

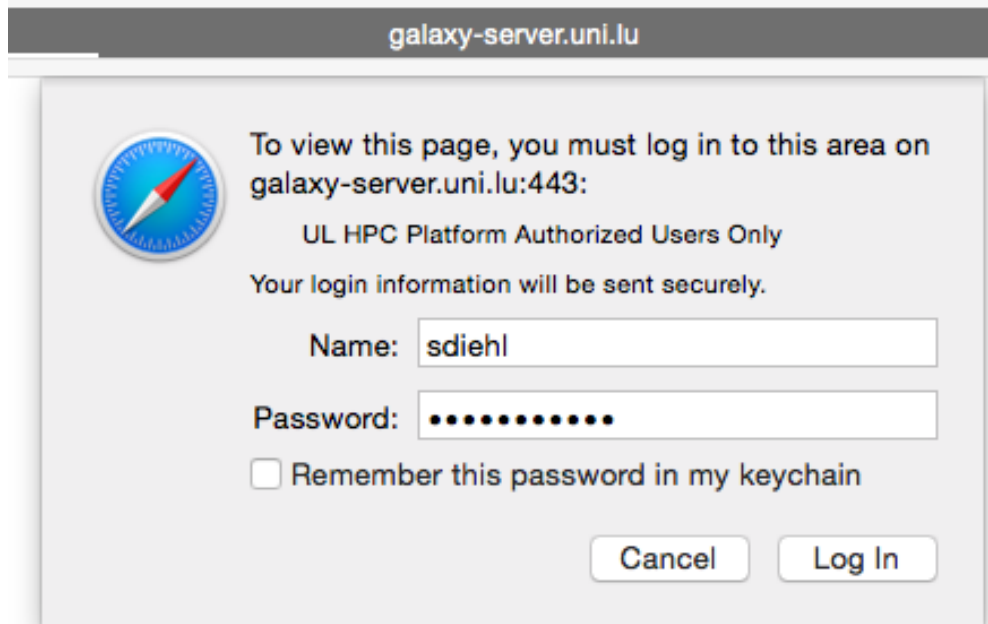
# Introduction to Galaxy

HPC @ Uni.lu and LCSB

Sarah Diehl

# Log in

- <http://galaxy-server.uni.lu>
- Accept certificate and login with your HPC credentials



The screenshot shows a login dialog box titled "galaxy-server.uni.lu". On the left is a circular icon with a blue background and a red and white compass needle. The text inside the dialog reads: "To view this page, you must log in to this area on galaxy-server.uni.lu:443:", "UL HPC Platform Authorized Users Only", and "Your login information will be sent securely." Below this are two input fields: "Name:" with the text "sdiehl" and "Password:" with a series of dots. At the bottom left is a checkbox labeled "Remember this password in my keychain". At the bottom right are two buttons: "Cancel" and "Log In".

galaxy-server.uni.lu

To view this page, you must log in to this area on galaxy-server.uni.lu:443:

UL HPC Platform Authorized Users Only

Your login information will be sent securely.

Name: sdiehl

Password: .....

☐ Remember this password in my keychain

Cancel Log In

Galaxy / Uni LU

Analyze DataWorkflowShared DataVisualizationHistory

Tools

search tools

Get DataSend DataLift-OverText ManipulationFilter and SortJoin, Subtract and GroupConvert FormatsOperate on Genomic IntervalsBED ToolsdeepToolsExtract FeaturesFetch SequencesFetch AlignmentsGet Genomic ScoresStatisticsGraph/Display DataEvolutionNGS: QC and manipulationNGS: MappingNGS: GATK ToolsNGS: PicardNGS: Variant AnalysisNGS: SimulationPhenotype AssociationDebugWorkflowsAll workflows

Hello world! It's running...

To customize this page edit static/welcome.html

Galaxy is an open, web-based platform for data intensive biomedical research. The Galaxy team is a part of BX at Penn State, and the Biology and Mathematics and Computer Science departments at Emory University. The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Emory University.

History

search datasets

Galaxy Introduction10 shown26.0 MB

10: Bar chart on data 99: Count on data 88: Intersect data 5 and data 77: Get flanks on data 6: Convert Genomic Intervals To BED on data 55: Replace Text on data 44: Replace Text on data 33: Text reformatting on data 12: UCSC Main on Mouse: refGene (genome)1: GSE37268\_mof3.out.h peak.txt

Top menu

Tools

Main frame

History

Galaxy / Uni LU

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 9%

Tools

search tools

Get Data

Send Data

Lift-Over

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Operate on Genomic Intervals

Subject the intervals of two datasets

Merge the overlapping 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

Coverage of a set of intervals on second set of intervals

Complement intervals of a dataset

Subtract Whole Dataset from another dataset

Intersect the intervals of two datasets

Concatenate two datasets into one dataset

Base Coverage of all intervals

Cluster the intervals of a dataset

BED Tools

deepTools

Concatenate two datasets into one dataset (Galaxy Tool Version 1.0.1)

Options

Concatenate

8: Intersect on data 5 and data 7

First dataset

with

8: Intersect on data 5 and data 7

Second dataset

Both datasets are same filetype?

Yes No

If unchecked Second dataset will be forced into format of First dataset

Execute

TIP: If your dataset does not appear in the pulldown menu -> 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

Both datasets are exactly the same filetype will preserve all extra fields in both files. Leaving this unchecked will force the second dataset to use the same column assignments for chrom, start, end and strand (but not the extra fields with the exception). In such cases, the output fields are truncated or padded with fields to maintain a fully tabular output.

Example

First dataset

Second dataset

Concatenated intervals

History

search datasets

Galaxy Introduction

10 shown

26.0 MB

10: History options

9: Count on data 8

8: Intersect on data 5 and data 7

7: Get flanks on data 2

6: Convert Genomic Intervals To BED on data 5

5: Replace Text on data 4

4: Replace Text on data 3

3: Text reformatting on data 1

2: UCSC Main on Mouse: refGene (genome)

1: GSE37268\_mofa\_out.h peak.txt

Upload data

Click category name to expand

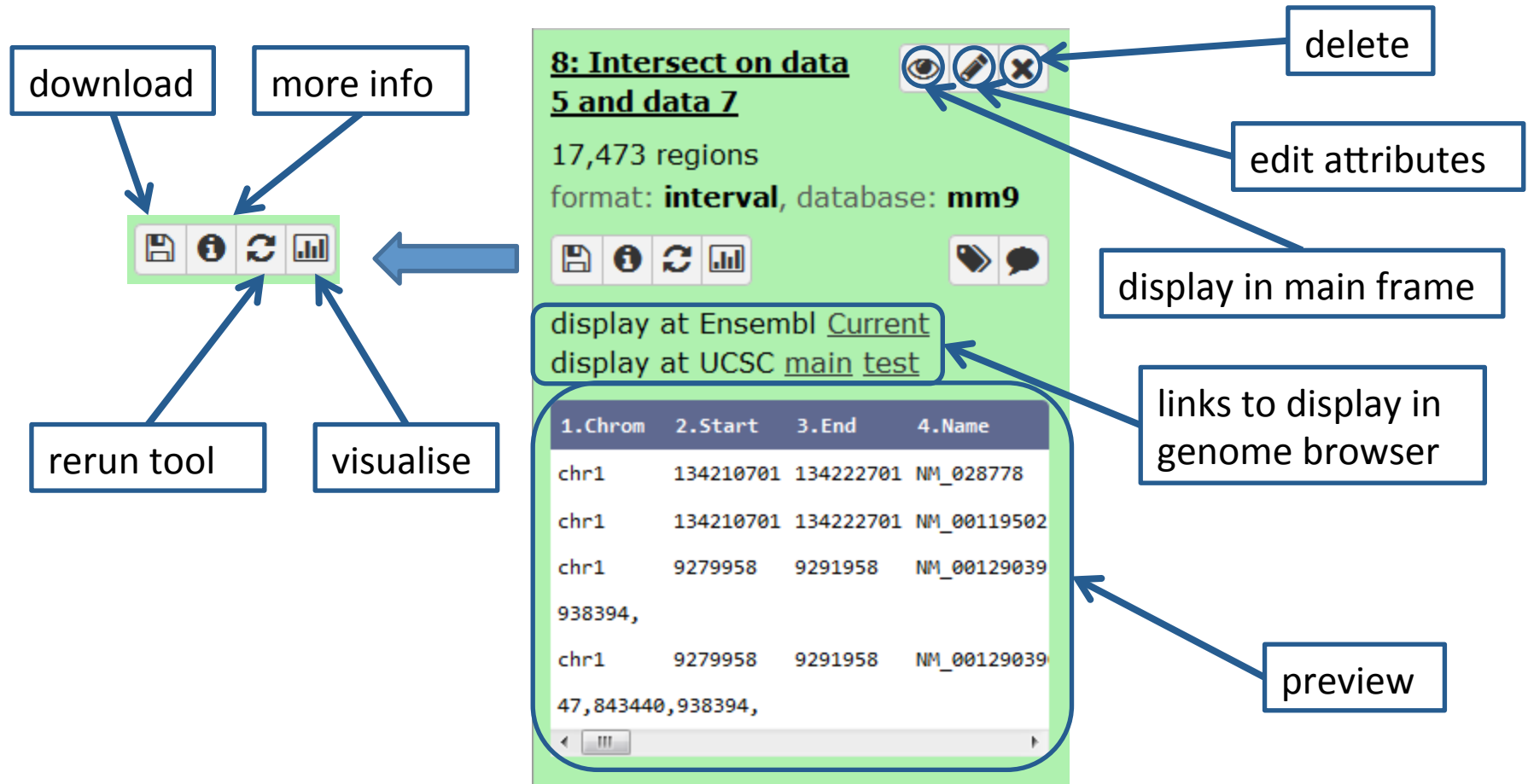
Click tool name to use

Tool options

Tool help

dataset





# Datasets










Click dataset name to expand





# Dataset states

- Waiting to be run
- Running
- Successfully finished
- Failed

 **6: Compute on data 1**   

 **7: Compute on data 1**   




**7: Compute on data 1**   

 **9: Bar chart on data 8**   

error

An error occurred with this dataset:

```
Traceback (most recent call last):
  File "/usr/local/galaxy/galaxy-dist/tools/pl
    import Gnuplot, Gnuplot.funcutils
ImportError: No module named Gnuplot
```

Send bug report

# Formats: tabular

13122	Hist1h2ai	-1.09803337373210	1.99391309961338	13
33790	Cenpi	-1.31045935685183	2.92807115314139	X
17603	Tcf19	-1.41017188366083	4.5199737219041	17
29570	Depdc1a	-1.74134731960069	5.22738553353615	3
32663	Anln	-1.76637339700090	4.82842251330819	9

- Column-based format
- Separated by tabs
- Similar to Excel

# Formats: interval

chr1	3660676	3661050	375	210	62.0876250438913	-2.00329386666667
chr1	3661326	3661500	175	102	28.2950833625942	-0.695557142857143
chr1	3661976	3662325	350	275	48.3062708406486	-1.29391285714286
chr1	3984926	3985075	150	93	34.1879823073944	-0.816992
chr1	4424801	4424900	100	70	26.8023246007435	-0.66282
chr1	4482601	4482775	175	77	32.2288894195497	-0.778994285714286
chr1	4775576	4775875	300	210	46.3134120503457	-1.27111133333333
chr1	4804026	4804125	100	85	28.2335379387586	-0.715186
chr1	4832226	4832325	100	97	29.0016223214396	-0.727826

- For genomic regions
- Column 1: chromosome
- Column 2: start position
- Column 3: end position
- Other columns: anything



# Formats: BED

chr1	134212701	134230065	NM_028778	0	+
chr1	134212701	134230065	NM_001195025	0	+
chr1	8352741	9289811	NM_027671	0	-
chr1	25124320	25886552	NM_175642	0	-
chr1	33510655	33726603	NM_008922	0	-
chr1	58714963	58752833	NM_175370	0	-

- Most common format for genomic regions  
[genome.ucsc.edu/FAQ/FAQformat.html#format1](http://genome.ucsc.edu/FAQ/FAQformat.html#format1)
- Column 1-3: same as interval
- Column 4: name
- Column 5: score
- Column 6: strand

# Formats: other

## Sequences

- FASTA  
[wikipedia.org/wiki/FASTA\\_format](http://wikipedia.org/wiki/FASTA_format)
- FASTQ  
[wikipedia.org/wiki/FASTQ\\_format](http://wikipedia.org/wiki/FASTQ_format)

## Coverage

- BedGraph  
[genome.ucsc.edu/goldenPath/help/bedgraph.html](http://genome.ucsc.edu/goldenPath/help/bedgraph.html)
- Wiggle  
[genome.ucsc.edu/goldenPath/help/wiggle.html](http://genome.ucsc.edu/goldenPath/help/wiggle.html)
- BigWig(gle)  
[genome.ucsc.edu/goldenPath/help/bigWig.html](http://genome.ucsc.edu/goldenPath/help/bigWig.html)

# Data Libraries

- Top menu -> Shared Data -> Data Libraries
- Access restricted by permissions
- Deposit for generally useful data
- Place for sharing
- Access to your project data (by request)

The screenshot shows the Galaxy / Uni LU interface. The top navigation bar includes links for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. A 'Using 8%' status indicator is on the right. The 'Data Libraries' section is active, displaying a search bar with the placeholder text 'search dataset name, info, message, dbke' and a magnifying glass icon. Below the search bar is a link for 'Advanced Search'. A table with the header 'Data library name↓' is shown, containing the text 'No Items'. A dropdown menu is open under the 'Shared Data' link, listing 'Data Libraries', 'Data Libraries Beta', 'Published Histories', 'Published Workflows', 'Published Visualizations', and 'Published Pages'.

**Galaxy / Uni LU** Analyze Data Workflow **Shared Data** Visualization Admin Help User Using 8%

## Data Libraries

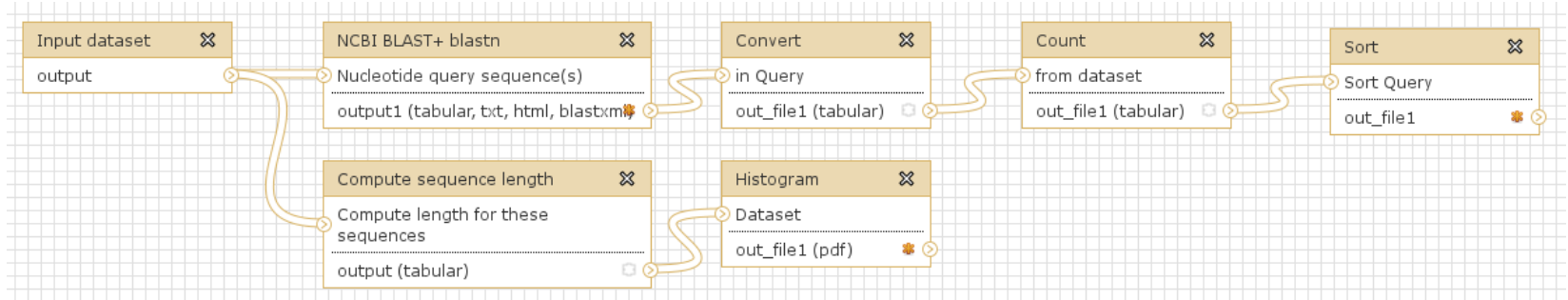
search dataset name, info, message, dbke 🔍

[Advanced Search](#)

Data library name↓
No Items

- Data Libraries
- Data Libraries Beta
- Published Histories
- Published Workflows
- Published Visualizations
- Published Pages

# Workflows



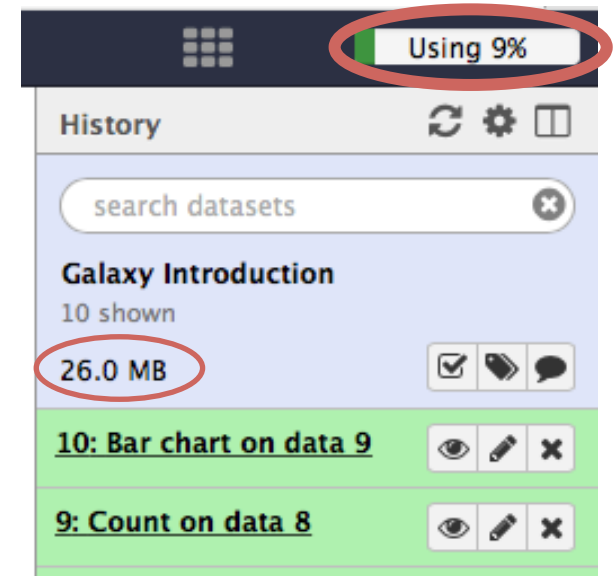
- Automate repeating analysis
- Help to stay organized
- Share them
- Use other people's

# Why use local server?

- Special tools
- More disk space and computing power for you
- Data stays in-house
- Fast and personalised help

# Warnings

- Watch your disk usage!
- For some tools or tasks it's perfectly normal to run for a day or more
- You cannot upload data > 2 GB through your browser



# Help

- Top Menu -> Help
- Send bug report
- Contact me