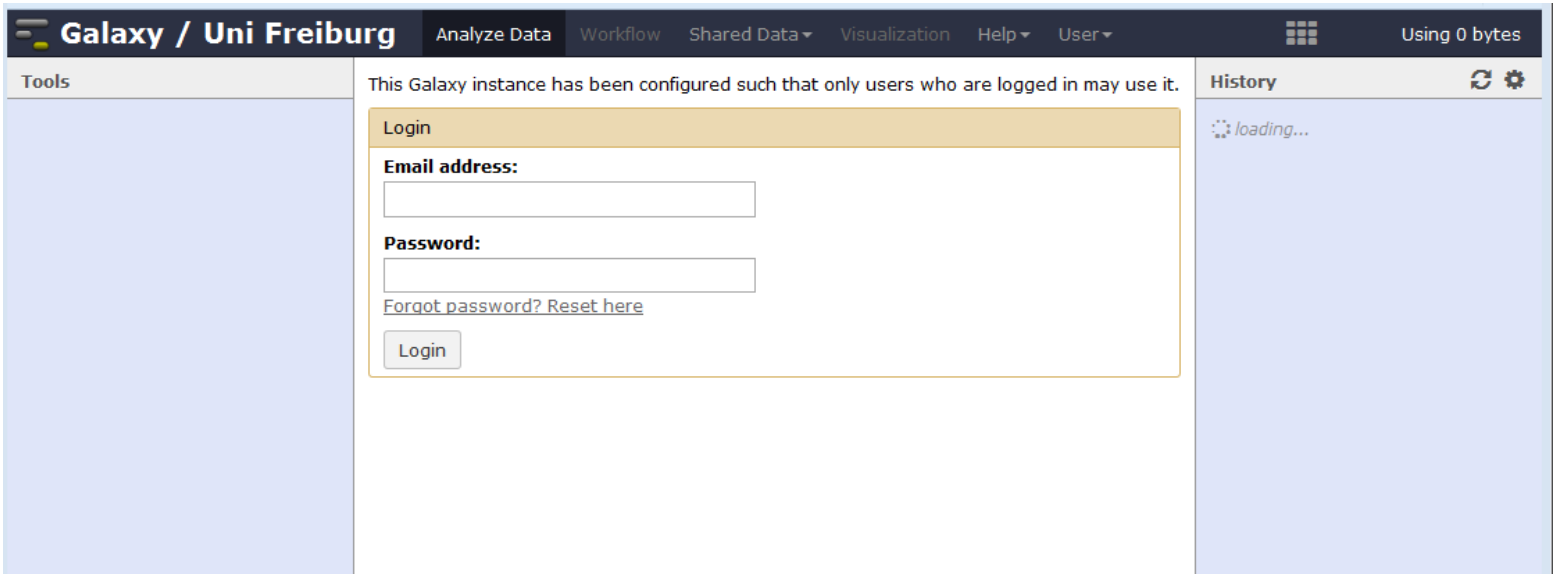
Introduction to Galaxy

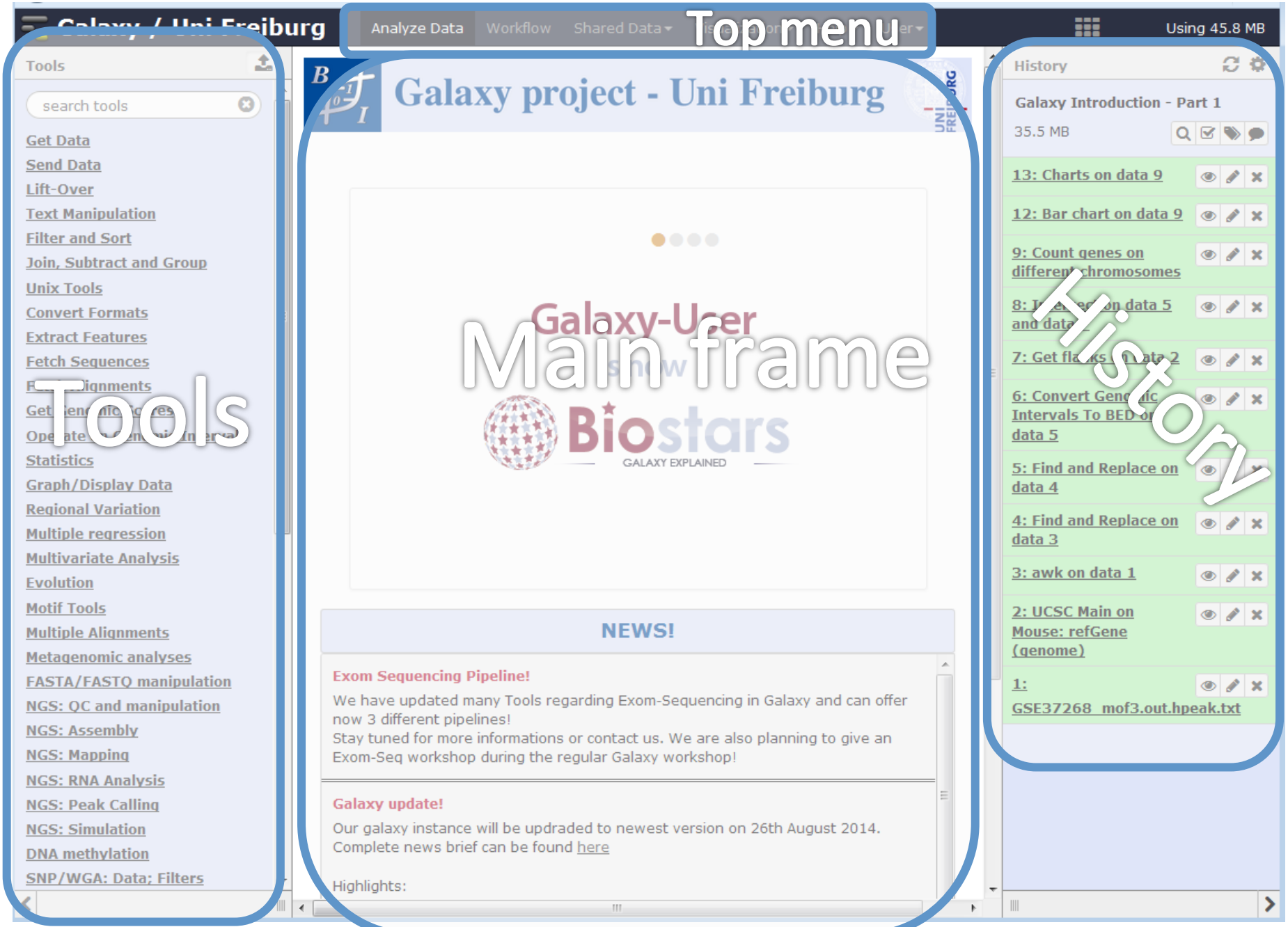
MPI for Immunobiology and Epigenetics

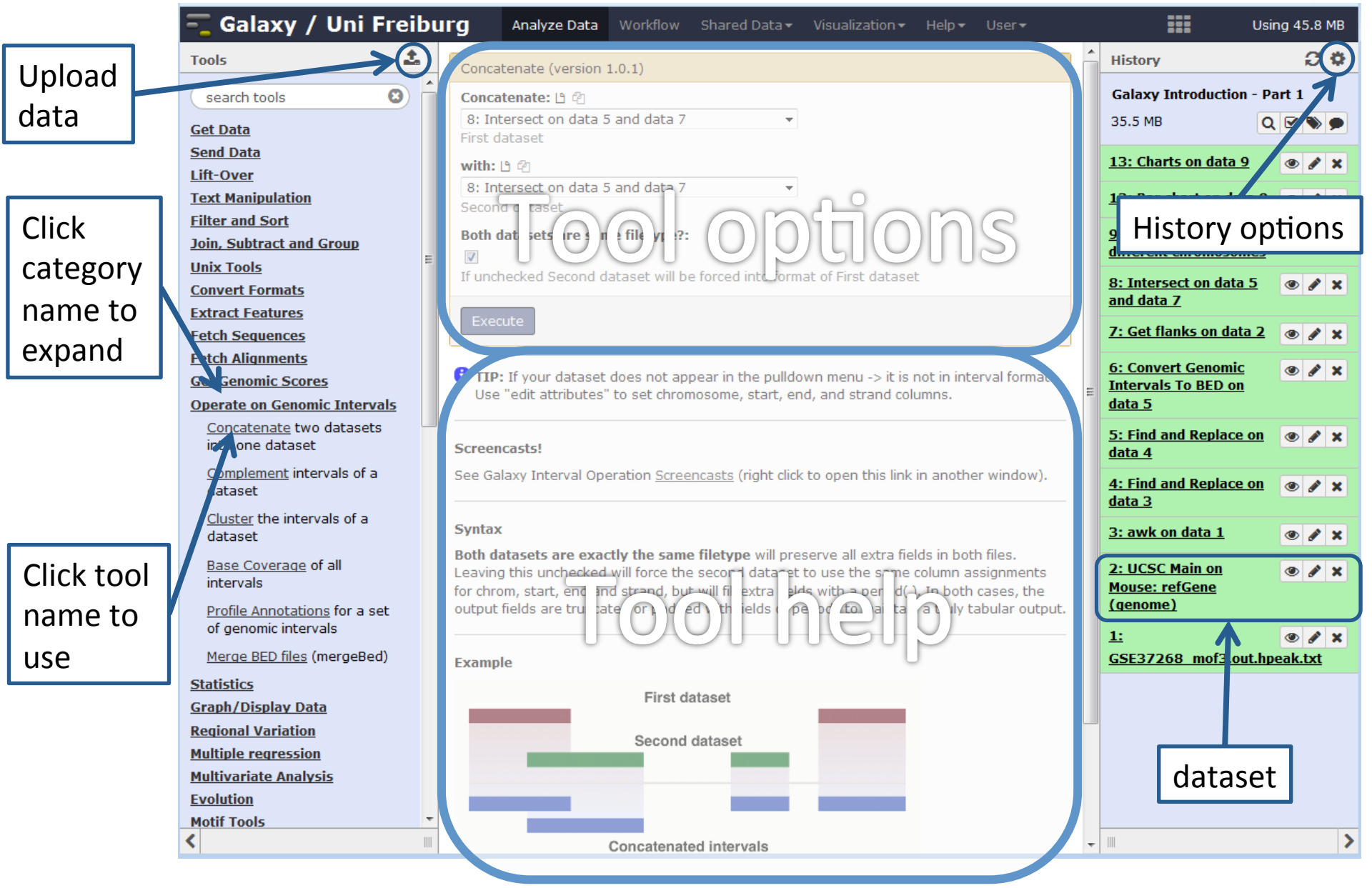
Sarah Diehl

Log in

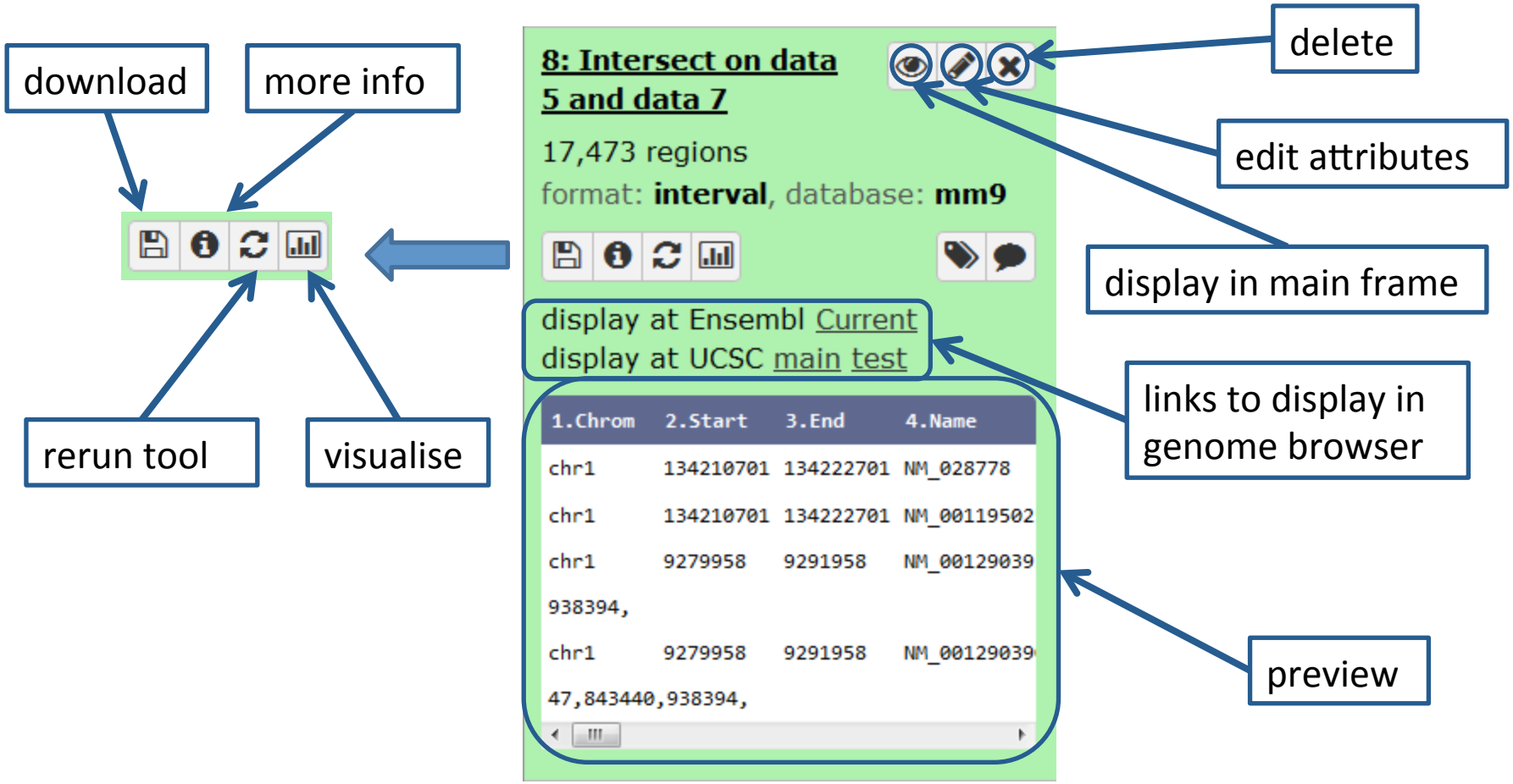
- http://galaxy.uni-freiburg.de




Galaxy / Uni Freiburg
Analyze Data
Workflow
Shared Data
Top menu
Using 45.8 MB
Tools
search tools
Get Data
Send Data
Lift-Over
Text Manipulation
Filter and Sort
Join, Subtract and Group
Unix Tools
Convert Formats
Extract Features
Fetch Sequences
Statistics
Graph/Display Data
Regional Variation
Multiple regression
Multivariate Analysis
Evolution
Motif Tools
Multiple Alignments
Metagenomic analyses
FASTA/FASTQ manipulation
NGS: QC and manipulation
NGS: Assembly
NGS: Mapping
NGS: RNA Analysis
NGS: Peak Calling
NGS: Simulation
DNA methylation
SNP/WGA: Data; Filters
Tools
Galaxy project - Uni Freiburg
Galaxy-User
Main frame
Biostars
GALAXY EXPLAINED
NEWS!
Exom Sequencing Pipeline!
We have updated many Tools regarding Exom-Sequencing in Galaxy and can offer now 3 different pipelines!
Stay tuned for more informations or contact us. We are also planning to give an Exom-Seq workshop during the regular Galaxy workshop!
Galaxy update!
Our galaxy instance will be updraded to newest version on 26th August 2014.
Complete news brief can be found here
Highlights:
History
Galaxy Introduction - Part 1
35.5 MB
13: Charts on data 9
12: Bar chart on data 9
9: Count genes on different chromosomes
7: Get flanks
6: Convert Genomic Intervals To BED
data 5
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
History


Upload data
Click category name to expand
Click tool name to use
Galaxy / Uni Freiburg
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 45.8 MB
Tools
search tools
Get Data
Send Data
Lift-Over
Text Manipulation
Filter and Sort
Join, Subtract and Group
Unix Tools
Convert Formats
Extract Features
Fetch Sequences
Fetch Alignments
Operate on Genomic Intervals
Concatenate two datasets
Complement intervals of a dataset
Cluster the intervals of a dataset
Base Coverage of all intervals
Profile Annotations for a set of genomic intervals
Merge BED files (mergeBed)
Statistics
Graph/Display Data
Regional Variation
Multiple regression
Multivariate Analysis
Evolution
Motif Tools
Concatenate (version 1.0.1)
Concatenate:
8: Intersect on data 5 and data 7
First dataset
with:
8: Intersect on data 5 and data 7
Tool options
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.
Tool help
Example
First dataset
Second dataset
Concatenated intervals
History
Galaxy Introduction - Part 1
35.5 MB
13: Charts on data 9
History options
8: Intersect on data 5 and data 7
7: Get flanks on data 2
6: Convert Genomic Intervals To BED on data 5
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
dataset

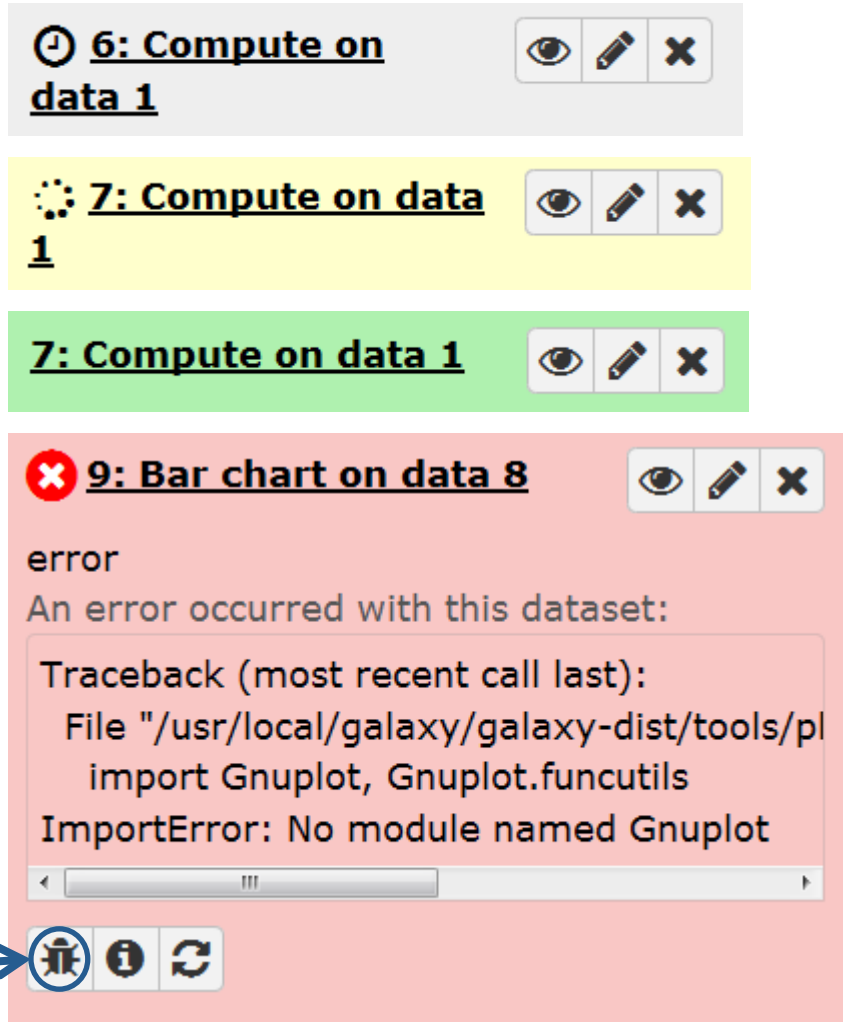
Datasets



Click dataset name to expand

Dataset states

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



Tools

- Useful categories:
  - Text manipulation
  - Join, Subtract and Group
  - Operate on Genomic Intervals
  - BED tools
- Only on our Galaxy:
  - BED tools
  - deepTools
  - NGS: Gemini
  - ChemicalToolBox
  - OpenMS
  - Etc.

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
- Column 1-3: same as interval
- Column 4: name
- Column 5: score
- Column 6: strand

Formats: other

Sequences

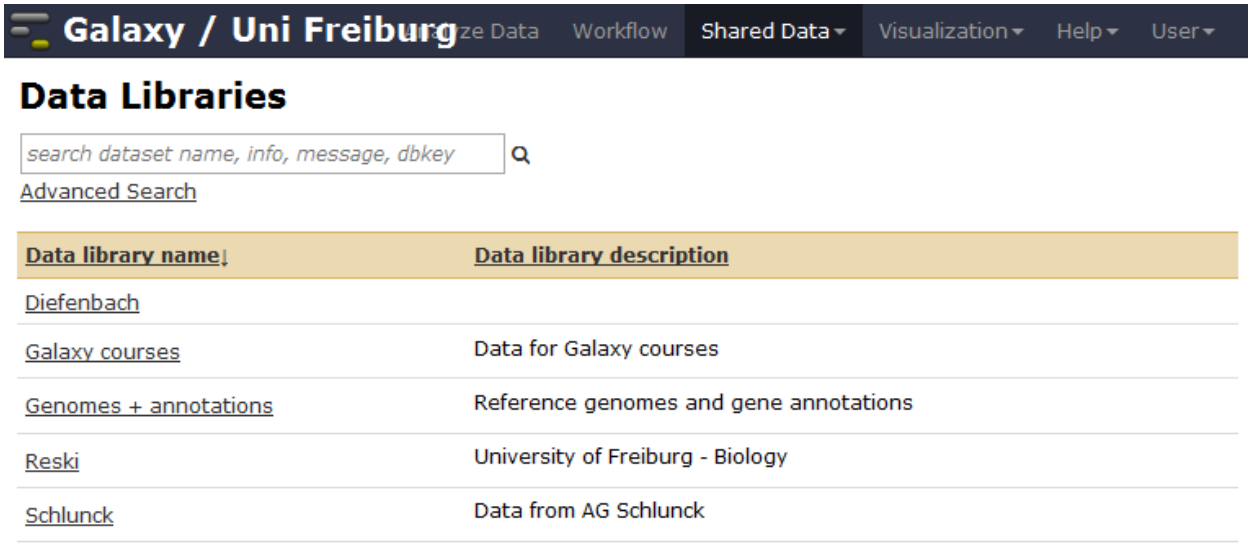
- FASTA wikipedia.org/wiki/FASTA_format
- FASTQ wikipedia.org/wiki/FASTQ_format

Coverage

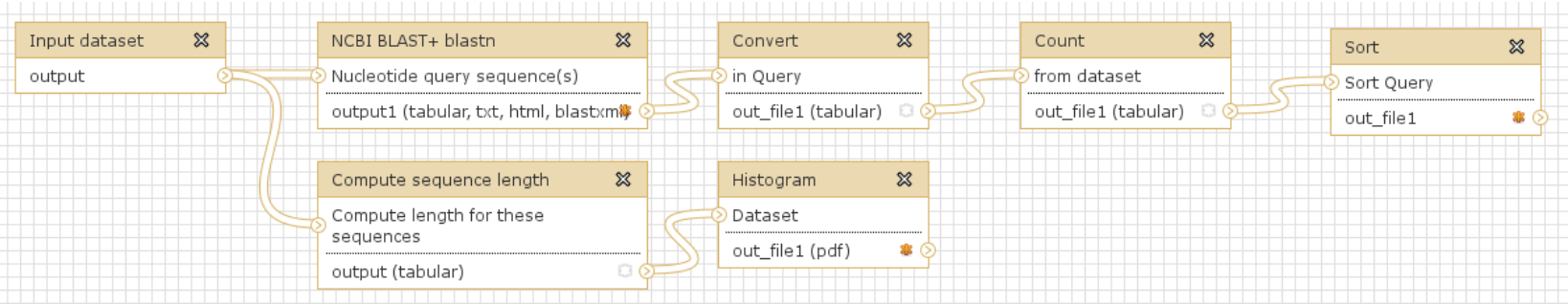
- BedGraph genome.ucsc.edu/goldenPath/help/bedgraph.html
- Wiggle genome.ucsc.edu/goldenPath/help/wiggle.html
- BigWig(gle) genome.ucsc.edu/goldenPath/help/bigWig.html

Data Libraries

- Top menu -> Shared Data -> Data Libraries
- Access restricted by permissions
- Deposit for generally useful data
- Place for sharing



Workflows

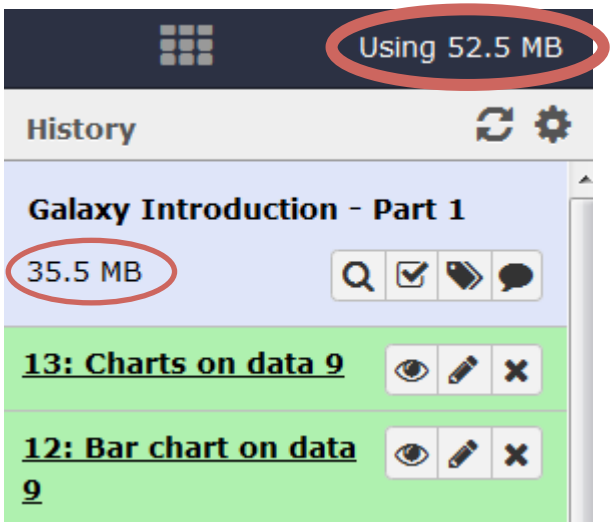


- Automate repeating analysis
- Help to stay organized
- Share them
- Use other people‘s (e.g. ones we provide you)

Why use local server?

- Special tools
- More disk space and computing power for you
- Data stays in-house
- Fast and personalized 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
- Galaxy Page: Getting help (coming soon)
- Send bug report
- Contact us