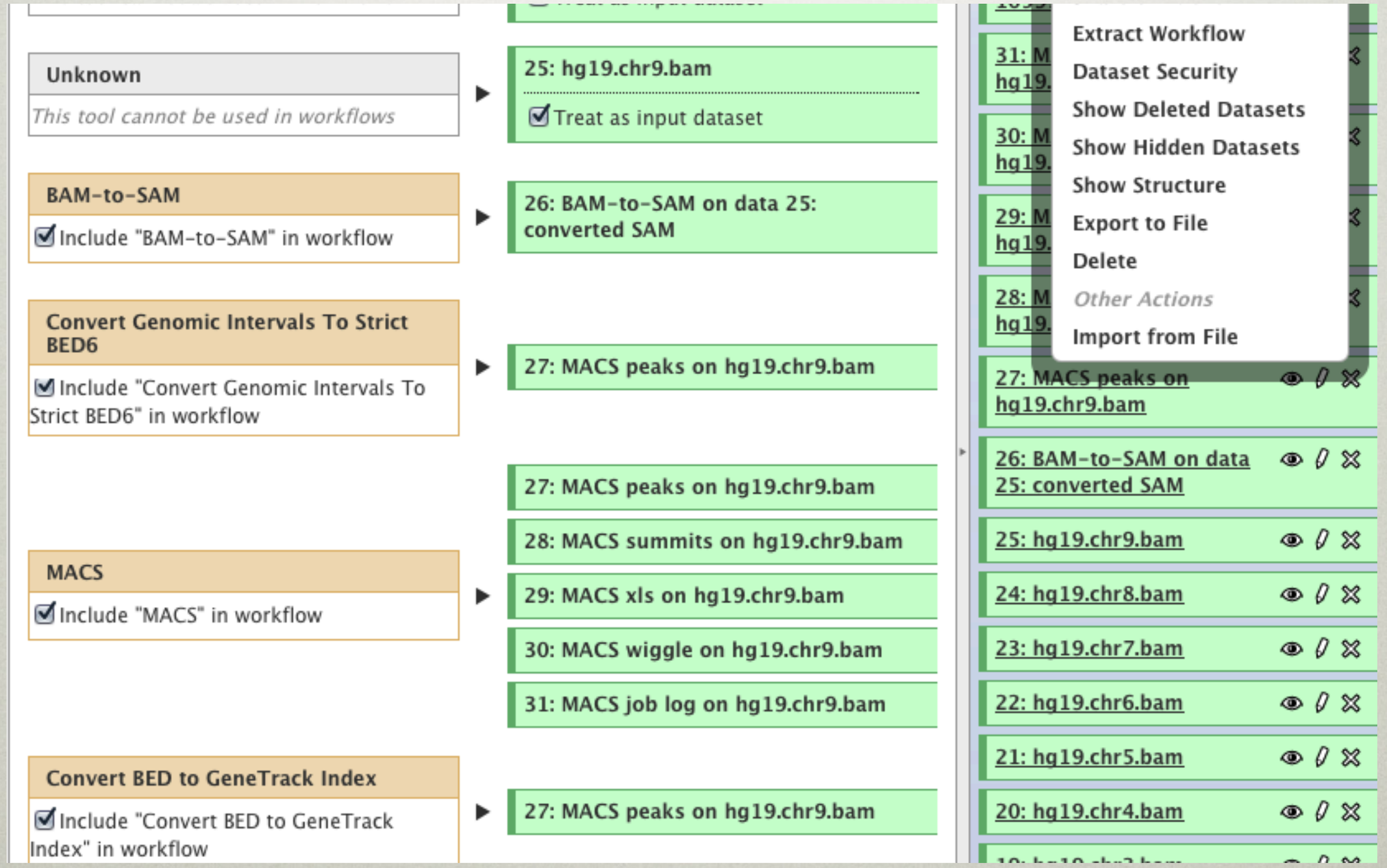
RNA-SEQ

- 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



Galaxy Workflows

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

Input dataset
output

Input dataset
output

Input dataset
output

Input dataset
output

BAM-to-SAM
BAM File to Convert
output1 (sam)

MACS
Treatment file
Input file
bedoutput (bed)
summitsoutput (bed)
xlsoutput (tabular)
wigoutput (wig)
log (txt)
xlsdiagreport (tabular)

Convert BED to GeneTrack Index
Choose BED file
output1 (genetrack)

Summary Statistics
Summary statistics on
out_file1 (tabular)

Convert Genomic Intervals To Strict BED6
Choose intervals
output1 (bed6)

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 × ChIP-Seq × hg19 × chr9 ×

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

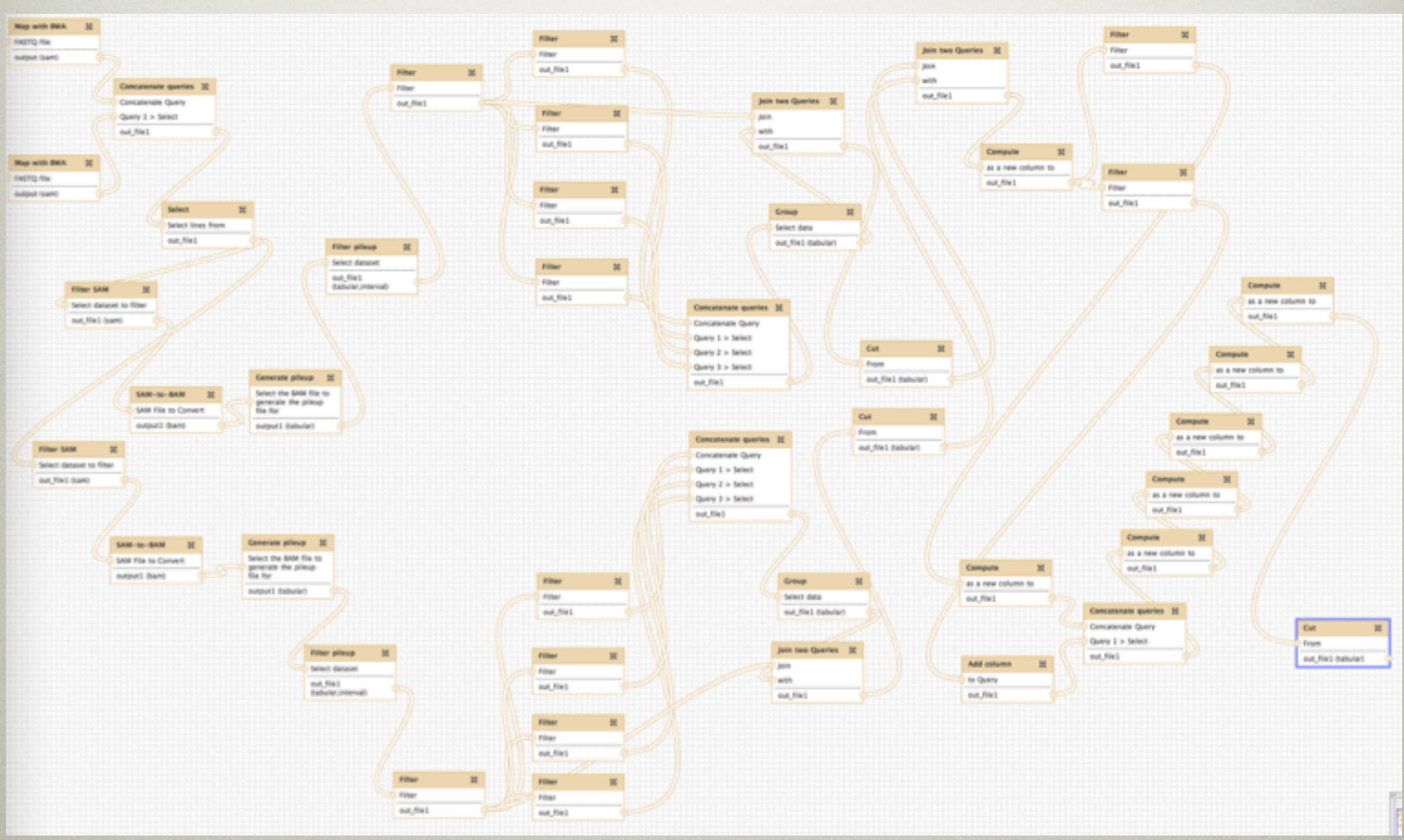
Annotation / Notes:
This is a partial peak calling with MACS using hg19 and chr9 data

gs at the exact same

al

nent for model building:

Galaxy Workflows



Visualization

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

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 KOSAKOVSKY POND[1,2,*], SAMIR WADHAWAN[3,6*], FRANCESCA CHIAROMONTE[4], GURUPRASAD ANANDA[1,3], WEN-YU CHUNG[1,3,7], JAMES TAYLOR[1,5], ANTON NEKRUTENKO[1,3] and THE GALAXY TEAM[1*]

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

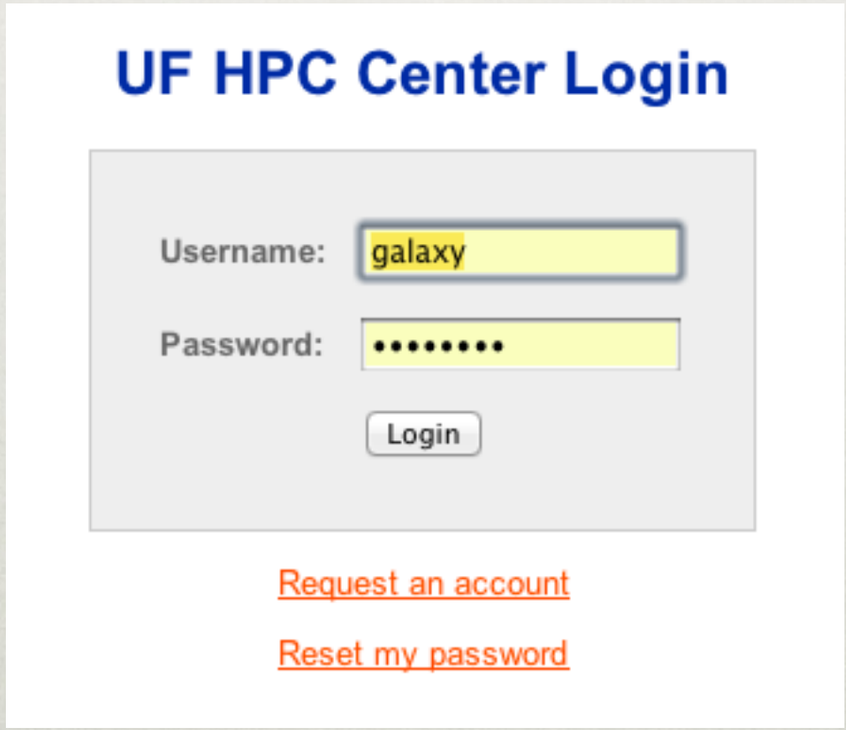

Galaxy / UF HPC /
Analyze Data
Workflow
Shared Data
Help
User
Tools
Options
Get Data
Send Data
ENCODE Tools
Lift-Over
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Fetch Alignments
Get Genomic Scores
Operate on Genomic Intervals
Statistics
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)
UF HPC center
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

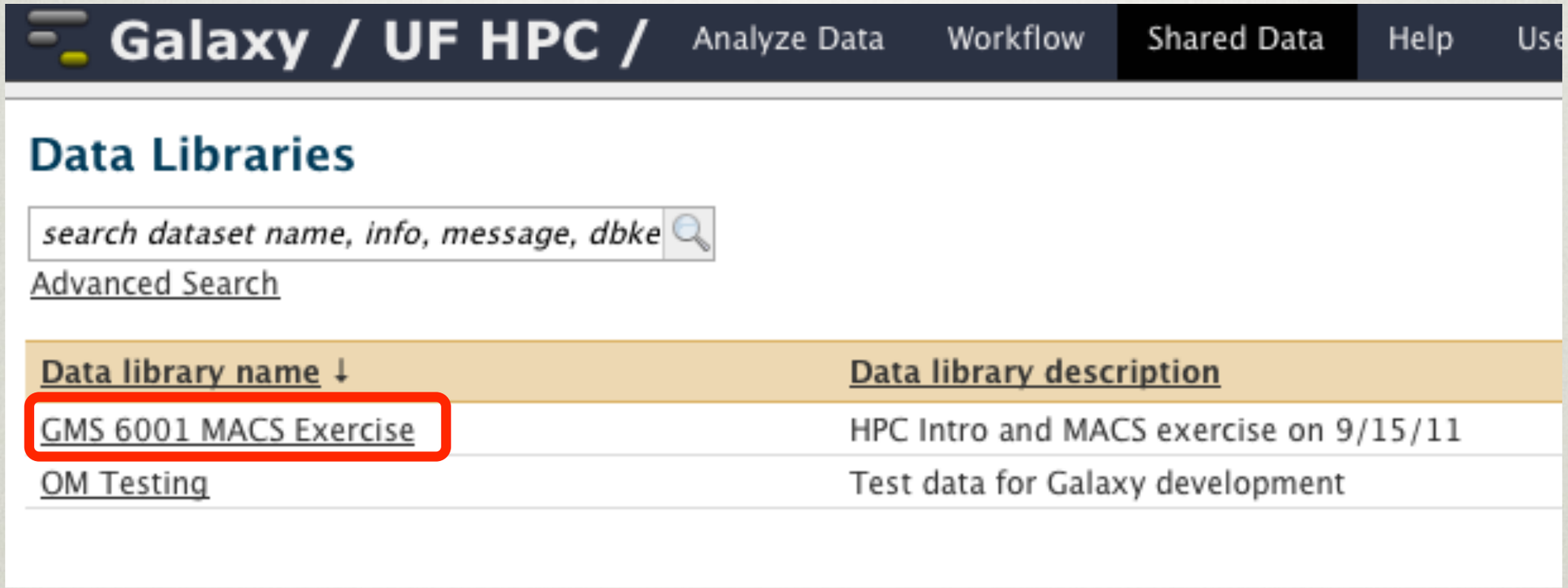
http://galaxy.hpc.ufl.edu



History/Shared Data

http://galaxy.hpc.ufl.edu

Shared Data


Galaxy / UF HPC /
Analyze Data
Workflow
Shared Data
Help
Use
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
OM Testing
Test data for Galaxy development

MACS – Load data

Data Library "GMS 6001 MACS Exercise"

MACS test data

Name	Message	Uploaded By	Date	File Size
2010-12-14_7_hg19_aln_sorted.bam		om@hpc.ufl.edu	2011-09-13	1.6 Gb
2010-12-14_7_hhv8_aln_sorted.bam		om@hpc.ufl.edu	2011-09-13	1.4 Gb
hg19.chr10.bam		om@hpc.ufl.edu	2011-09-14	80.8 Mb
hg19.chr11.bam		om@hpc.ufl.edu	2011-09-14	82.5 Mb
hg19.chr12.bam		om@hpc.ufl.edu	2011-09-14	74.9 Mb
hg19.chr13.bam		om@hpc.ufl.edu	2011-09-14	50.9 Mb
hg19.chr14.bam		om@hpc.ufl.edu	2011-09-14	36.1 Mb
hg19.chr15.bam		om@hpc.ufl.edu	2011-09-14	48.1 Mb
hg19.chr16.bam		om@hpc.ufl.edu	2011-09-14	55.9 Mb
hg19.chr17.bam		om@hpc.ufl.edu	2011-09-14	64.5 Mb
hg19.chr18.bam		om@hpc.ufl.edu	2011-09-14	33.5 Mb
hg19.chr19.bam		om@hpc.ufl.edu	2011-09-14	39.6 Mb
hg19.chr1.bam		om@hpc.ufl.edu	2011-09-14	148.5 Mb
hg19.chr20.bam		om@hpc.ufl.edu	2011-09-14	38.5 Mb
hg19.chr21.bam		om@hpc.ufl.edu	2011-09-14	17.5 Mb
hg19.chr22.bam		om@hpc.ufl.edu	2011-09-14	16.9 Mb
hg19.chr2.bam		om@hpc.ufl.edu	2011-09-14	126.3 Mb
hg19.chr2.sam		om@hpc.ufl.edu	2011-09-14	488.0 Mb
hg19.chr3.bam		om@hpc.ufl.edu	2011-09-14	118.0 Mb
hg19.chr4.bam		om@hpc.ufl.edu	2011-09-14	85.7 Mb
hg19.chr5.bam		om@hpc.ufl.edu	2011-09-14	102.7 Mb
hg19.chr6.bam		om@hpc.ufl.edu	2011-09-14	65.7 Mb
hg19.chr7.bam		om@hpc.ufl.edu	2011-09-14	89.9 Mb
hg19.chr8.bam		om@hpc.ufl.edu	2011-09-14	85.9 Mb
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

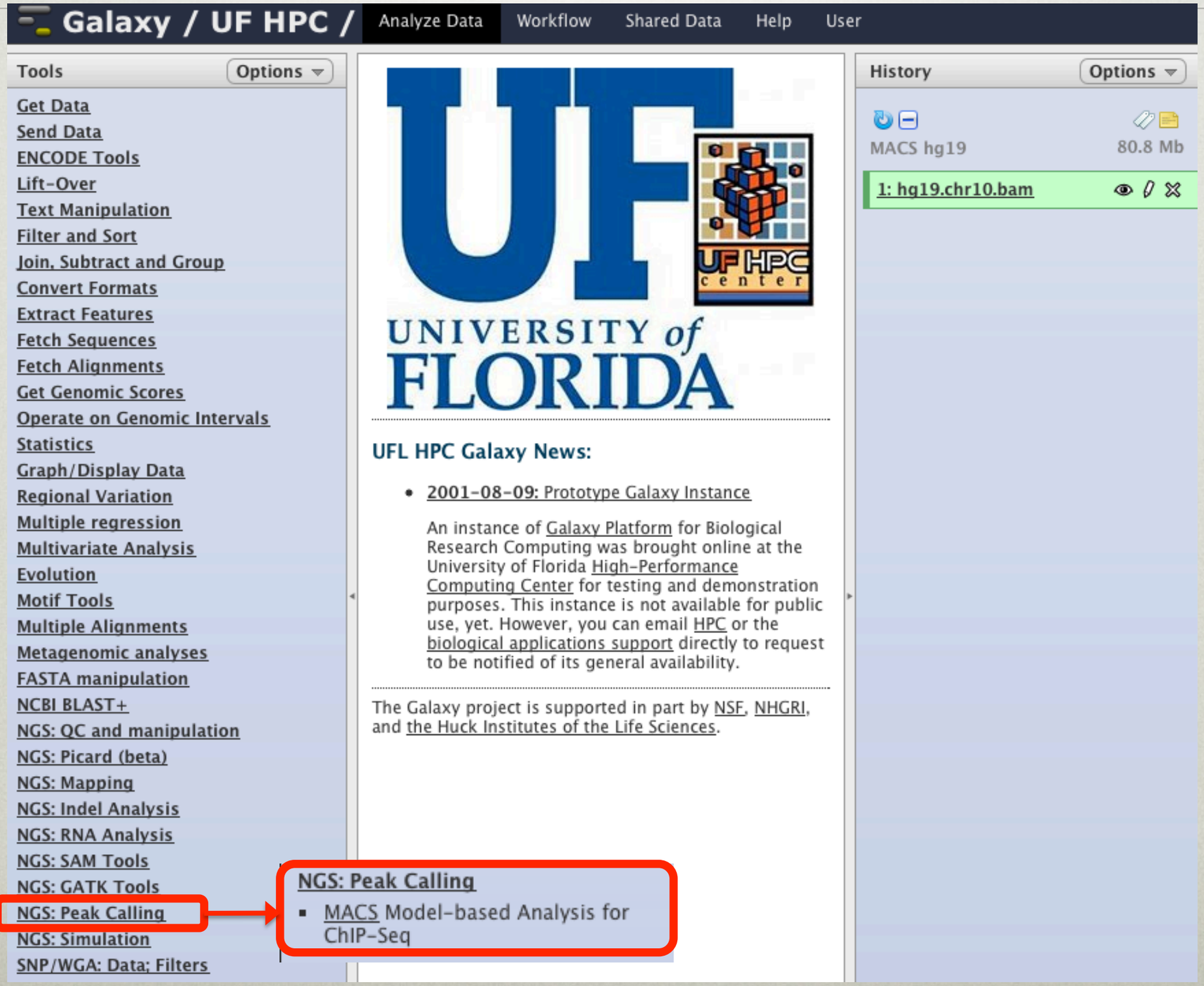
MACS hg19 80.8 Mb

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

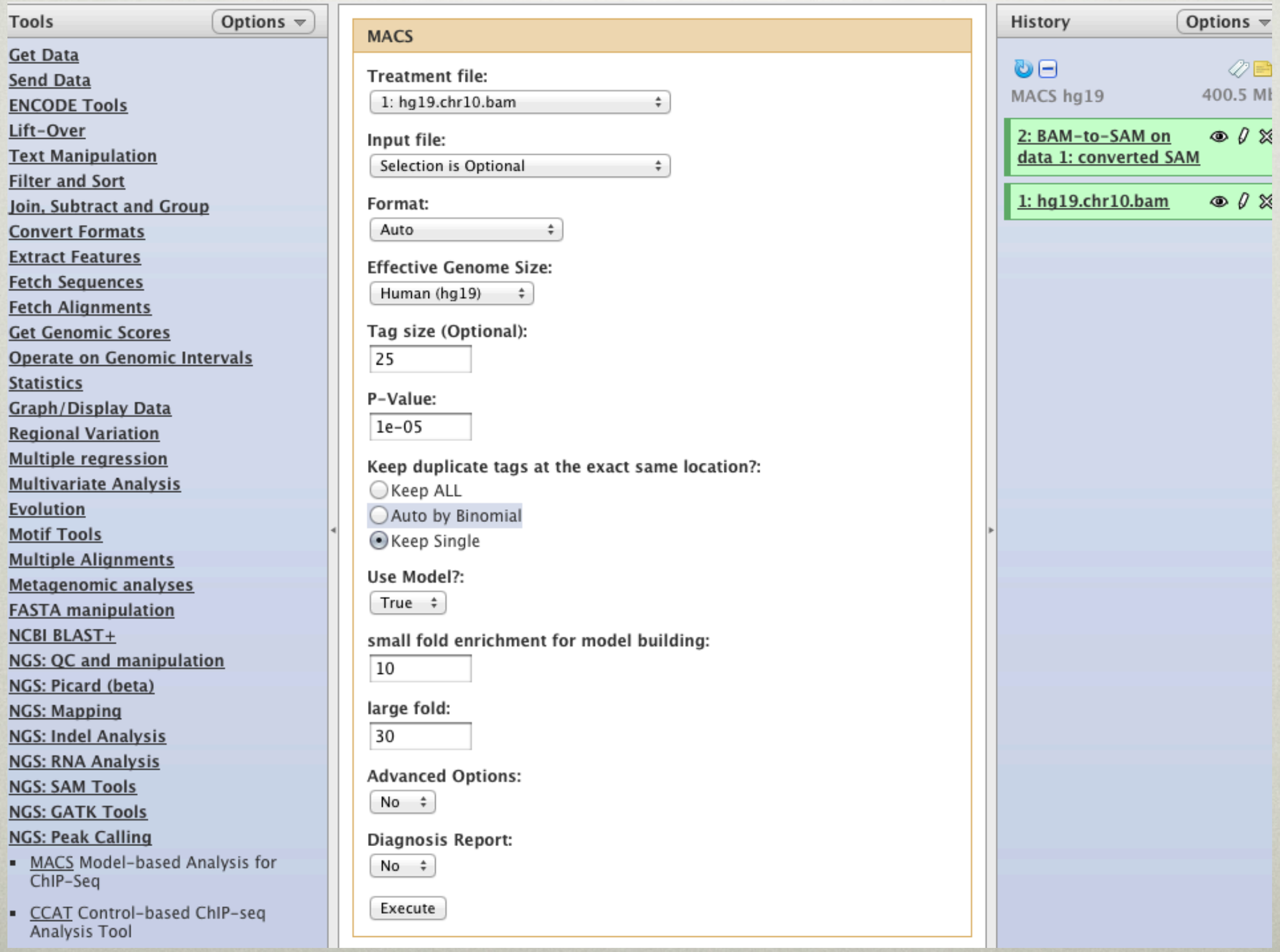
1: hg19.chr10.bam

```
@SQ     SN:chrUn_gl000248       LN:39786
@SQ     SN:chrUn_gl000249       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 --tryharc
HWUSI-EAS1654_0011:7:101:15569:18265#0/1        16      chr10   60132   0       40M     *       0       0       TTAATTGACG
HWUSI-EAS1654_0011:7:70:15927:14469#0/1 16      chr10   60156   0       40M     *       0       0       TTGAGTTCGGTTGAGTTT
HWUSI-EAS1654_0011:7:77:2963:15680#0/1  16      chr10   60189   0       40M     *       0       0       TCTTCCACAAGGGATTGT
HWUSI-EAS1654_0011:7:56:13734:7224#0/1  0       chr10   60478   0       40M     *       0       0       CGCCTTTGGAAGGAGCAT
HWUSI-EAS1654_0011:7:27:17072:12387#0/1 0       chr10   60491   255     40M     *       0       0       AGCATTTATCCCCAGCAA
HWUSI-EAS1654_0011:7:41:17138:6404#0/1  0       chr10   60491   255     40M     *       0       0       AGCATATATCCCCAGCAA
HWUSI-EAS1654_0011:7:82:1846:20862#0/1  0       chr10   60491   255     40M     *       0       0       AGCATATATCCCCAGCAA
HWUSI-EAS1654_0011:7:96:15342:7286#0/1  16      chr10   60513   0       40M     *       0       0       TCCGGTTTTTTGAAGTCT
HWUSI-EAS1654_0011:7:91:12232:19734#0/1 0       chr10   60618   0       40M     *       0       0       AATCTTTGTGTTATACAT
HWUSI-EAS1654_0011:7:4:3181:2553#0/1    0       chr10   60918   0       40M     *       0       0       GAAGTATCAATATGCCTT
HWUSI-EAS1654_0011:7:43:15903:1917#0/1  16      chr10   61129   0       40M     *       0       0       TTTGTATTGGTAGGATAA
HWUSI-EAS1654_0011:7:11:8621:5776#0/1   16      chr10   61203   0       40M     *       0       0       ATGAGGCCTCACTCTGTC
HWUSI-EAS1654_0011:7:44:10220:2559#0/1  0       chr10   62846   0       40M     *       0       0       ATACTGGGGAGGAGCTGT
HWUSI-EAS1654_0011:7:95:15568:1582#0/1  16      chr10   63074   0       40M     *       0       0       TGAAAAGCCAATGGCTGC
HWUSI-EAS1654_0011:7:32:3459:2766#0/1   0       chr10   63307   0       40M     *       0       0       AAAGGACATATAATCTTG
HWUSI-EAS1654_0011:7:89:18639:10678#0/1 0       chr10   63307   0       40M     *       0       0       AAAGGACATATAATCTTG
HWUSI-EAS1654_0011:7:109:19350:5114#0/1 16      chr10   63367   0       40M     *       0       0       TACAGCTGATGCTTTCTG
HWUSI-EAS1654_0011:7:68:3139:9708#0/1   16      chr10   63523   0       40M     *       0       0       CCCAAAGATGGTTCACAT
HWUSI-EAS1654_0011:7:27:15884:16777#0/1 0       chr10   63723   0       40M     *       0       0       CAGTCTTCAGCCCTAGAC
HWUSI-EAS1654_0011:7:64:1626:5553#0/1   0       chr10   63915   0       40M     *       0       0       AGCTAATCAGGGAGGGGC
HWUSI-EAS1654_0011:7:21:7450:15409#0/1  0       chr10   64133   0       40M     *       0       0       GACAAGGTCTTTGATTTA
HWUSI-EAS1654_0011:7:67:16528:10957#0/1 0       chr10   64143   0       40M     *       0       0       TTGATTTAACCCAATCCA
HWUSI-EAS1654_0011:7:61:13906:4548#0/1  0       chr10   64190   0       40M     *       0       0       TATGAGCAAAGTCTCCAA
HWUSI-EAS1654_0011:7:42:8178:16716#0/1  16      chr10   64272   0       40M     *       0       0       ATTCTAAAAGCCAGGAAA
HWUSI-EAS1654_0011:7:64:10597:5725#0/1  16      chr10   64305   0       40M     *       0       0       TAATCTAGGAAAACTTCC
HWUSI-EAS1654_0011:7:50:6555:17127#0/1  16      chr10   64375   0       40M     *       0       0       TGGGAAATTCATCACAAA
HWUSI-EAS1654_0011:7:119:10260:12806#0/1        0       chr10   64633   0       40M     *       0       0       TGAAGGAAAG
HWUSI-EAS1654_0011:7:117:5789:14945#0/1 0       chr10   64648   0       40M     *       0       0       GTCATTTTCAGACAAACA
HWUSI-EAS1654_0011:7:49:10018:10089#0/1 0       chr10   64844   0       40M     *       0       0       AAAGAGCATGATGAAAGC
HWUSI-EAS1654_0011:7:81:5640:15183#0/1  0       chr10   65212   0       40M     *       0       0       ACTGGAGCTCCCAAATTT
HWUSI-EAS1654_0011:7:67:19633:13182#0/1 0       chr10   65483   0       40M     *       0       0       GCCATAAAATGAGTCTCA
HWUSI-EAS1654_0011:7:77:17790:3617#0/1  0       chr10   65626   0       40M     *       0       0       CTGAATGAGCATTGGGCC
HWUSI-EAS1654_0011:7:60:8645:3117#0/1   16      chr10   65700   0       40M     *       0       0       TATCGAAACCTCTGGGAT
HWUSI-EAS1654_0011:7:60:6931:13228#0/1  0       chr10   66024   0       40M     *       0       0       TCACTAAGAAATGAAACA
HWUSI-EAS1654_0011:7:76:1871:15186#0/1  0       chr10   66045   0       40M     *       0       0       GATATTACAACTGACACC
HWUSI-EAS1654_0011:7:114:9163:3182#0/1  0       chr10   66273   0       40M     *       0       0       CGGATTCACAGCAGAATT
```

MACS (NGS: Peak Calling)



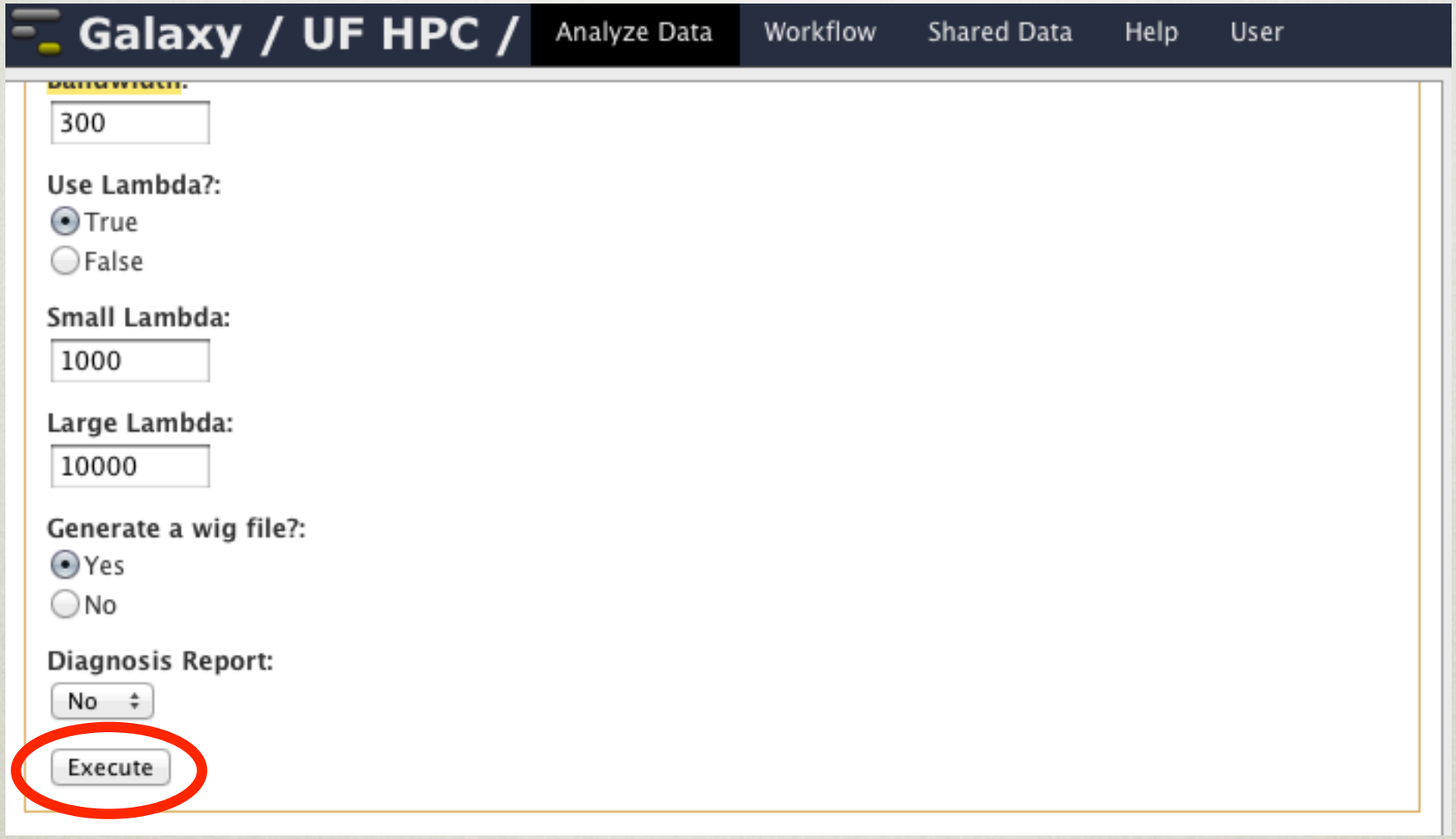
Submission form



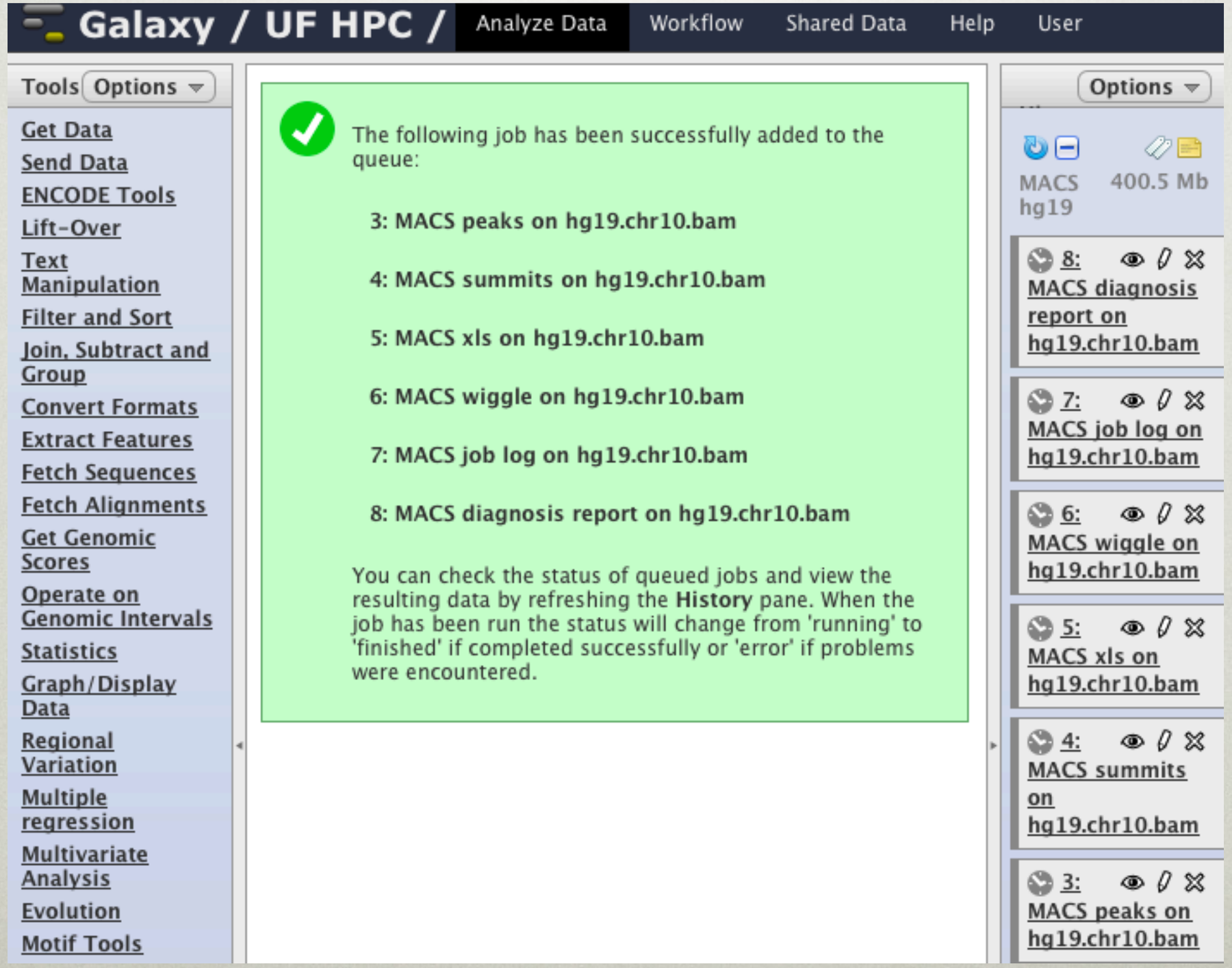
MACS options

- 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 ½ window size without the model: default is 300

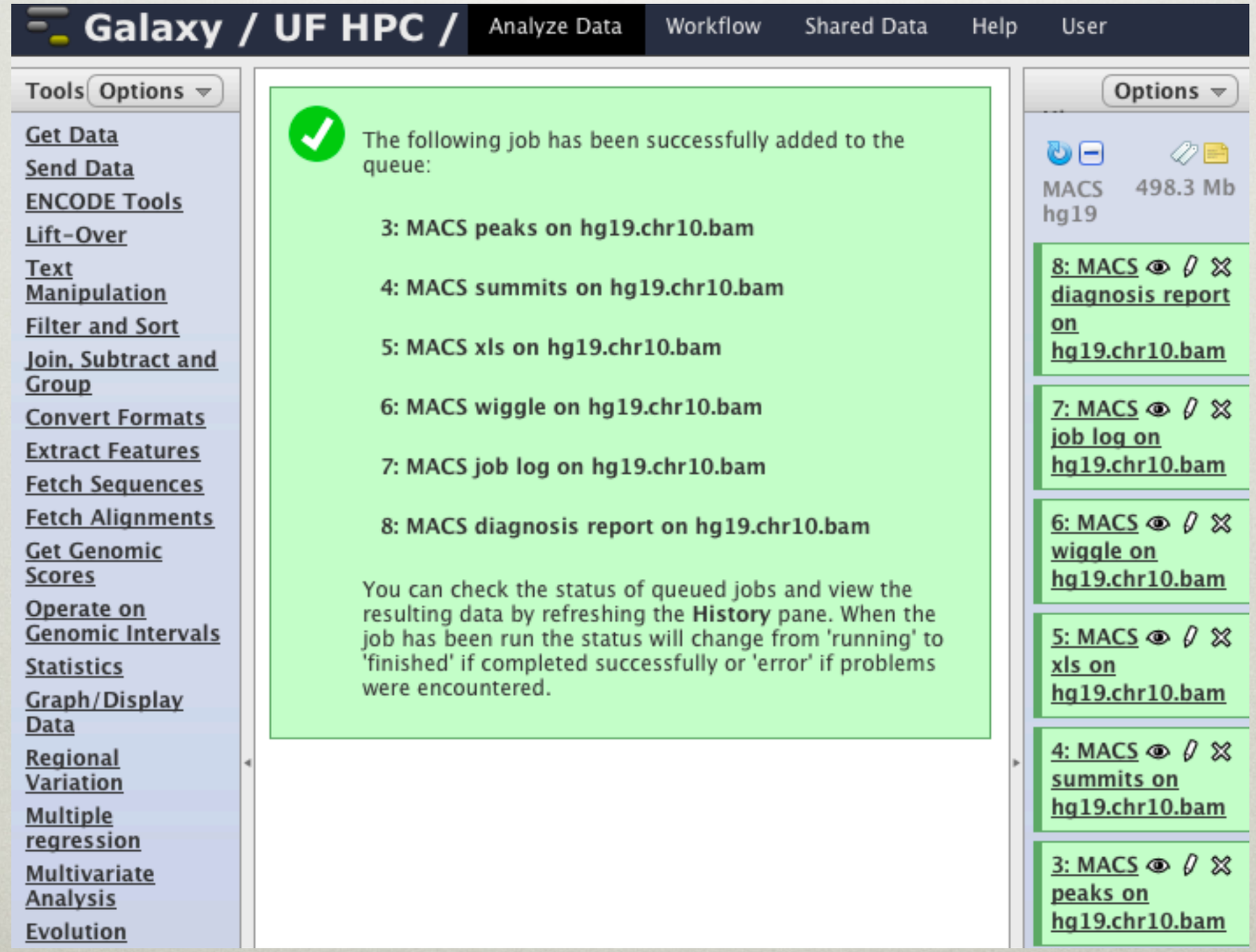
Submit the job to cluster



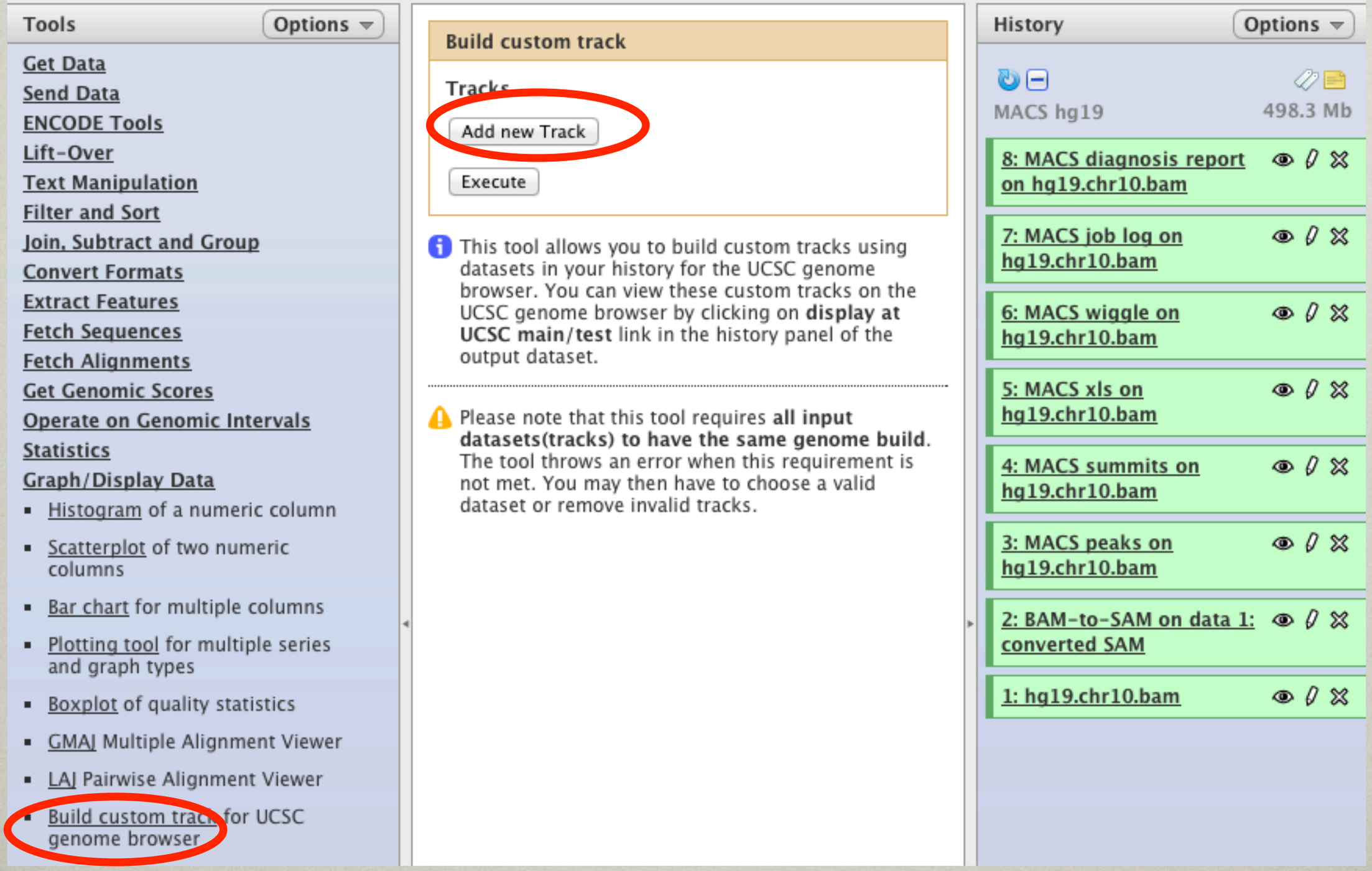
Cluster job run



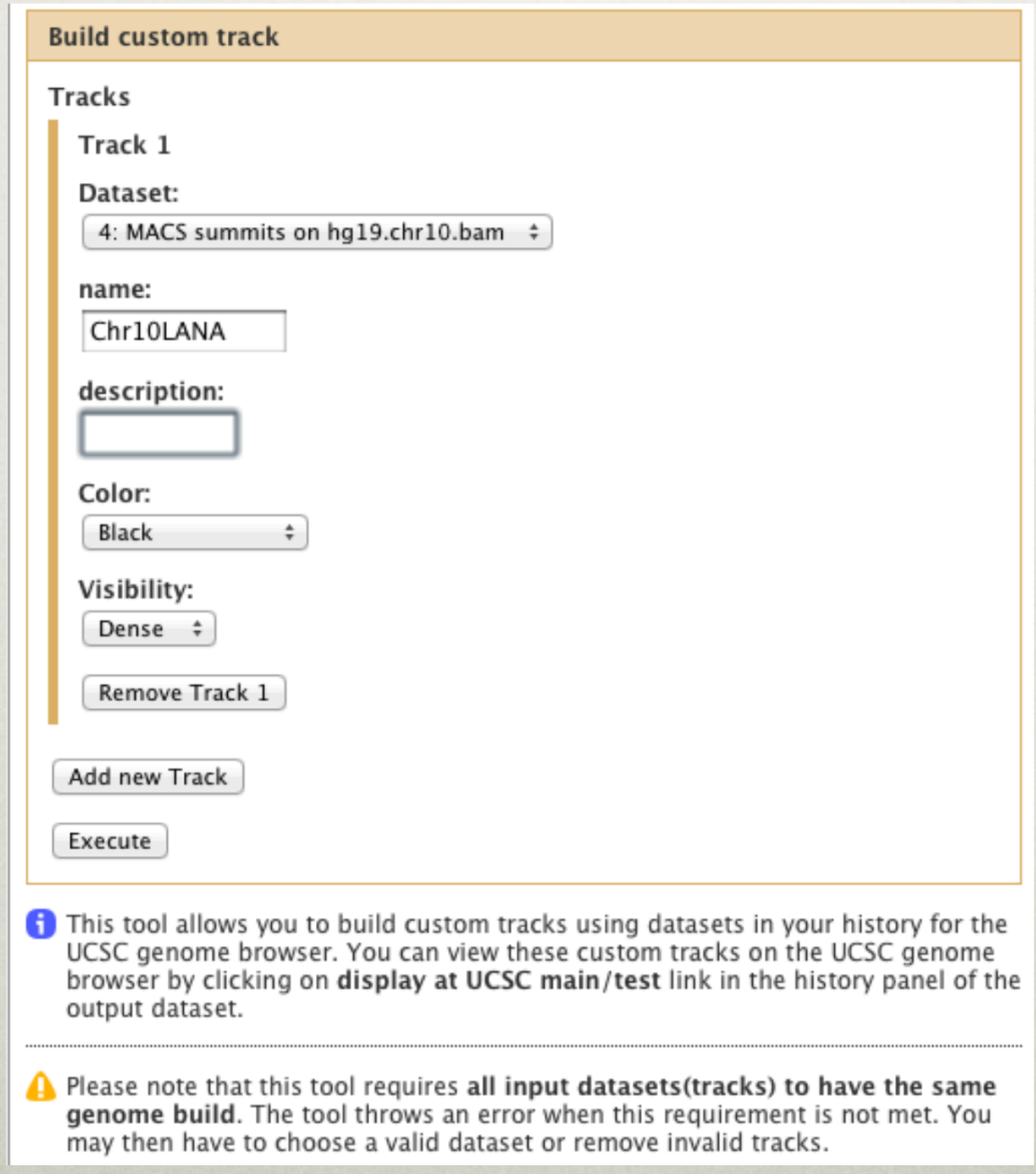
Job completion



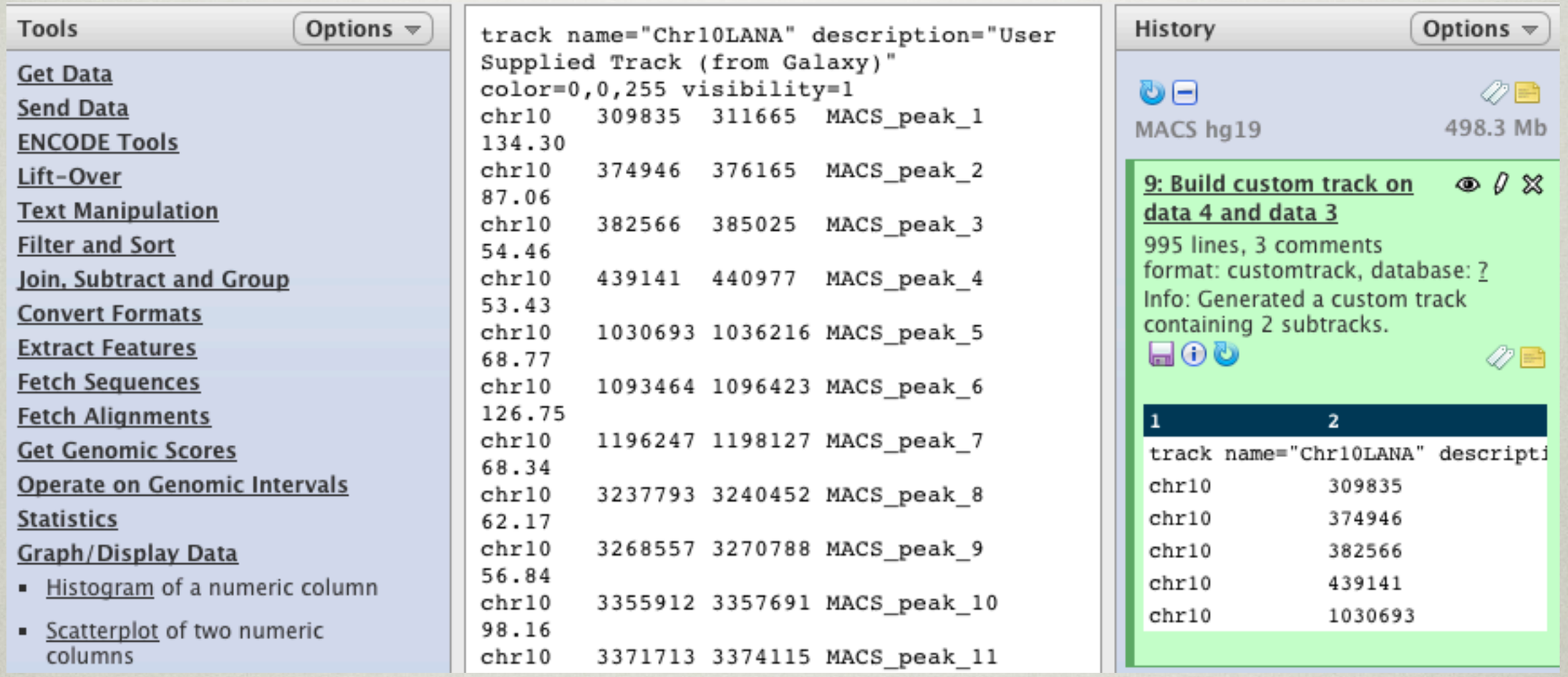
Build a genome browser track



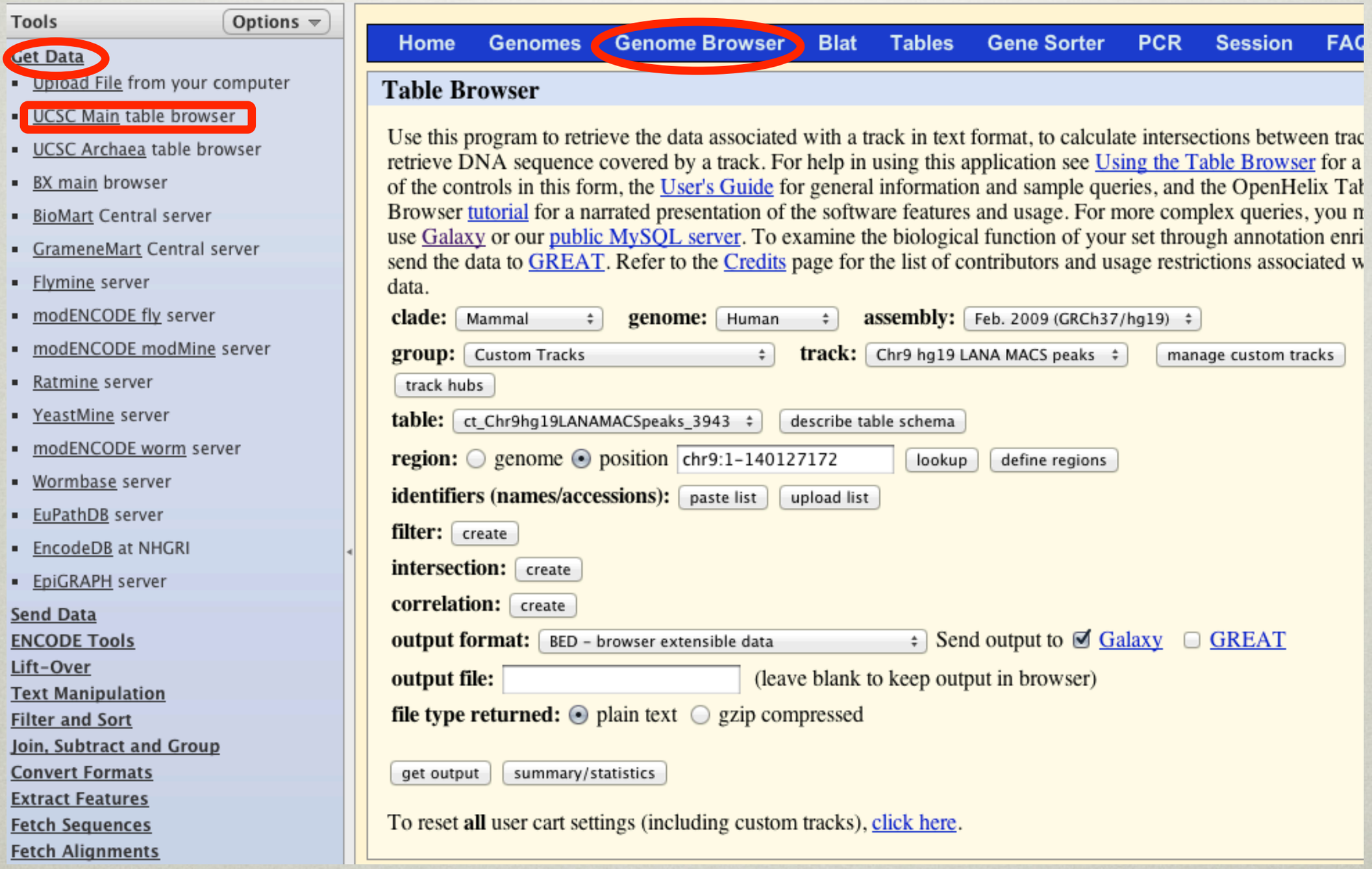
Submit a track build job



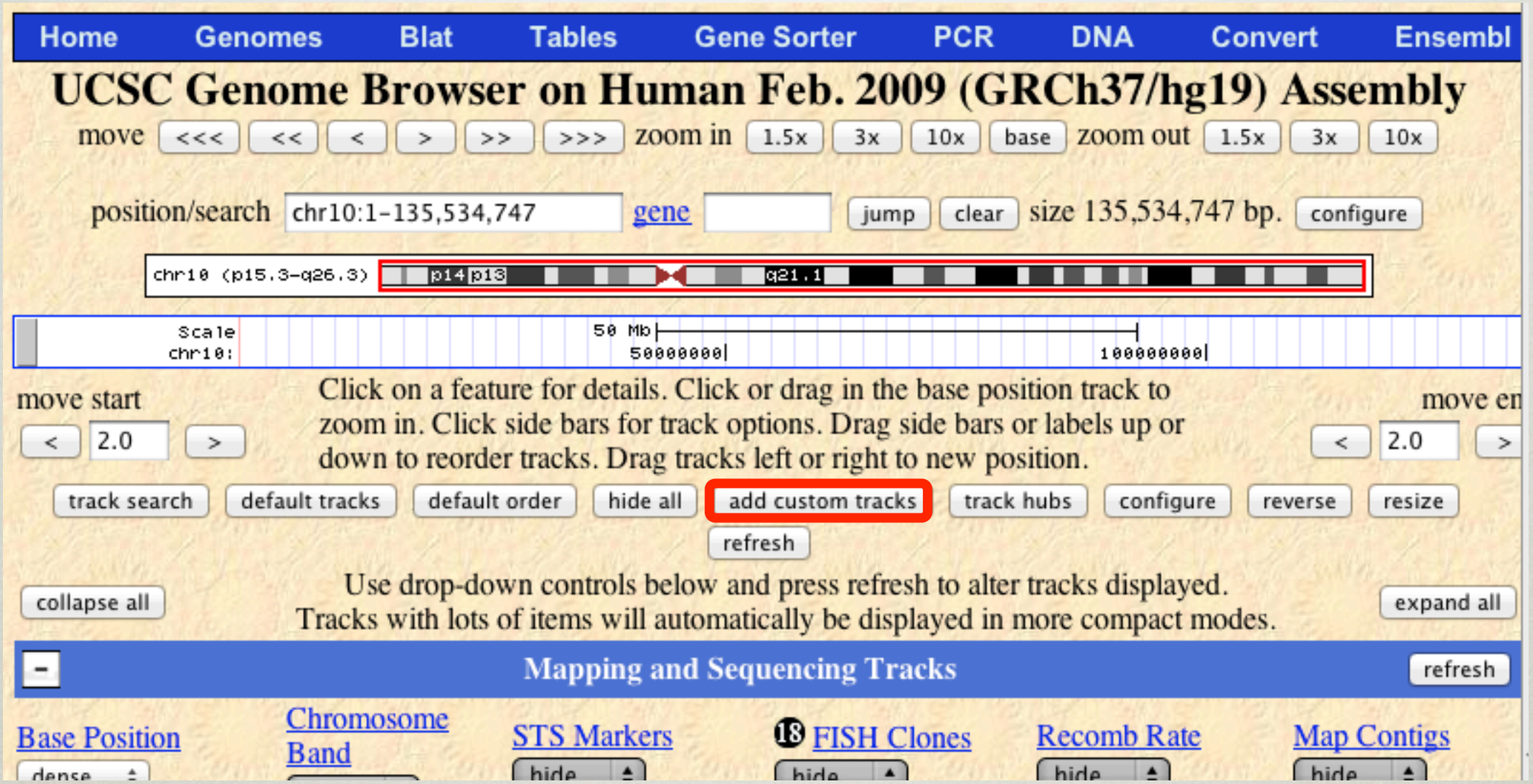
Open the track



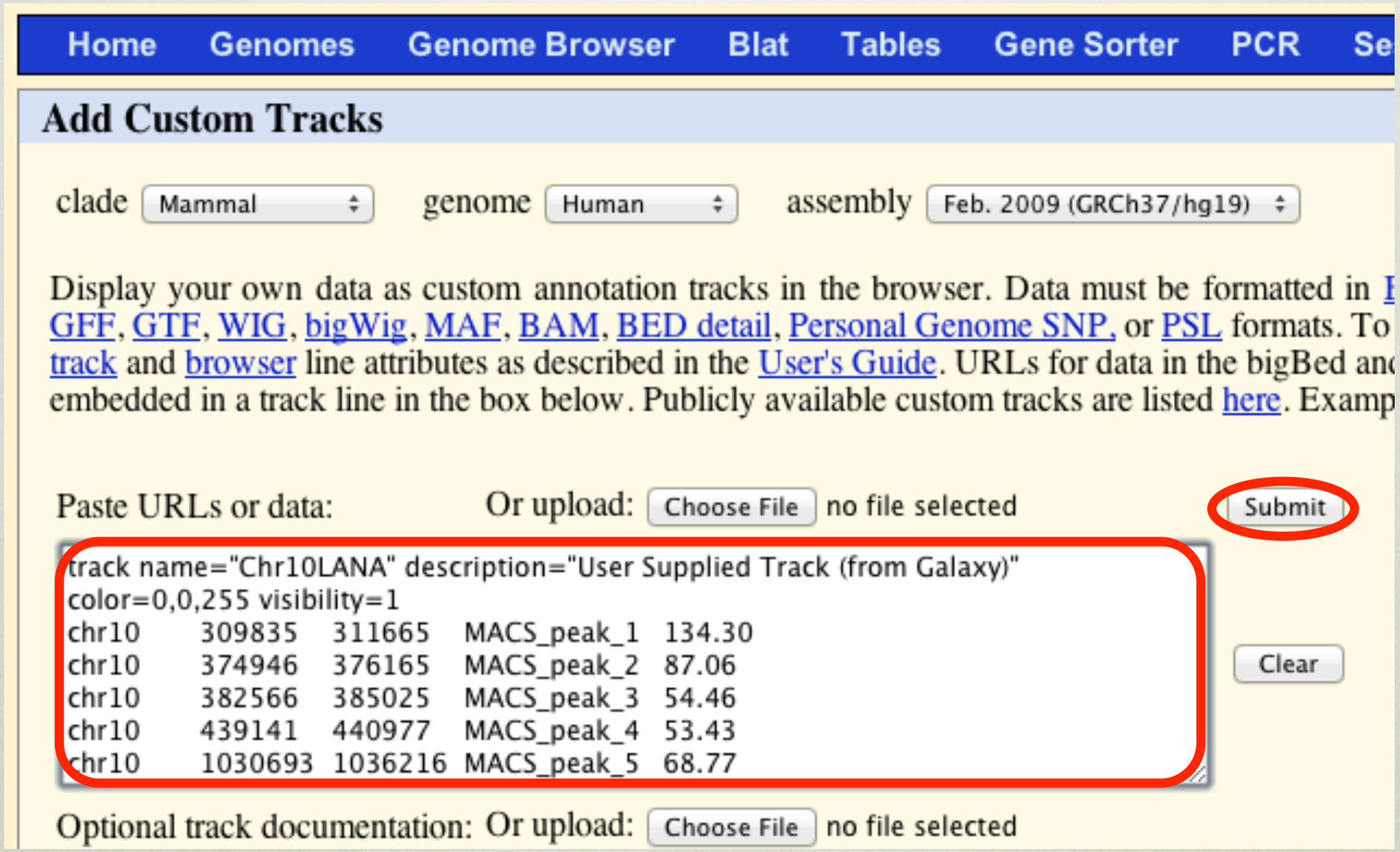
Genome Browser



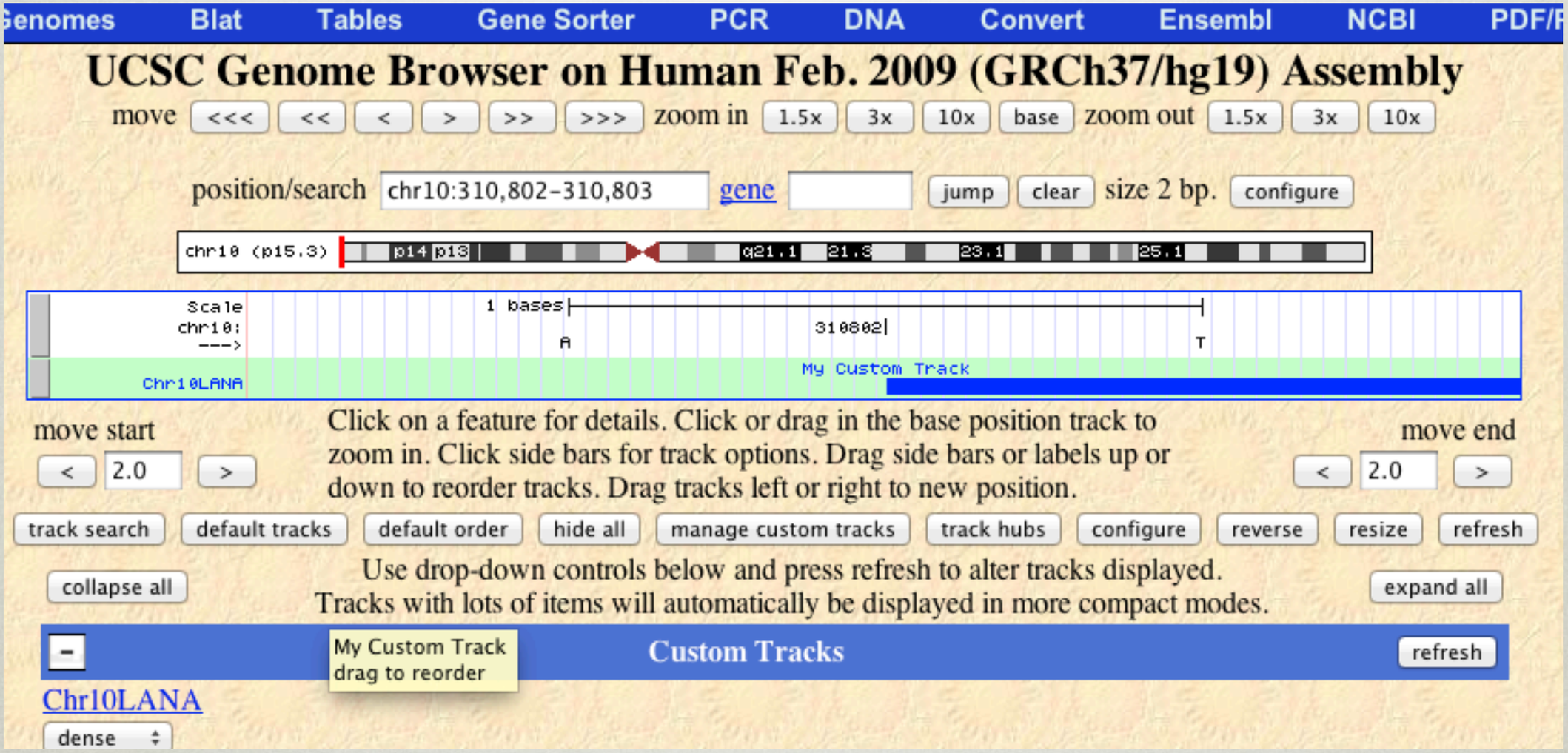
Add a custom track



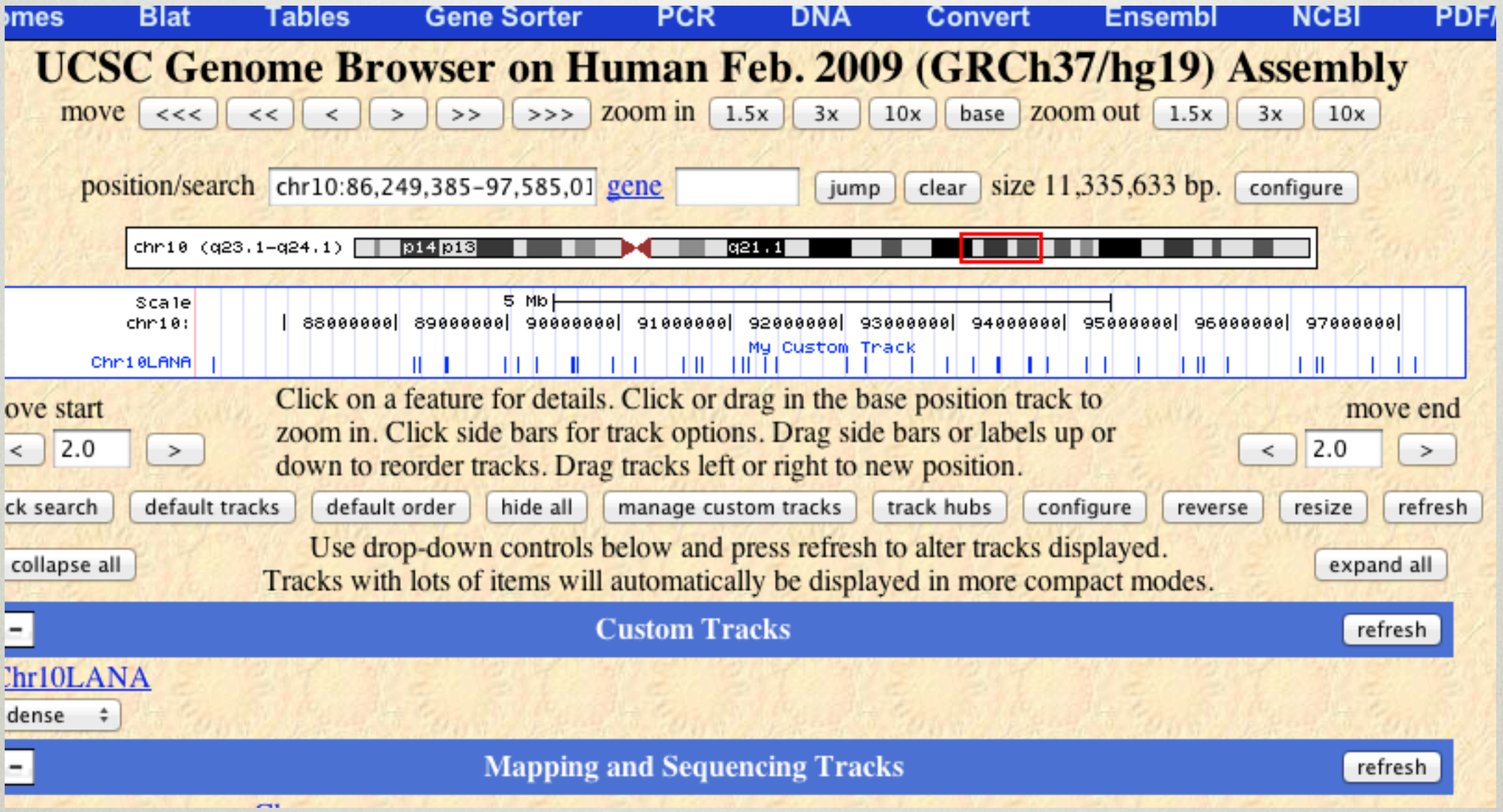
Paste track data



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
A
310802
T
My Custom Track
Chr10LANA
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.
Tracks with lots of items will automatically be displayed in more compact modes.
expand all
-
My Custom Track
drag to reorder
Custom Tracks
refresh
Chr10LANA
dense

Zoom in, pan, etc.



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 - Biological applications support
  - magitz@ufl.edu - Training