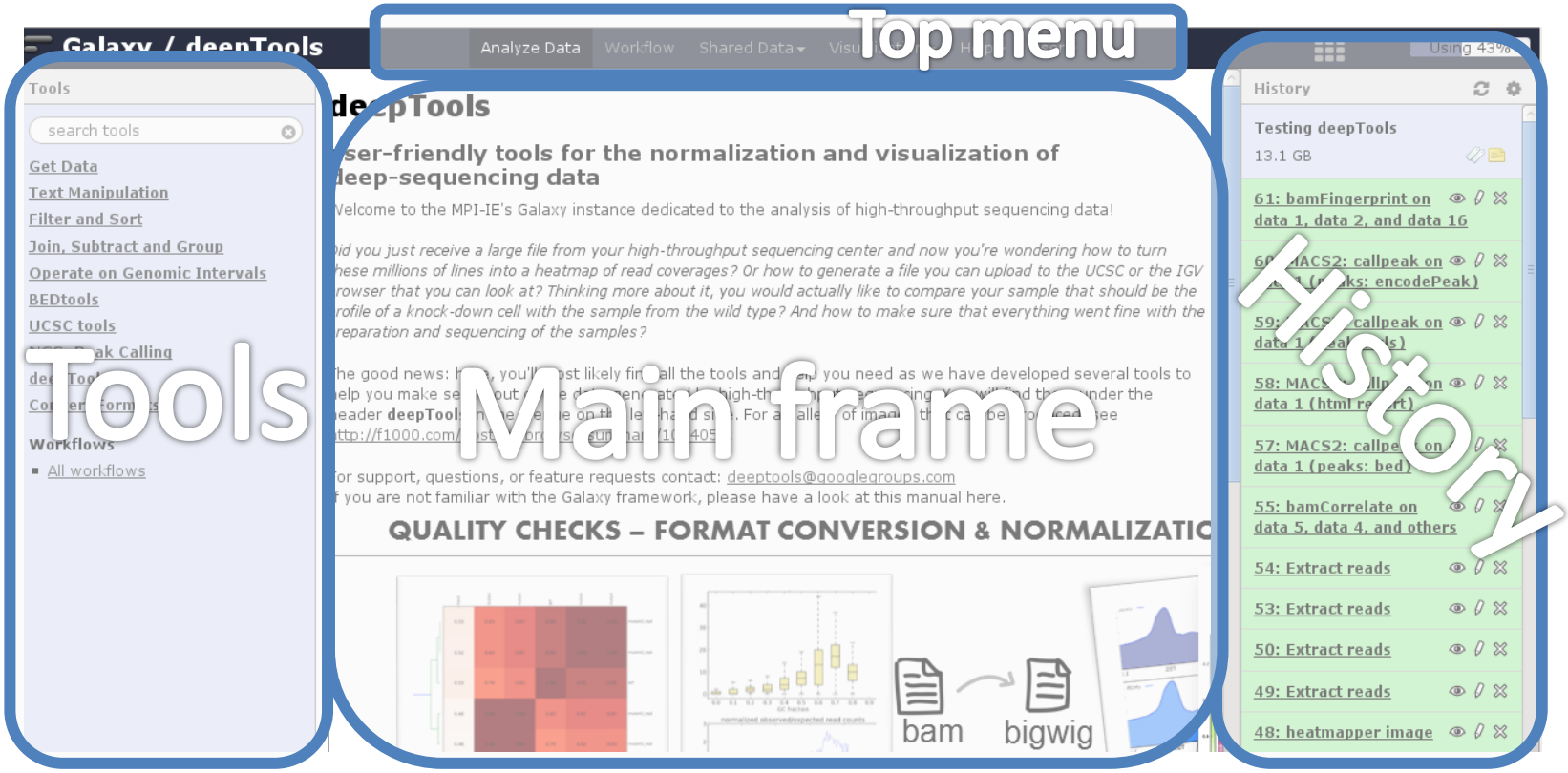
MAX PLANCK INSTITUTE
OF IMMUNOBIOLOGY AND EPIGENETICS

Introduction to the Galaxy framework

Sarah Diehl

Bioinformatics Facility, MPI-IE, Freiburg

The start site
deeptools.ie-freiburg.mpg.de



only function that requires login

Data Libraries

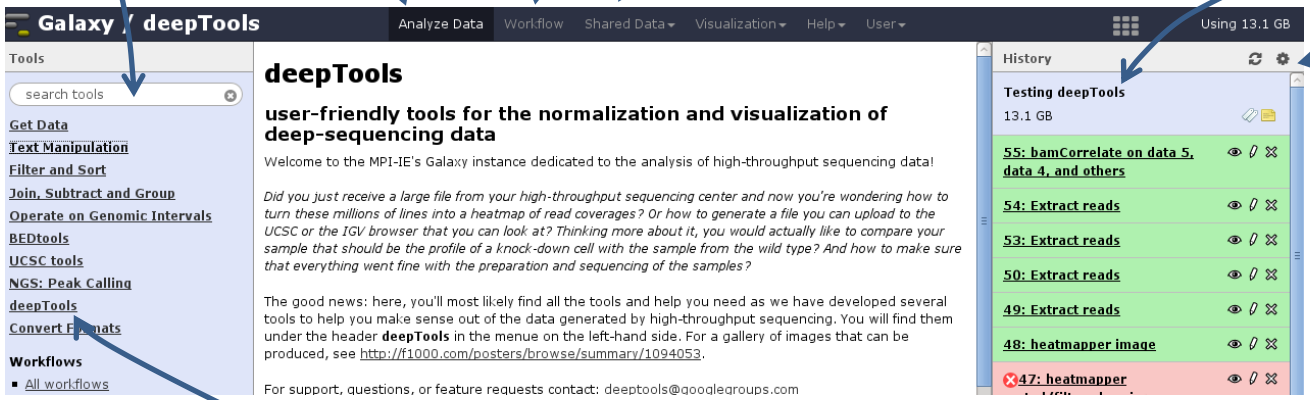
Published Histories
Published Workflows
Published Visualizations
Published Pages

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

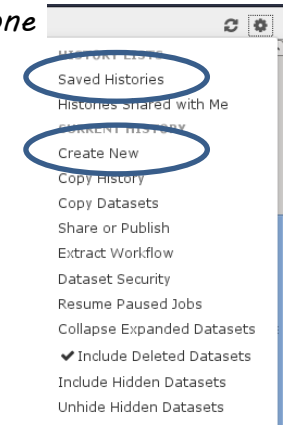
way to go for sample data, workflows and informative pages

free text tool search

give meaningful history name (just click on it)



to create new history or to access a saved one



TOOL PANEL

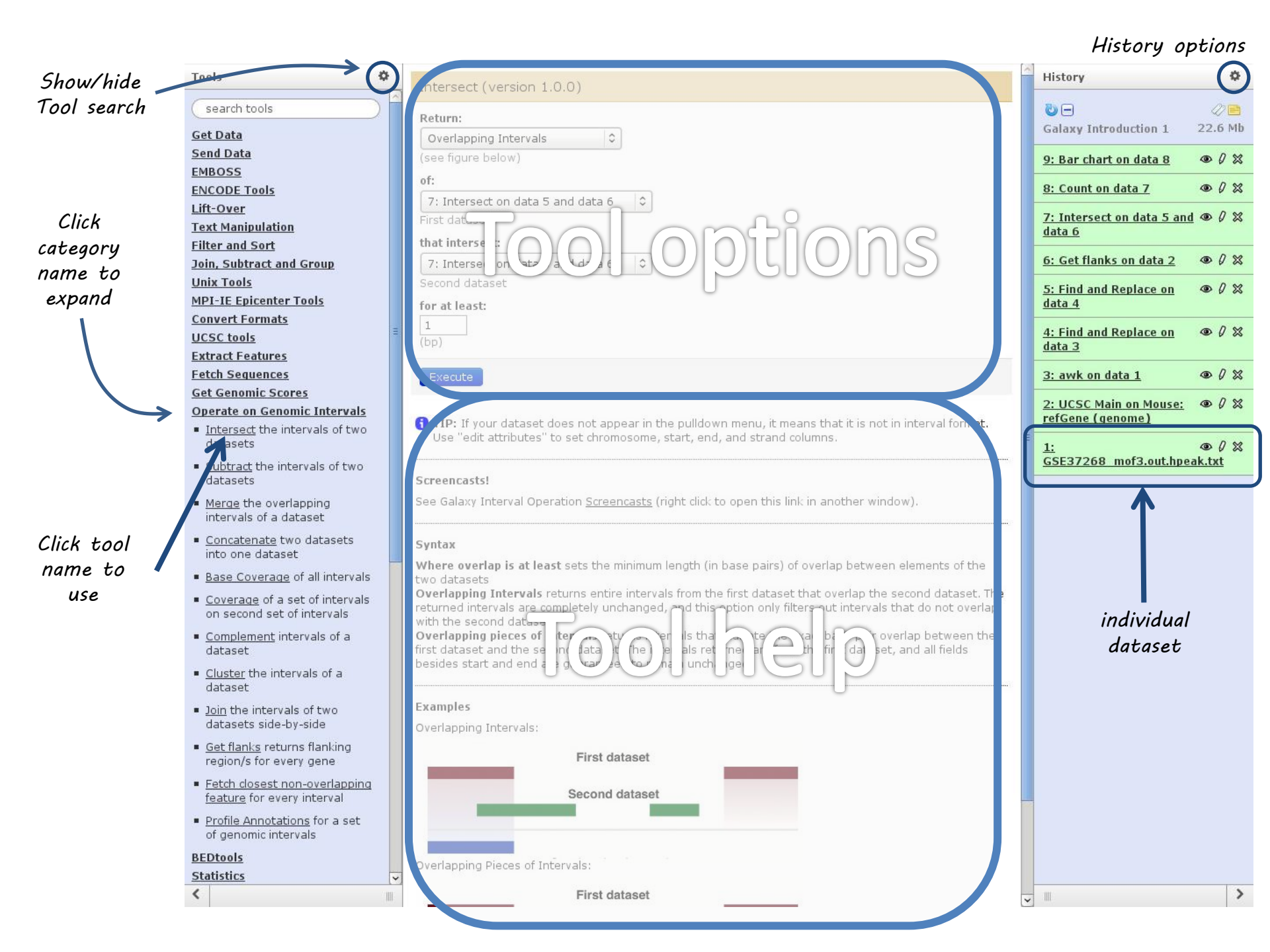
→ contains all tools installed in this Galaxy instance

deepTools = central tools for NGS data analysis

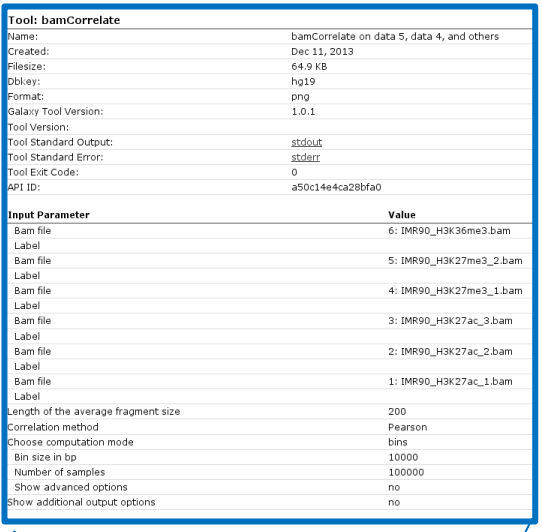
HISTORY PANEL

→ 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)


History options
Show/hide Tool search
Click category name to expand
Click tool name to use
Tool options
Tool help
individual dataset
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 intersect:
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 intervals returns intervals that represent the overlap between the first dataset and the second dataset. The intervals returned have the first dataset, and all fields besides start and end are unchanged.
Examples
Overlapping Intervals:
First dataset
Second dataset
Overlapping Pieces of Intervals:
First dataset
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

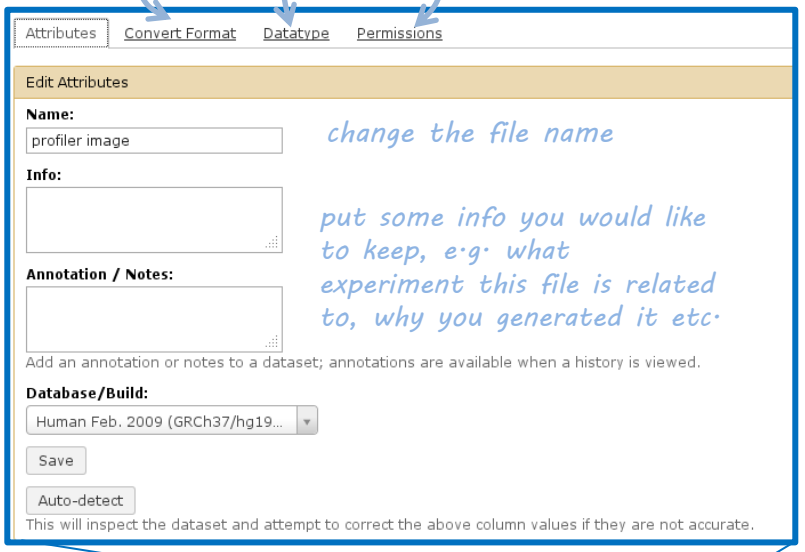


Tool: bamCorrelate	
Name:	bamCorrelate on data 5, data 4, and others
Created:	Dec 11, 2013
Filesize:	64.9 KB
Dbkey:	hg19
Format:	png
Galaxy Tool Version:	1.0.1
Tool Version:	
Tool Standard Output:	stdout
Tool Standard Error:	stderr
Tool Exit Code:	0
API ID:	a50c14e4ca28bfa0

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

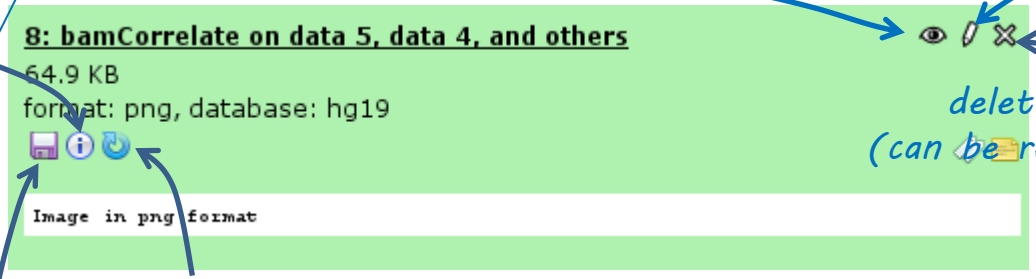
details about how this file was generated

have a look here as well



edit attributes

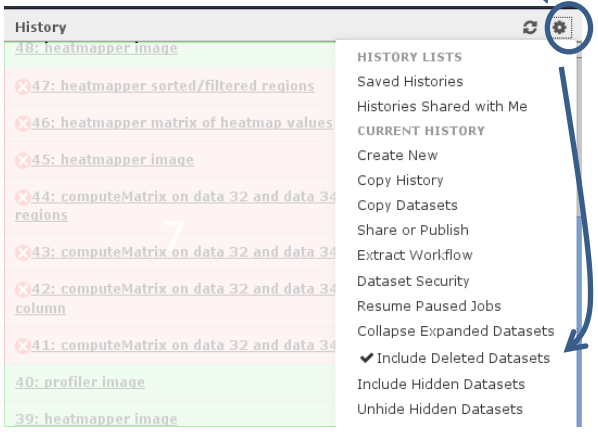
view the file



delete the file (can be recovered)

download the file

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



States of data sets

Waiting to be run

5: Find and Replace on data 4

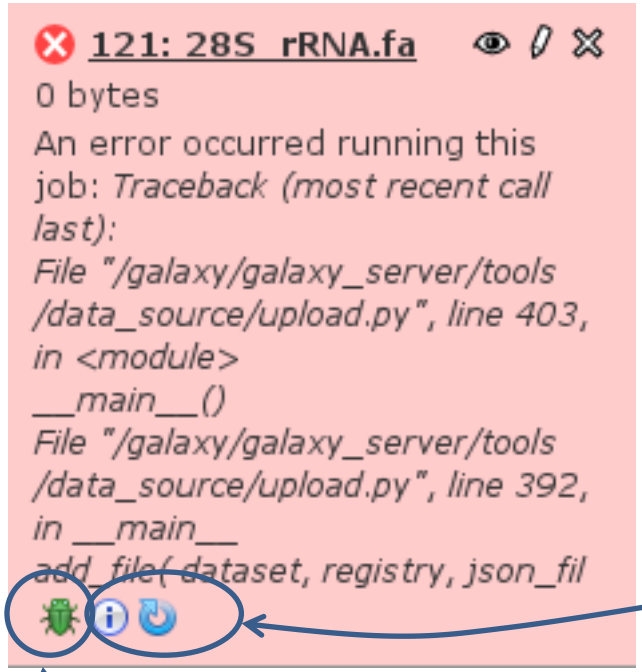
Running

3: Compute on data 2

Finished successfully

7: Intersect on data 5 and data 6

Failed



there are several reasons for a failure:

- wrong input data
- input data set doesn't have the correct format
- program bug
- …

check whether all the parameters you chose were correct

send a bug report to us

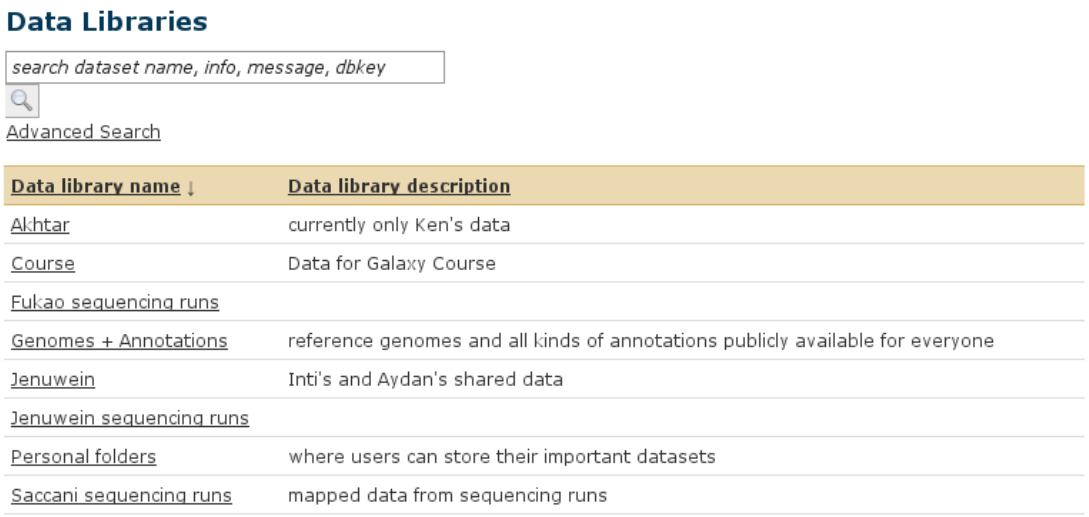
Tools

- 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

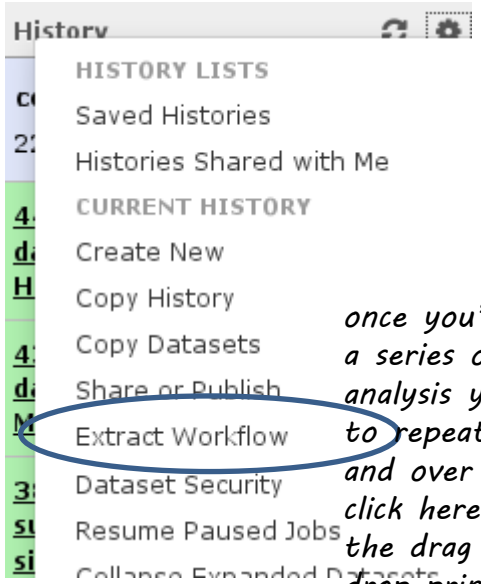
Advanced Search

Data library name ↓	Data library description
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

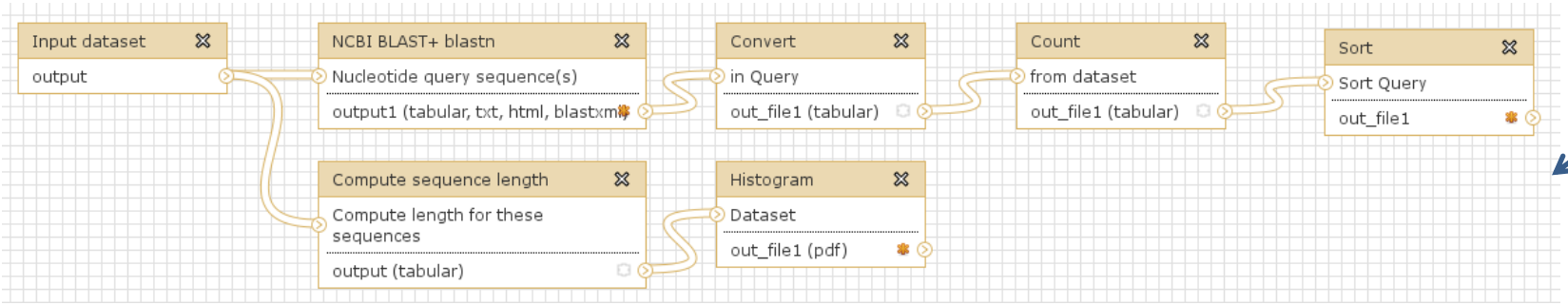
for registered users only!

Workflows

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



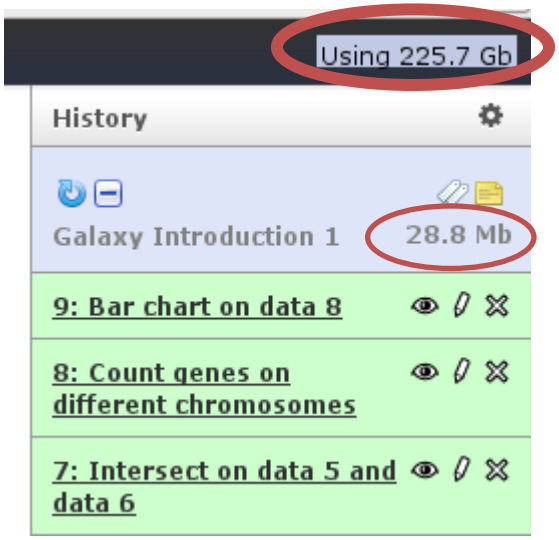
once you've done a series of analysis you'd like to repeat over and over again, click here & enjoy the drag and drop principle of the workflows



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