



Introduction to the Galaxy framework

Sarah Diehl

Bioinformatics Facility, MPI-IE, Freiburg

The start site

deeptools.ie-freiburg.mpg.de

The screenshot displays the Galaxy / deeptools website interface. A blue rounded rectangle highlights the top navigation bar, which includes links for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualize', and 'History'. A label 'Top menu' points to this bar. Another blue rounded rectangle highlights the left sidebar, which contains a search bar and a list of tool categories: 'Get Data', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Operate on Genomic Intervals', 'BEDtools', 'UCSC tools', 'Peak Calling', 'deeptools', 'Conversion formats', and 'Workflows'. A label 'Tools' points to this sidebar. A third blue rounded rectangle highlights the main content area, which features the 'deeptools' logo, a welcome message, a list of tools, and a section titled 'QUALITY CHECKS - FORMAT CONVERSION & NORMALIZATION'. A label 'Main frame' points to this central area. On the right, a 'History' panel shows a list of recent jobs, including 'Testing deeptools', '61: bamFingerprint on data 1, data 2, and data 16', '60: MACS2: callpeak on data 1 (peaks: encodePeak)', '59: MACS2: callpeak on data 1 (peaks: encodePeak)', '58: MACS2: callpeak on data 1 (html report)', '57: MACS2: callpeak on data 1 (peaks: bed)', '55: bamCorrelate on data 5, data 4, and others', '54: Extract reads', '53: Extract reads', '50: Extract reads', '49: Extract reads', and '48: heatmapper image'. A label 'History' points to this panel. A large, stylized '4/5/2017' watermark is overlaid on the right side of the image.

Galaxy / deeptools

Analyze Data Workflow Shared Data Visualize History

Top menu

Tools

search tools

Get Data

Text Manipulation

Filter and Sort

Join, Subtract and Group

Operate on Genomic Intervals

BEDtools

UCSC tools

Peak Calling

deeptools

Conversion formats

Workflows

All workflows

deeptools

User-friendly tools for the normalization and visualization of deep-sequencing data

Welcome to the MPI-IE's Galaxy instance dedicated to the analysis of high-throughput sequencing data!

Did you just receive a large file from your high-throughput sequencing center and now you're wondering how to turn these millions of lines into a heatmap of read coverages? Or how to generate a file you can upload to the UCSC or the IGV browser that you can look at? Thinking more about it, you would actually like to compare your sample that should be the profile of a knock-down cell with the sample from the wild type? And how to make sure that everything went fine with the preparation and sequencing of the samples?

The good news: here, you'll most likely find all the tools and what you need as we have developed several tools to help you make sense out of the data generated by high-throughput sequencing. You will find them under the header **deeptools** in the menu on the left hand side. For a gallery of images that can be used to read more see <http://f1000.com/protocols/articles/suppl/article/100405>

For support, questions, or feature requests contact: deeptools@googlegroups.com

If you are not familiar with the Galaxy framework, please have a look at this manual here.

QUALITY CHECKS - FORMAT CONVERSION & NORMALIZATION

bam bigwig

History

Testing deeptools

13.1 GB

61: bamFingerprint on data 1, data 2, and data 16

60: MACS2: callpeak on data 1 (peaks: encodePeak)

59: MACS2: callpeak on data 1 (peaks: encodePeak)

58: MACS2: callpeak on data 1 (html report)

57: MACS2: callpeak on data 1 (peaks: bed)

55: bamCorrelate on data 5, data 4, and others

54: Extract reads

53: Extract reads

50: Extract reads

49: Extract reads

48: heatmapper image

only function that requires login

if you get lost, this is the button that will always lead you back to this start page

free text tool search

way to go for sample data, workflows and informative pages

give meaningful history name (just click on it)

to create new history or to access a saved one

TOOL PANEL
 → contains all tools installed in this Galaxy instance

HISTORY PANEL
 → contains all files that one produces or uploads
 does not only contain the files, but also most relevant information about how they were generated (think of it like a log)

deepTools
 user-friendly tools for the normalization and visualization of deep-sequencing data
 Welcome to the MPI-IE's Galaxy instance dedicated to the analysis of high-throughput sequencing data!

Did you just receive a large file from your high-throughput sequencing center and now you're wondering how to turn these millions of lines into a heatmap of read coverages? Or how to generate a file you can upload to the UCSC or the IGV browser that you can look at? Thinking more about it, you would actually like to compare your sample that should be the profile of a knock-down cell with the sample from the wild type? And how to make sure that everything went fine with the preparation and sequencing of the samples?

The good news: here, you'll most likely find all the tools and help you need as we have developed several tools to help you make sense out of the data generated by high-throughput sequencing. You will find them under the header **deepTools** in the menu on the left-hand side. For a gallery of images that can be produced, see <http://f1000.com/posters/browse/summary/1094053>.

For support, questions, or feature requests contact: deeptools@googlegroups.com

Data Libraries

- Published Histories
- Published Workflows
- Published Visualizations
- Published Pages

History

Testing deepTools	13.1 GB
55: bamCorrelate on data 5, data 4, and others	0 0
54: Extract reads	0 0
53: Extract reads	0 0
50: Extract reads	0 0
49: Extract reads	0 0
48: heatmapper image	0 0
47: heatmapper	0 0

History List

- Saved Histories
- Histories Shared with Me
- Current History
- Create New
- Copy History
- Copy Datasets
- Share or Publish
- Extract Workflow
- Dataset Security
- Resume Paused Jobs
- Collapse Expanded Datasets
- ✓ Include Deleted Datasets
- Include Hidden Datasets
- Unhide Hidden Datasets

Show/hide
Tool search

Click
category
name to
expand

Click tool
name to
use

Tools

search tools

Get Data
Send Data
EMBOSS
ENCODE Tools
Lift-Over
Text Manipulation
Filter and Sort
Join, Subtract and Group
Unix Tools
MPI-IE Epicenter Tools
Convert Formats
UCSC tools
Extract Features
Fetch Sequences
Get Genomic Scores
Operate on Genomic Intervals

- **Intersect** the intervals of two datasets
- **Subtract** the intervals of two datasets
- **Merge** the overlapping intervals of a dataset
- **Concatenate** two datasets into one dataset
- **Base Coverage** of all intervals
- **Coverage** of a set of intervals on second set of intervals
- **Complement** intervals of a dataset
- **Cluster** the intervals of a dataset
- **Join** the intervals of two datasets side-by-side
- **Get flanks** returns flanking region/s for every gene
- **Fetch closest non-overlapping feature** for every interval
- **Profile Annotations** for a set of genomic intervals

BEDtools
Statistics

Intersect (version 1.0.0)

Return:
Overlapping Intervals
(see figure below)

of:
7: Intersect on data 5 and data 6

First dataset
that intersects:
7: Intersect on data 5 and data 6

Second dataset

for at least:
1
(bp)

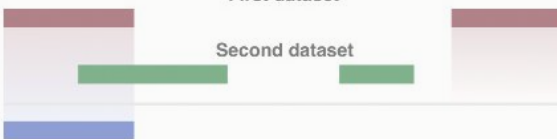
Execute

ⓘ TIP: If your dataset does not appear in the pulldown menu, it means that it is not in interval format. Use "edit attributes" to set chromosome, start, end, and strand columns.


Screencasts!
See Galaxy Interval Operation [Screencasts](#) (right click to open this link in another window).

Syntax
Where overlap is at least sets the minimum length (in base pairs) of overlap between elements of the two datasets
Overlapping Intervals returns entire intervals from the first dataset that overlap the second dataset. The returned intervals are completely unchanged, and this option only filters out intervals that do not overlap with the second dataset.
Overlapping pieces of returns intervals that overlap the second dataset, but only the part that overlaps between the first dataset and the second dataset. The intervals returned are the same as the first dataset, and all fields besides start and end are guaranteed to remain unchanged.

Examples
Overlapping Intervals:



Overlapping Pieces of Intervals:



History options

History

Galaxy Introduction 1 22.6 Mb

9: Bar chart on data 8

8: Count on data 7

7: Intersect on data 5 and data 6

6: Get flanks on data 2

5: Find and Replace on data 4

4: Find and Replace on data 3

3: awk on data 1

2: UCSC Main on Mouse: refGene (genome)

1: GSE37268_mof3.out.hpeak.txt

Information stored for each data set

Tool: bamCorrelate	
Name:	bamCorrelate on data 5, data 4, and others
Created:	Dec 11, 2013
Filesize:	64.9 KB
Dbley:	hg19
Format:	png
Galaxy Tool Version:	1.0.1
Tool Version:	
Tool Standard Output:	stdout
Tool Standard Error:	stderr
Tool Exit Code:	0
API ID:	a50c14e4ca28bfa0
Input Parameter	
Value	
Bam file	6: IMR90_H3K36me3.bam
Label	
Bam file	5: IMR90_H3K27me3_2.bam
Label	
Bam file	4: IMR90_H3K27me3_1.bam
Label	
Bam file	3: IMR90_H3K27ac_3.bam
Label	
Bam file	2: IMR90_H3K27ac_2.bam
Label	
Bam file	1: IMR90_H3K27ac_1.bam
Label	
Length of the average fragment size	200
Correlation method	Pearson
Choose computation mode	bins
Bin size in bp	10000
Number of samples	100000
Show advanced options	no
Show additional output options	no

details about how
this file was
generated

8: bamCorrelate on data 5, data 4, and others

64.9 KB

format: png, database: hg19



Image in png format

download
the file

re-run an analysis with the
exact same parameters
!!extremely useful!!

have a look here as well

Attributes Convert Format Datatype Permissions

Edit Attributes

Name:
 change the file name

Info:

Annotation / Notes:
 put some info you would like to keep, e.g. what experiment this file is related to, why you generated it etc.

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build:

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

view the file

edit attributes

delete the file
(can be recovered)

History

- 48: heatmapper image
- 47: heatmapper sorted/filtered regions
- 46: heatmapper matrix of heatmap values
- 45: heatmapper image
- 44: computeMatrix on data 32 and data 34 regions
- 43: computeMatrix on data 32 and data 34
- 42: computeMatrix on data 32 and data 34 column
- 41: computeMatrix on data 32 and data 34
- 40: profiler image
- 39: heatmapper image

HISTORY LISTS

- Saved Histories
- Histories Shared with Me
- CURRENT HISTORY
- Create New
- Copy History
- Copy Datasets
- Share or Publish
- Extract Workflow
- Dataset Security
- Resume Paused Jobs
- Collapse Expanded Datasets
- ✓ Include Deleted Datasets
- Include Hidden Datasets
- Unhide Hidden Datasets

States of data sets

Waiting to be run

🕒 5: Find and Replace on data 4 🔍 ✎ ✕

Running

⚙️ 3: Compute on data 2 🔍 ✎ ✕

Finished successfully

7: Intersect on data 5 and data 6 🔍 ✎ ✕

Failed

❌ 121: 28S rRNA.fa 🔍 ✎ ✕

0 bytes

An error occurred running this job: *Traceback (most recent call last):*

File "/galaxy/galaxy_server/tools/data_source/upload.py", line 403, in <module>

__main__()

File "/galaxy/galaxy_server/tools/data_source/upload.py", line 392, in __main__

add_file(dataset, registry, json_fil



check whether all the parameters you chose were correct

send a bug report to us

Tools

- Useful categories:
 - Text manipulation
 - Join, Subtract and Group
 - Operate on Genomic Intervals
- Only on our Galaxy:
 - **deepTools → NGS data processing & visualization**
 - BEDtools

Data Libraries

- Top menu -> Shared Data -> Data Libraries
- Access restricted by permissions

Data Libraries

search dataset name, info, message, dbkey



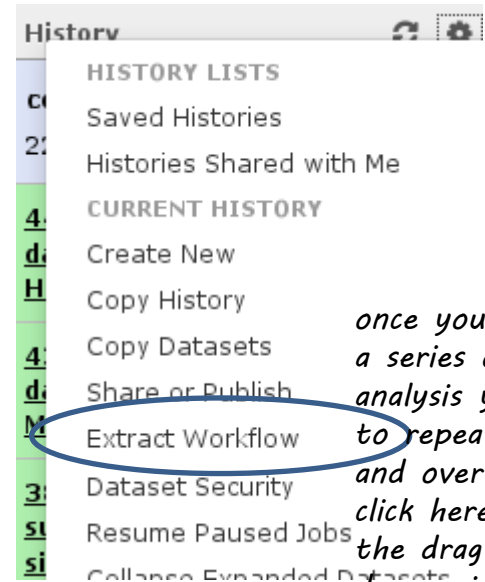
[Advanced Search](#)

<u>Data library name ↓</u>	<u>Data library description</u>
Akhtar	currently only Ken's data
Course	Data for Galaxy Course
Fukao sequencing runs	
Genomes + Annotations	reference genomes and all kinds of annotations publicly available for everyone
Jenuwein	Inti's and Aydan's shared data
Jenuwein sequencing runs	
Personal folders	where users can store their important datasets
Saccani sequencing runs	mapped data from sequencing runs

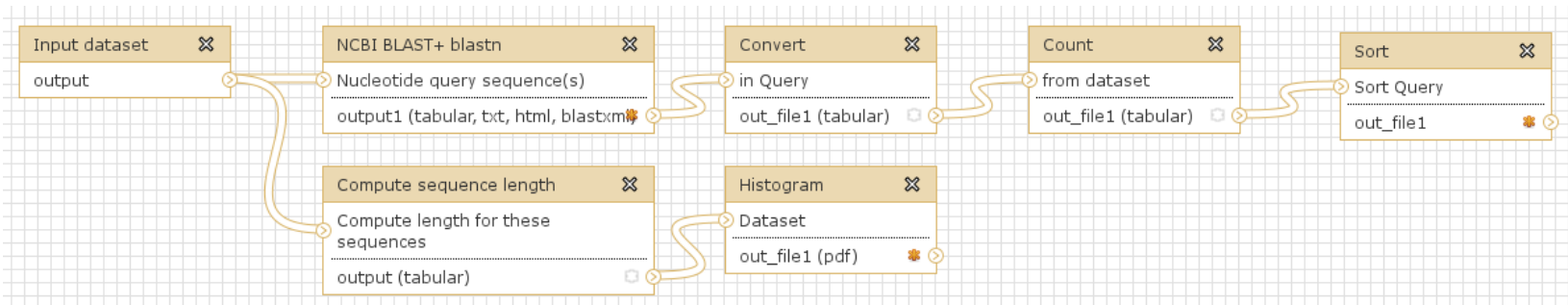
for
registered
users only!

Workflows

- automate repeating analysis
- help to stay organized
- share them
- use other people's (e.g. those we provide)



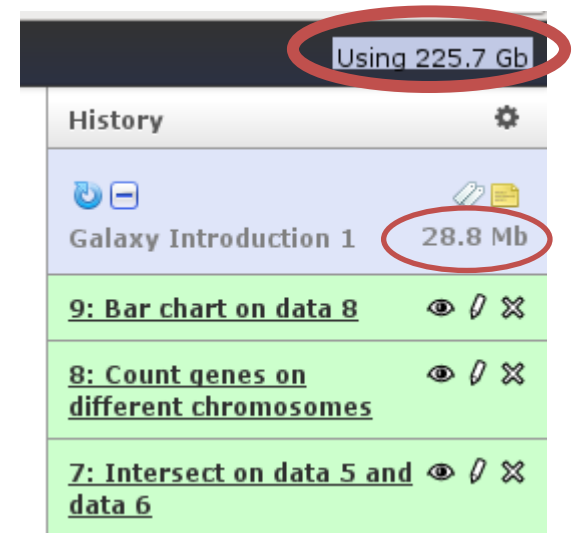
*once you've done
a series of
analysis you'd like
to repeat over
and over again,
click here & enjoy
the drag and
drop principle of
the workflows*



*don't forget to save and give the workflow a meaningful name! you can
always re-use and edit them via the **WORKFLOW** menu*

Some words of caution

- Watch your disk usage!
- depending on the size of the data and the tool you're using, tasks can run between seconds to ca. 15 minutes
- You cannot upload data > 2 GB through your browser (use FTP instead)



Help

- general Galaxy help:
wiki.galaxyproject.org/Learn
- specific deepTools Galaxy:
deeptools@googlegroups.com
- send bug reports and we will get in touch

