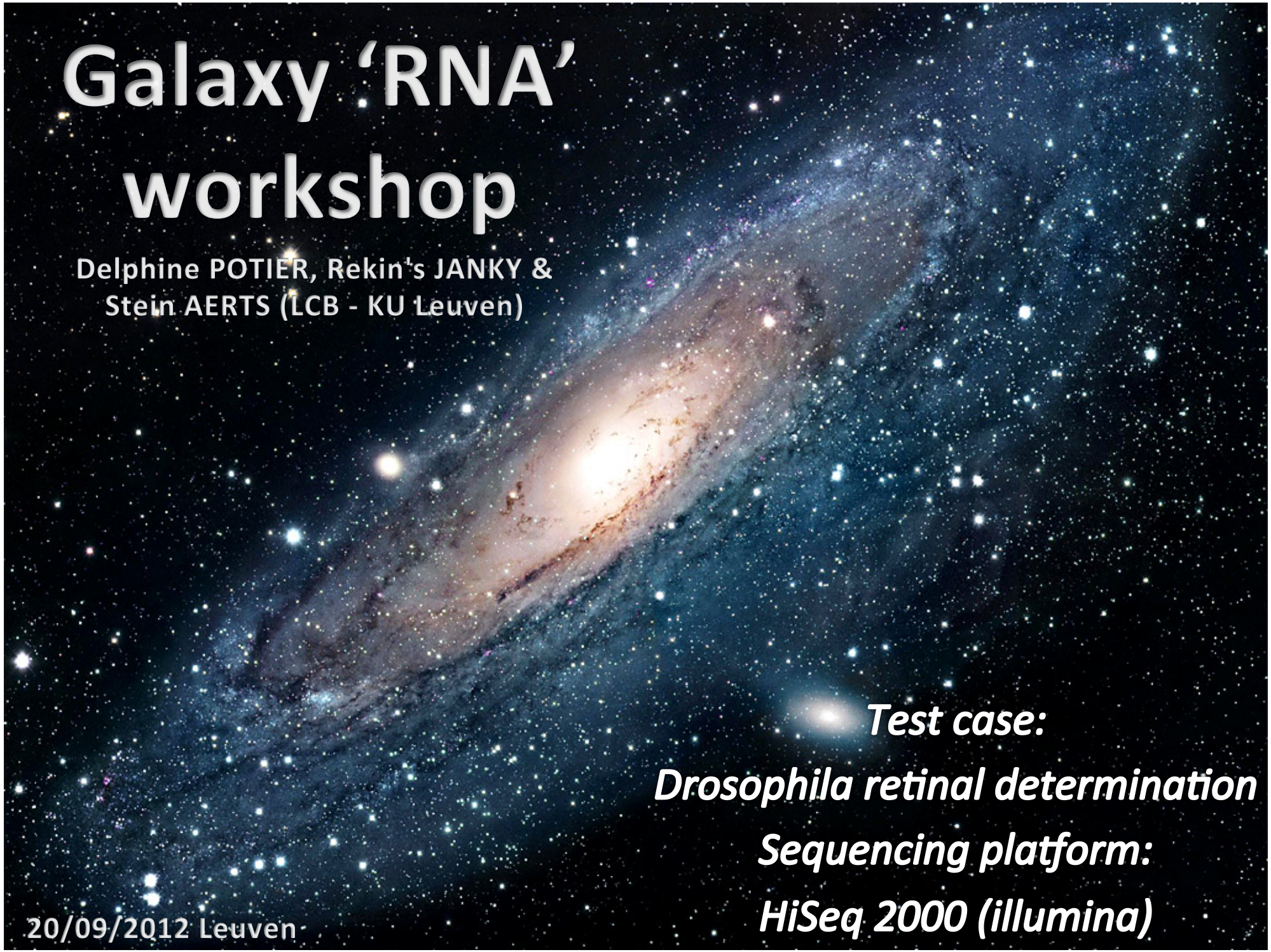
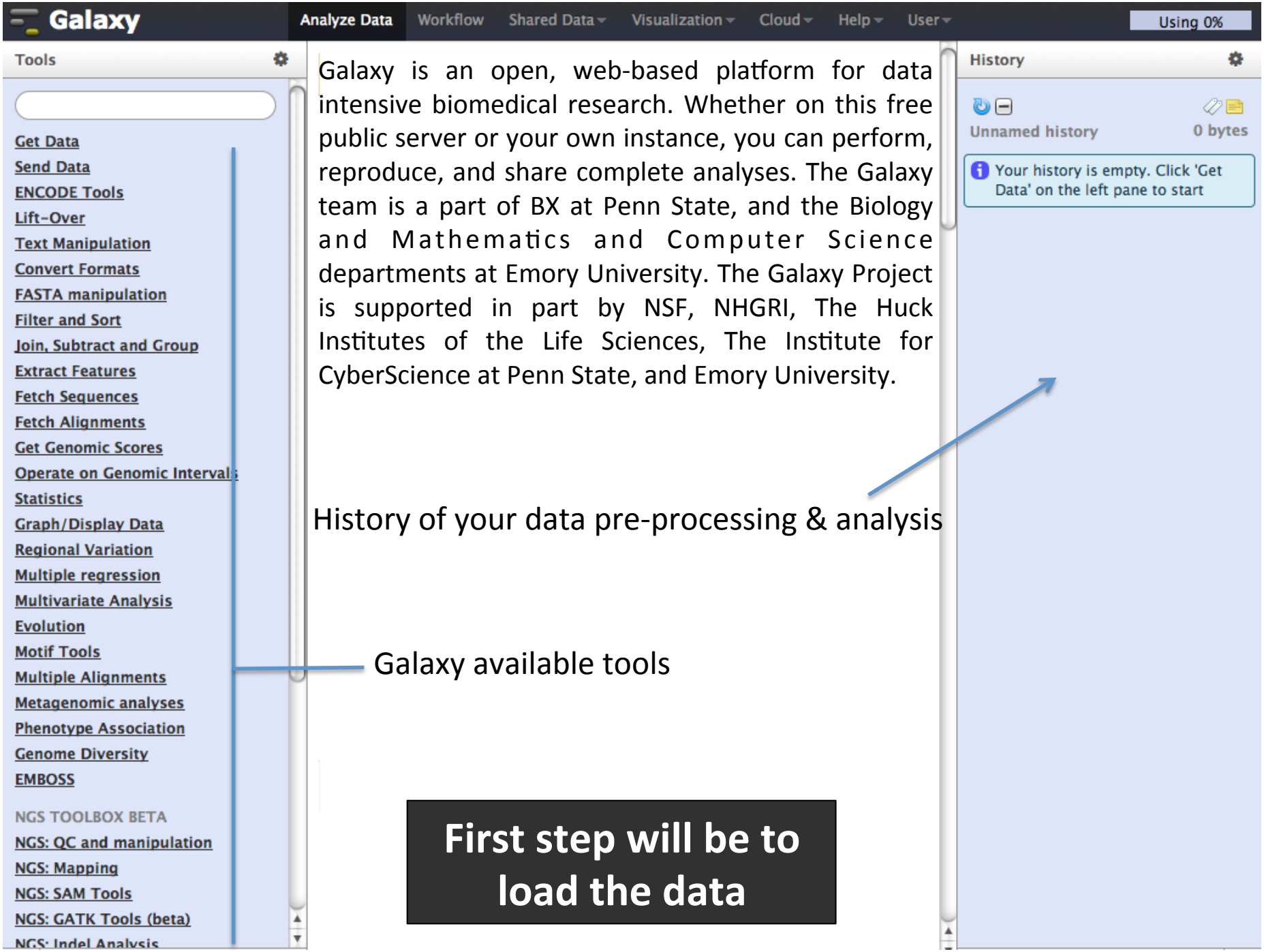

Galaxy 'RNA' workshop
Delphine POTIER, Rekin's JANKY & Stein AERTS (LCB - KU Leuven)
Test case:
Drosophila retinal determination
Sequencing platform:
HiSeq 2000 (illumina)
20/09/2012 Leuven


Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Cloud
Help
User
Using 0%
Tools
Get Data
Send Data
ENCODE Tools
Lift-Over
Text Manipulation
Convert Formats
FASTA manipulation
Filter and Sort
Join, Subtract and Group
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
Phenotype Association
Genome Diversity
EMBOSS
NGS TOOLBOX BETA
NGS: QC and manipulation
NGS: Mapping
NGS: SAM Tools
NGS: GATK Tools (beta)
NGS: Indel Analysis
Galaxy is an open, web-based platform for data intensive biomedical research. Whether on this free public server or your own instance, you can perform, reproduce, and share complete analyses. 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 NSF, NHGRI, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Emory University.
History
Unnamed history
0 bytes
Your history is empty. Click 'Get Data' on the left pane to start
History of your data pre-processing & analysis
Galaxy available tools
First step will be to load the data

RNA-seq data:

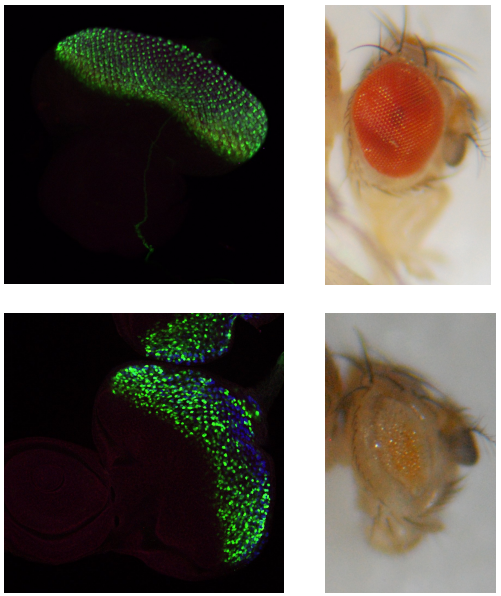
Eye development, glass mutant VS wild type eye/antennal imaginal discs

- The transcription factor '*glass*' is required for normal development of photoreceptor cells in eye imaginal disc during Drosophila retinal determination (larval stage 3).

- We will study the differential expression in gl mutants compared to wild type eye/antennal discs. To this aim will provide you single-end HiSeq2000 RNA-seq data from:

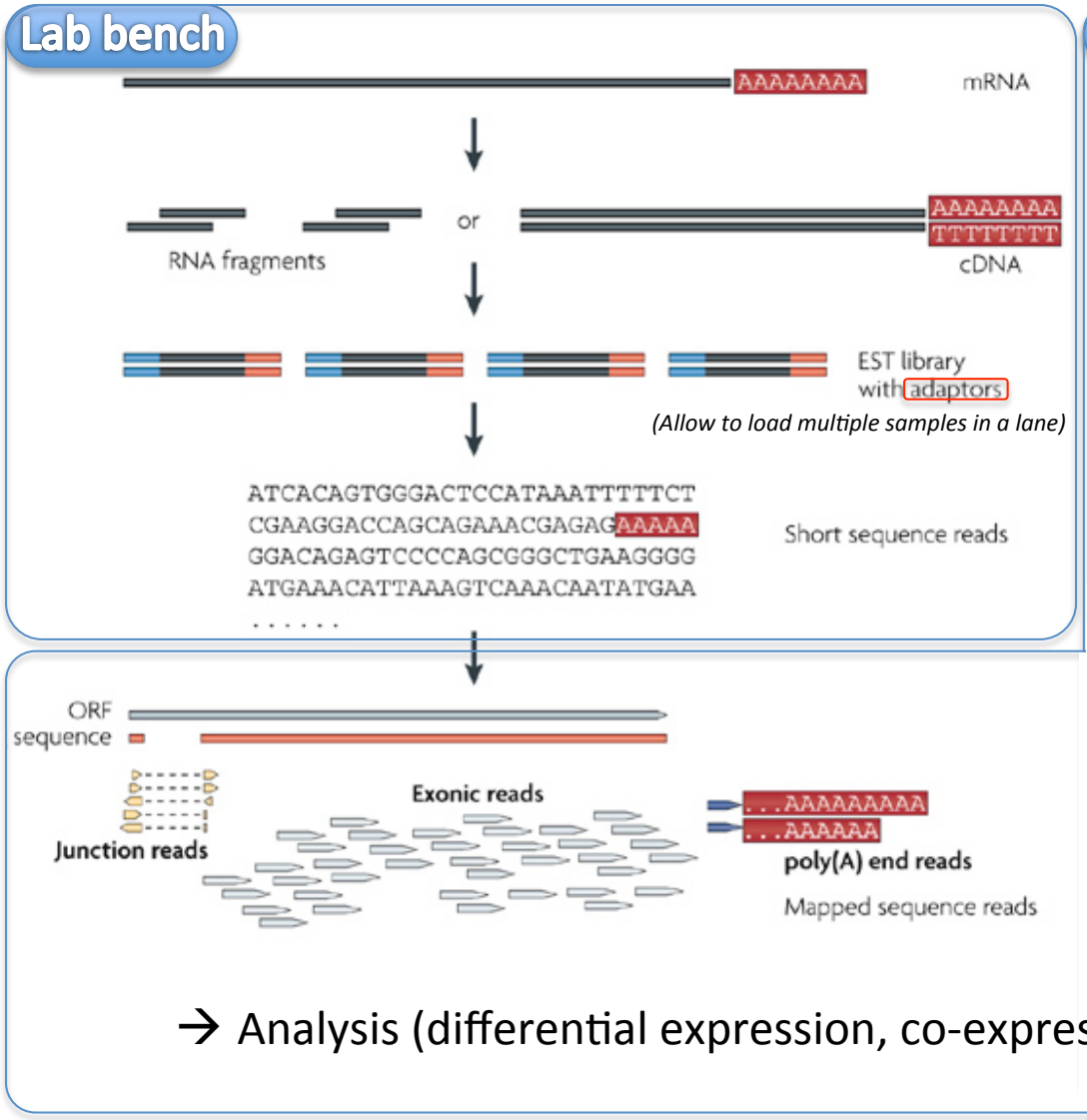
  – Wild type eye/antennal disc
  (2 replicates)

  – Glass mutant eye/antennal disc
  (2 replicates)



✳ *You will work on a partial dataset to reduce calculation time*
Results for the entire dataset are available here:
http://galaxy.bits.vib.be/u/delphine-potier/h/ngs-workshop

Bioinformatics treatment linked to lab work



Computer

→ Demultiplexing (usally down by the platform)

→ Adapter clipping (Fast-x clipper)

→ Checking reads quality (FastQC)

→ Mapping (tophat)

→ Visualization

→ Analysis (differential expression, co-expression, cis-regulation etc...)

Modified from Zhong Wang et al., RNA-Seq: a revolutionary tool for transcriptomics, Nat Rev Genet (2009)

Understand .fastq reads

@HWI-ST571_69:4:1101:11162:2311:ACAGTG/1	***Illumina sequence identifier***
ATTTATTTCTAATTGTTTATTATGTTTTTATTTCTT	***Sequence***
+HWI-ST571_69:4:1101:11162:2311:ACAGTG/1	
Gggggfggggeggggggggggggggggggggggg	***Encoded sequencing quality score for each base***

Illumina sequence identifier:

- **HWI-ST571_69** the unique instrument name
- **4** flowcell lane
- **1101** tile number within the flowcell lane
- **11162** 'x'-coordinate of the cluster within the tile
- **2311** 'y'-coordinate of the cluster within the tile
- **ACAGTG** index number for a multiplexed sample (0 for no indexing)
- **/1** the member of a pair, /1 or /2 *(paired-end or mate-pair reads only)*

From raw data to pre-processed data

- Demultiplexing

  → Gives the .fastq files from which you will start

  - 24 different adapters allow to load 24 samples in one lane (~120-150 million base sequenced/lane), the more samples loaded the less base sequenced/sample
  - Unused adapter shouldn't be found (quality check)

- Adapter cleaning (Fastx-clipper)
  - After demultiplexing, primers and/or adapter can still be part of the reads, you have to clip them to keep only RNA sequence. In our case:
    - Multiplex PCR primer

    CATTTGTTTGGCTATTAATTGAACAAATGAGATCGGAAGAGCACACGTCT

    - Sample specific True-seq adapters

    GATCGGAAGAGCACACGTCTGAACTCCAGTCACXXXXXXATCTCGTATGCCGTCTTCTGCTTG

```
cat MyFile.fastq /software/fastx_toolkit-0.0.13/src/fastx_clipper/fastx_clipper -Q33 -a
CATTTGTTTGGCTATTAATTGAACAAATGAGATCGGAAGAGCACACGTCT -M15 -n -v -l 20 | /software/fastx_toolkit-0.0.13/src/fastx_clipper/
fastx_clipper GATCGGAAGAGCACACGTCTGAACTCCAGTCACXXXXXXATCTCGTATGCCGTCTTCTGCTTG -M15 -n -v -l 20 >MyCleanFile.fastq
```

Keep in mind that the options that we are using today fits to our data (e.g.:RNA –seq, single end sequencing...)
You should be careful to all options (it can dramatically change your results)

Note that all these steps can be done in an automatic pipeline with command lines in a shell. (Quicker for numerous and big datasets)

From raw data to pre-processed data

- Reads quality checking (FastQC)
  - A report (containing graphs representing basic statistics, sequence quality, GC content, N content, Overrepresented sequences etc...) will help to know if each step want well.
  - Example of FastQC reports:
    - Good illumina dataset: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/good_sequence_short_fastqc/fastqc_report.html
    - Poor illumina dataset: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/bad_sequence_fastqc/fastqc_report.html

```
/software/FastQC/fastqc MyCleanFile.fastq
```

- Mapping (TopHat)

  - TopHat is a fast splice junction mapper for RNA-Seq reads. It aligns RNA-Seq reads to genomes using the ultra high-throughput short read aligner Bowtie, and then analyzes the mapping results to identify splice junctions between exons.

```
/software/tophat/bin/tophat -p 20 –o MyMapped_file.tophat /bowtie_index/dmel-all-chromosome-r5.45  MyCleanFile.fastq
```

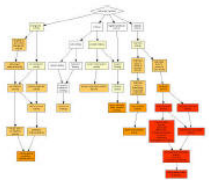
Data analysis

Getting to differential expression

- Cuffdiff
  - find significant changes in transcript/gene expression starting from BAM files and a GTF file using FPKM (To test whether an observed difference in a gene's expression is significant, Cuffdiff compares the log ratio of gene's expression in two conditions against the log of one.)

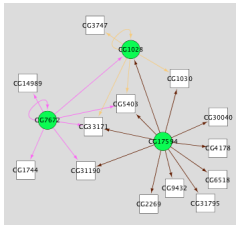
```
cuffdiff --no-update-check -q -p 4 -c 10 --FDR 0.050000 –N --labels wild_types,mutants transcript.gtf
Dmel_rep1.bam,Dmel_rep2.bam mut_rep1.bam,mut_rep2.bam
```

- DESeq, EDAseq, EdgeR
  - Other differential expression calculation tools available in galaxy and/or R



Data analysis

complementary analysis



- Gorilla - http://cbl-gorilla.cs.technion.ac.il/

  → tool for identifying and visualizing enriched GO terms in (ranked) lists of genes

  Available for Arabidopsis thaliana, Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster, Danio rerio, Homo sapiens, Mus musculus and Rattus norvegicus.

- i-regulon - cytoscape plugin

  → computational framework uniting cis-regulatory sequence analysis with network biology.

  iRegulon analyzes a *human*, *mouse*, or *fly* gene network or gene signature and identifies enriched cis-regulatory motifs.

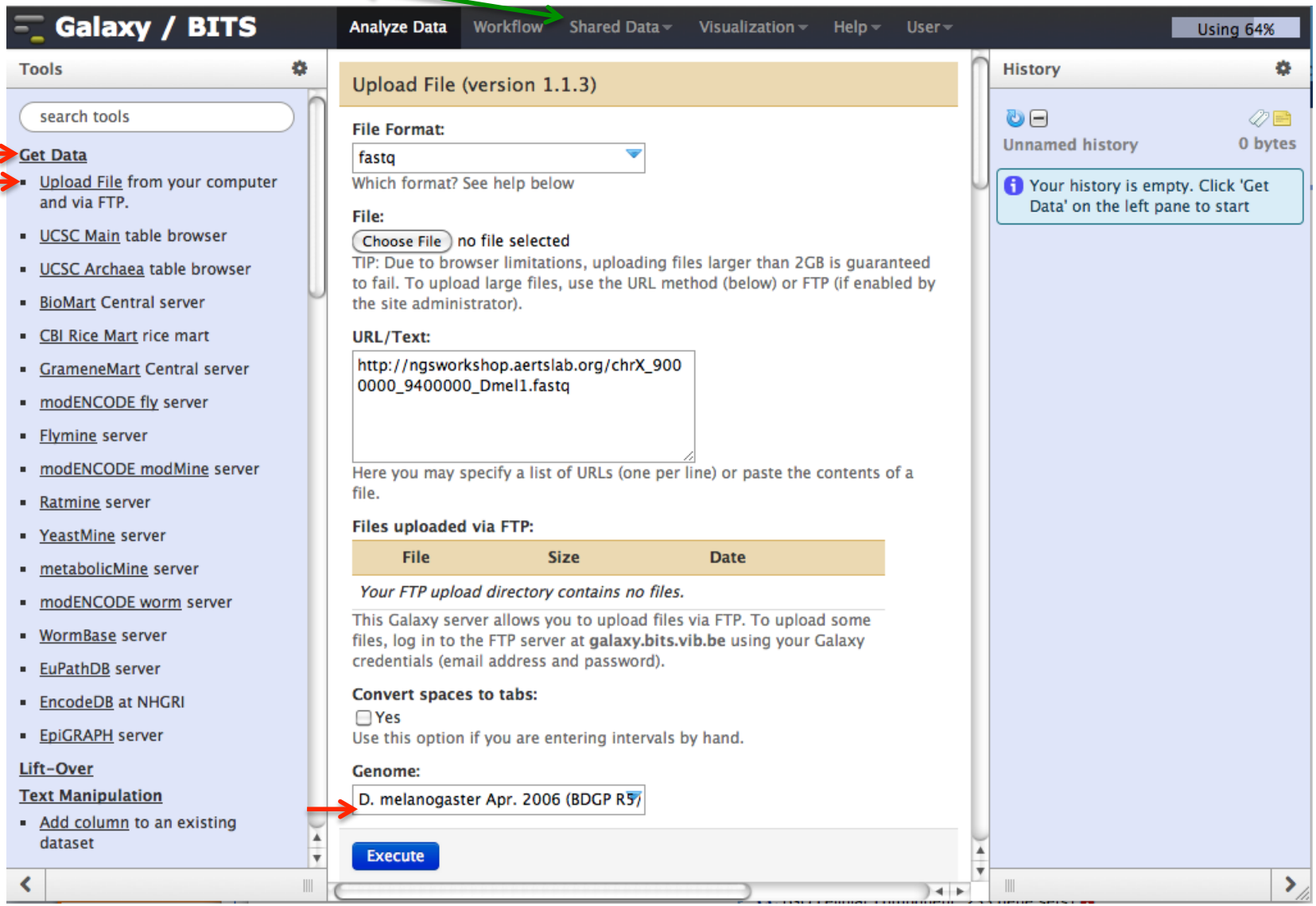
  For each candidate TF, iRegulon determines the optimal sub-network as direct targetome of the TF.

Workshop in details

1-load data

(available on http://ngsworkshop.aertslab.org/ or in the library)

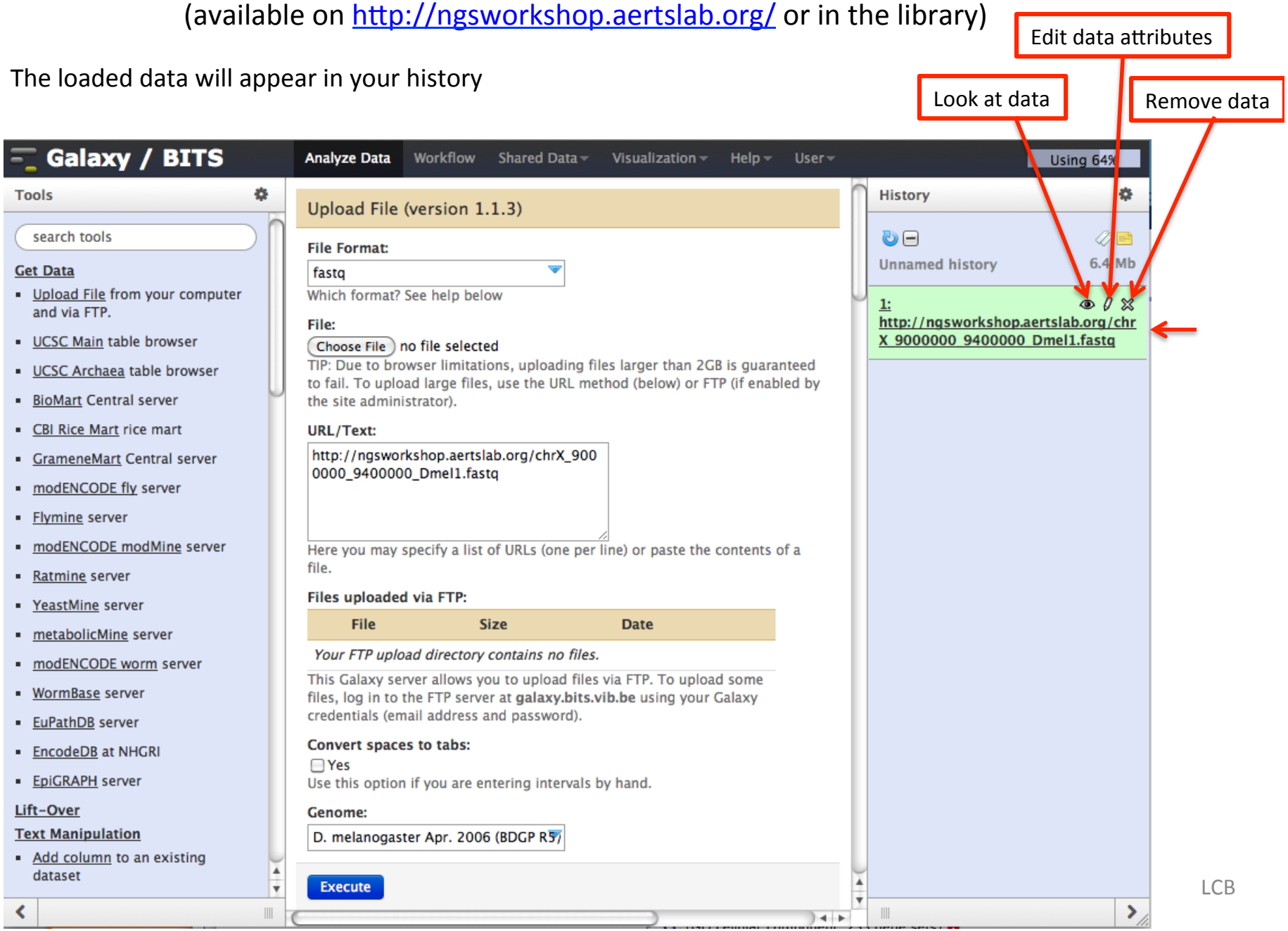
Choose your format, file to download and related genome (here we are working with RNA-seq data from D. melanogaster)
Or select datasets from “shared data → data libraries → Galaxy RNA workshop 20 sep 2012” we prepared for you



1-load data

(available on http://ngsworkshop.aertslab.org/ or in the library)

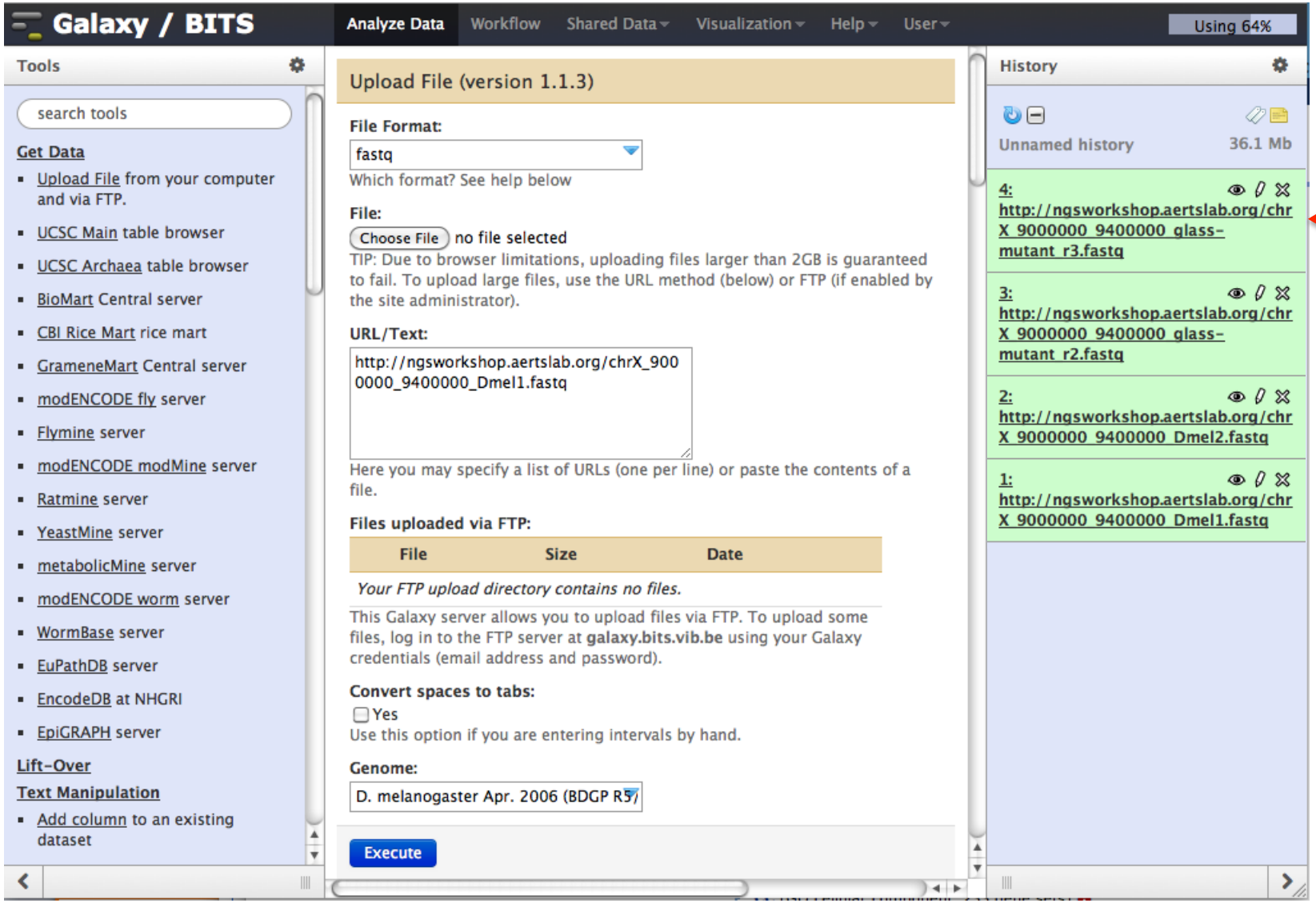
The loaded data will appear in your history



1-load data

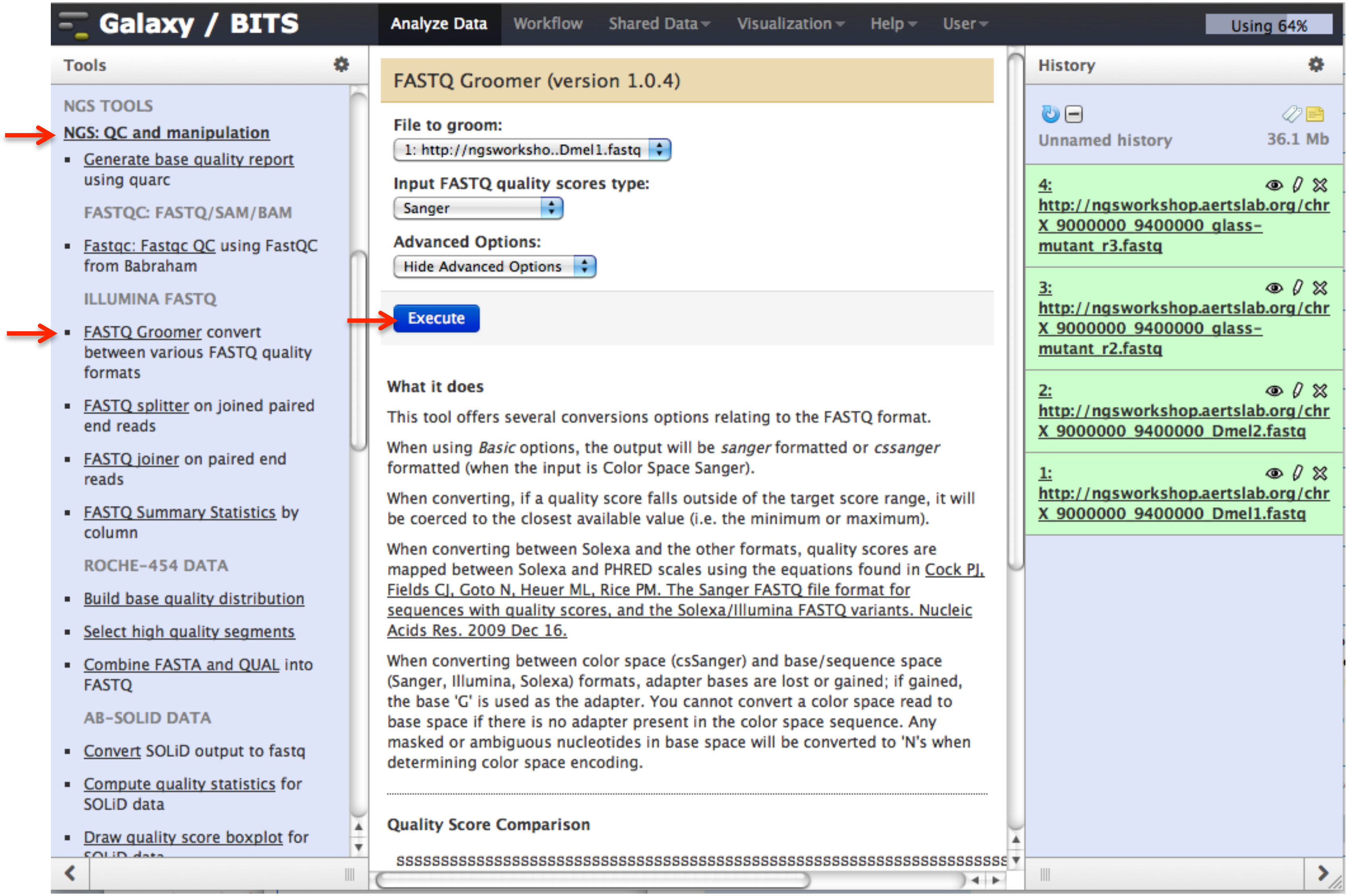
(available on http://ngsworkshop.aertslab.org/ or in the library)

The loaded data will appear in your history, you should start with 4 datasets 2 wild types and 2 mutants



2- FASTQ Groomer

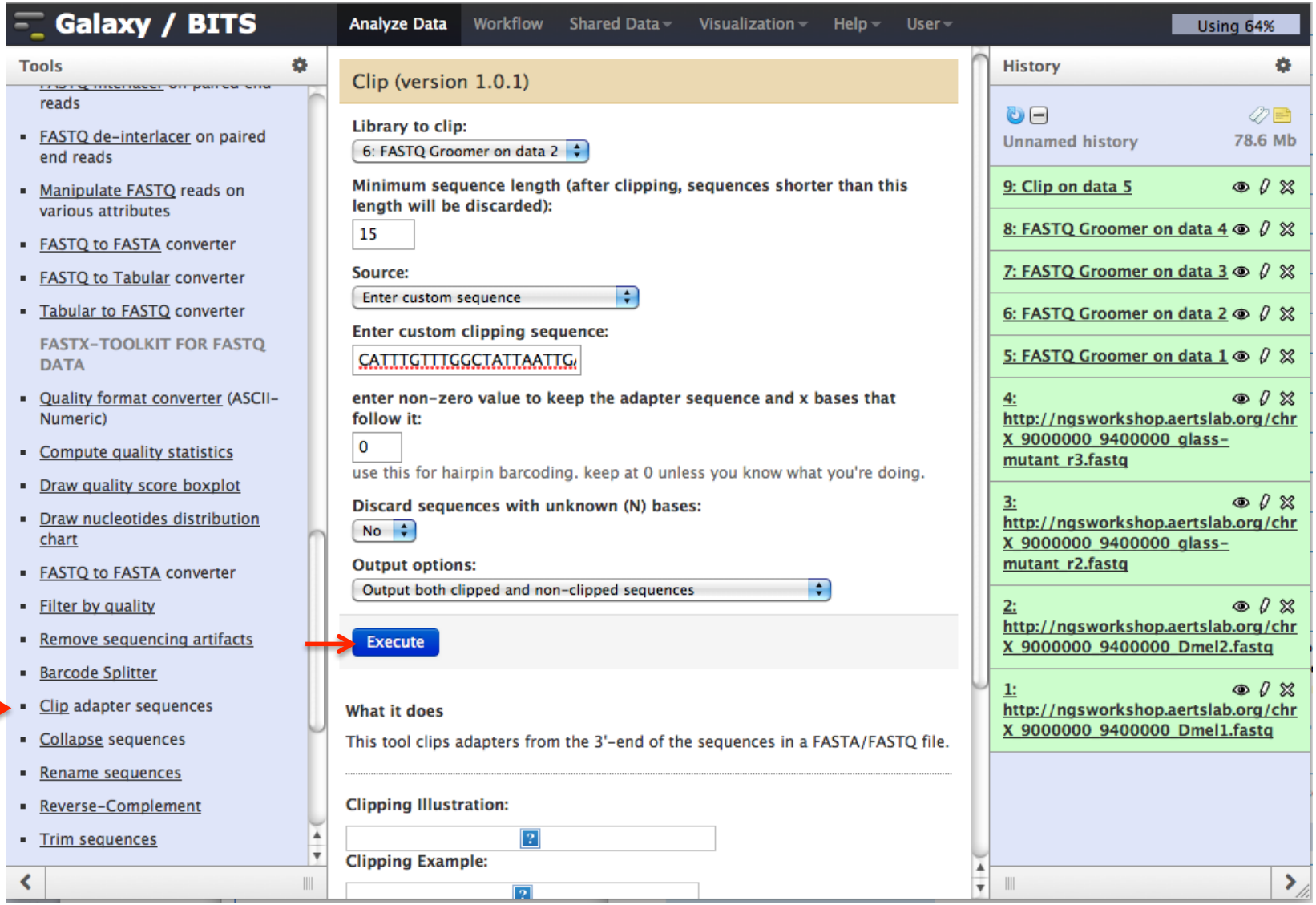
Use the FASTQ Groomer to get the data in a galaxy-tools readable format



3- Clip adapter sequences

FASTX-clipper

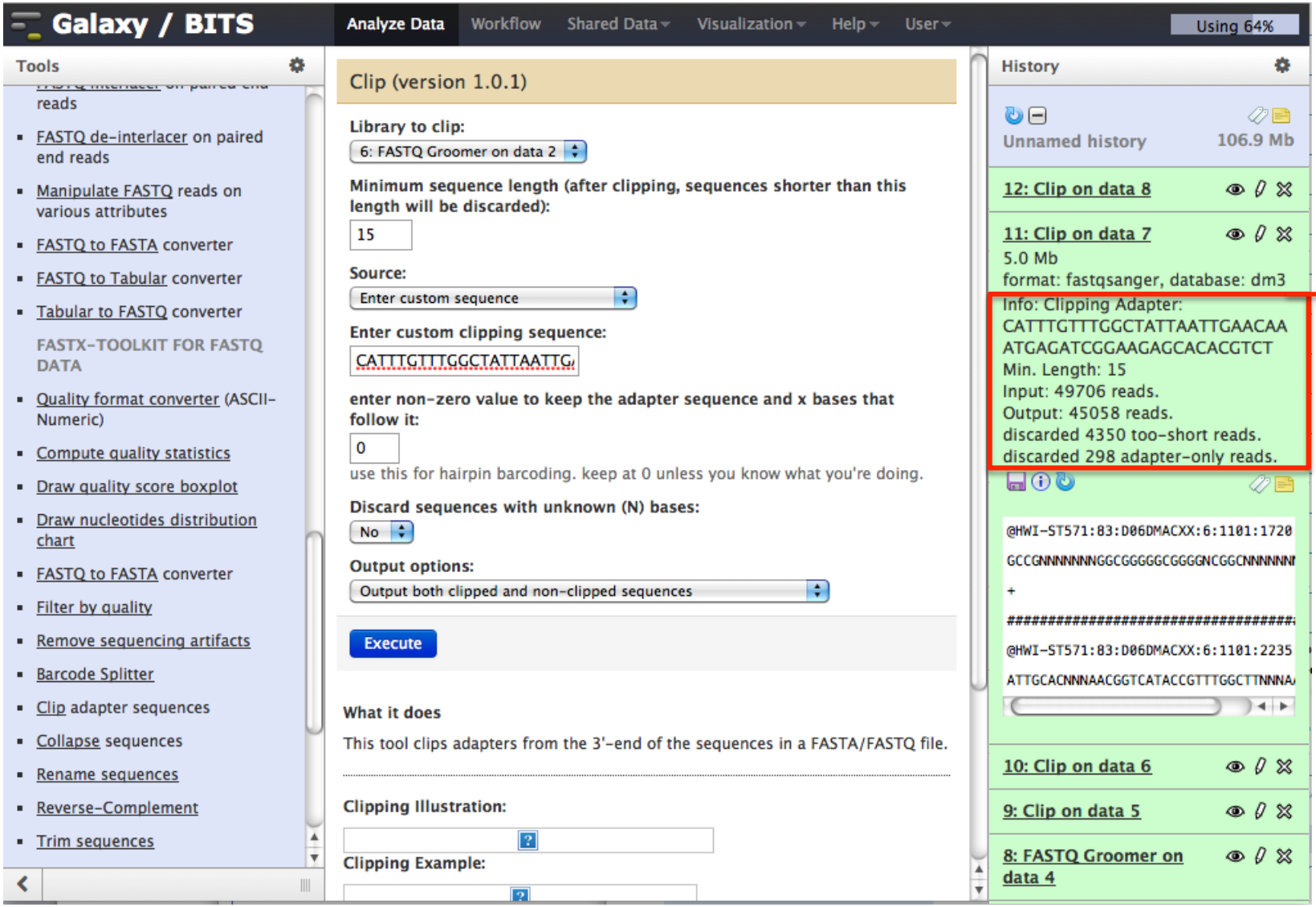
Adapters/primer can still be part of some reads, use FastX-clipper to remove them with the following options.



3- Clip adapter sequences

FASTX-clipper

Data should be clipped 2 times, first with the Multiplex PCR primer and second with the specific True-seq adapter



Infos about the number of discarded sequences

4- Check reads quality with fastQC

- After clipping you can check the quality of your reads with the fastQC quality control tool ("NGS: QC and manipulation → Fastqc: Fastqc QC")

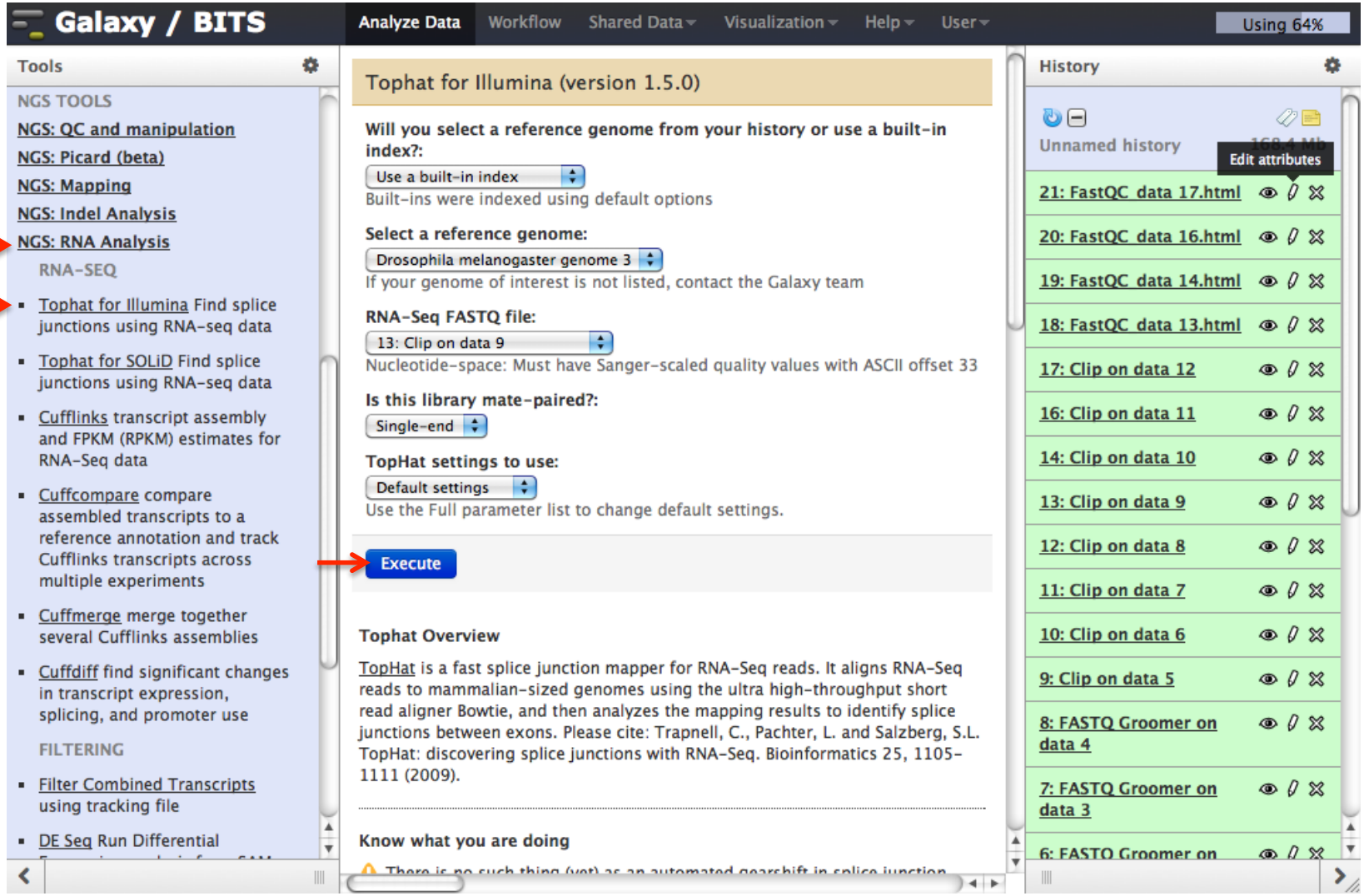
- For some exemples of good and bad results:

http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

- If you think that the mean quality of some base of your reads is too low you can trim these reads with ("NGS: QC and manipulation → FASTQ Trimmer")

5- Mapping of the reads

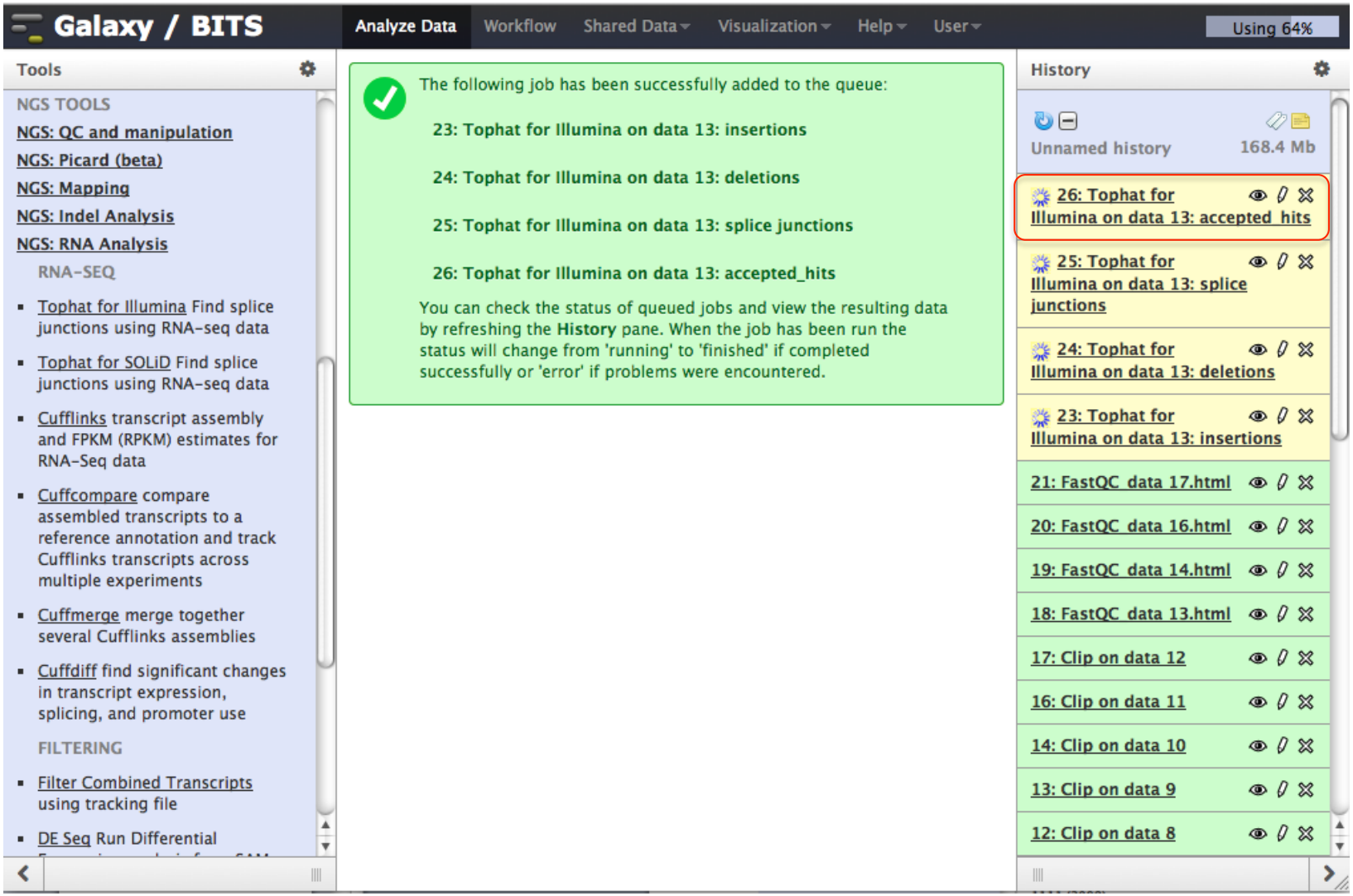
Tophat



5- Mapping of the reads

Tophat

4 files are produced by Tophat, our mapped reads are in the “accepted_hits” one



6- Get differential expression
cuffdiff

For differential expression calculation we will switch to the full dataset that can be found here:

* http://ngsworkshop.aertslab.org/ → load the 4 Tophat* files via URL box in "get data"
→ load the gtf file Flybase2006.gtf

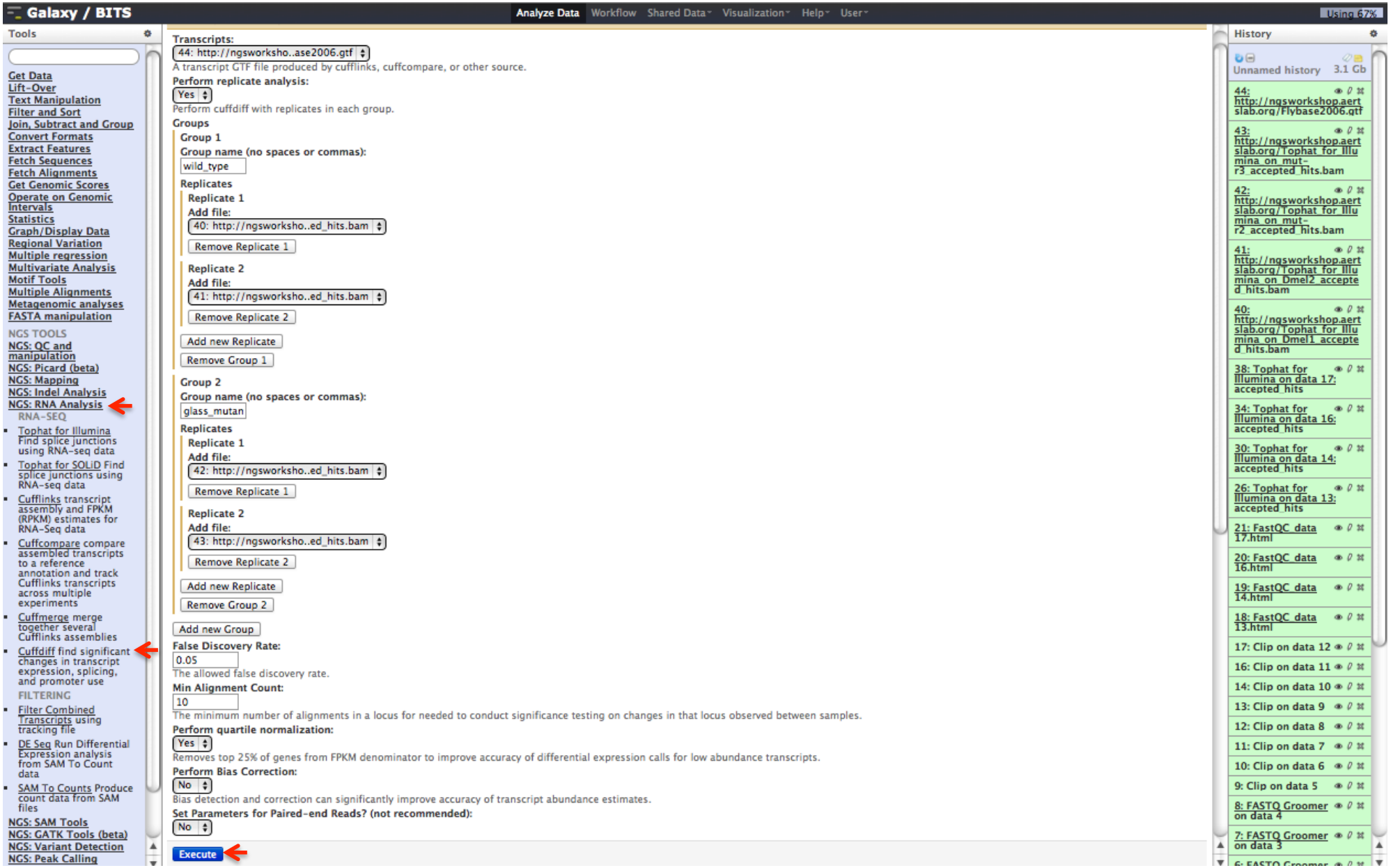
* Or in "shared data → data libraries → Galaxy RNA workshop 20 sep 2012"

Use replicates → add new group : wild_type → add new replicates : Dmel1
→ add new replicates : Dmel2
→ add new group : glass_mutant → add new replicates : mut_r2
→ add new replicates : mut_r3

Other options to change :
- Perform quartile normalization : yes

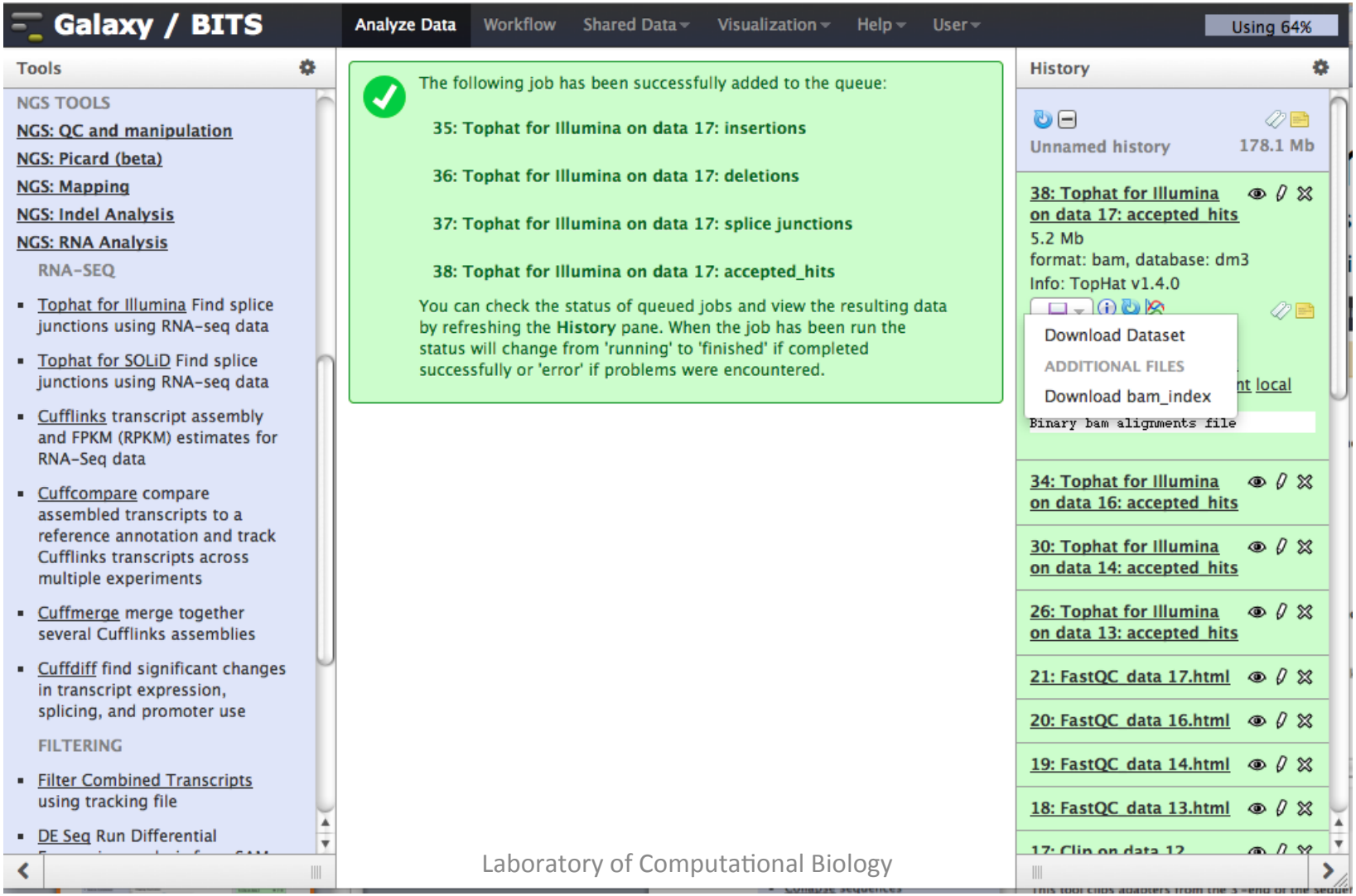
6- Get differential expression

cuffdiff (DESeq, EDAseq etc... Can also be used)



7- Visualization of your mapped reads

While cuffdiff is running you can visualize your mapped data:
Download bam and bai to have a look in local IGV, or use the visualization button (new visualisation).



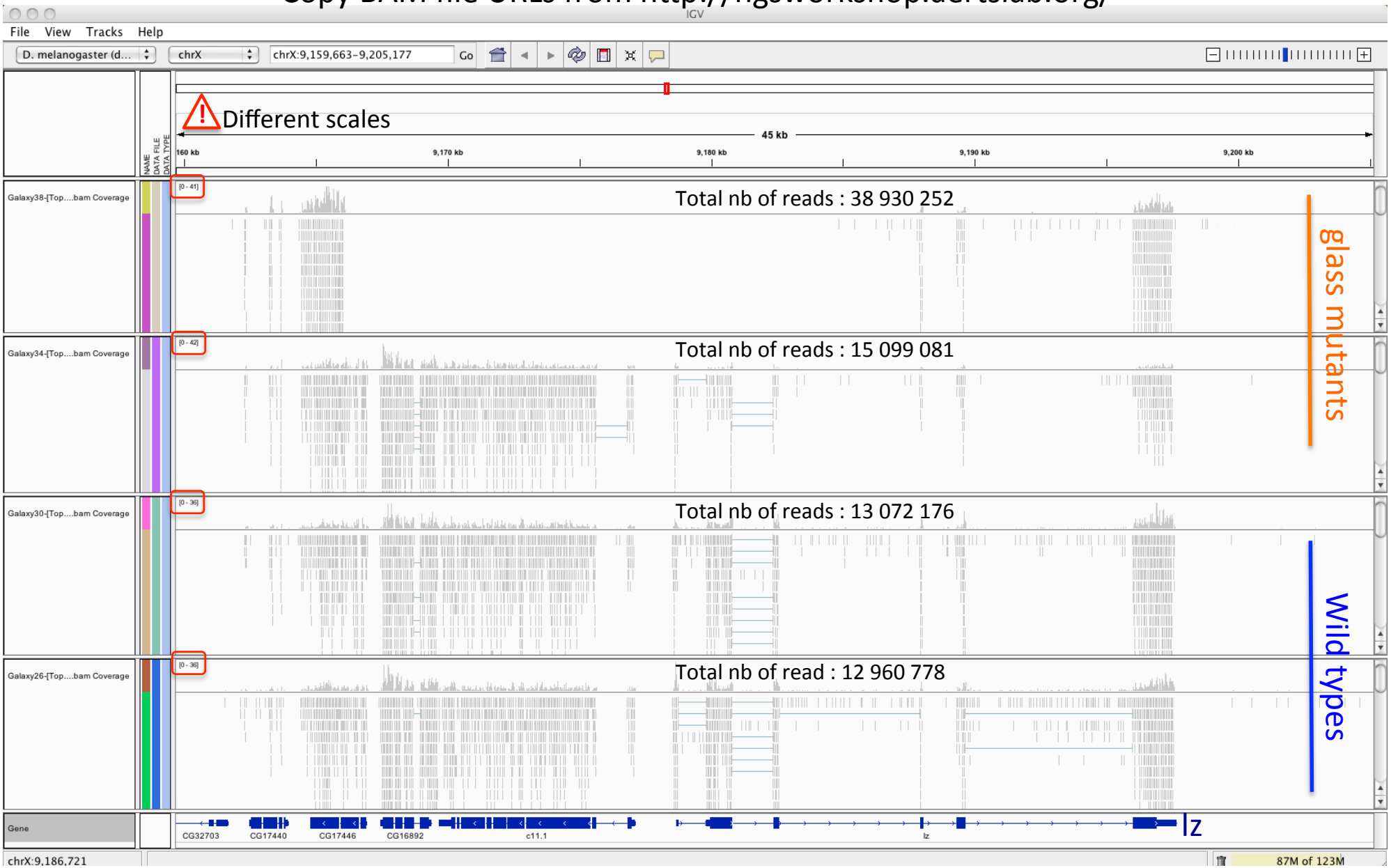
7- Visualization in galaxy

lz example (only known glass target)

7- Visualization in IGV

using the entire BAM file (File: Load from URL)

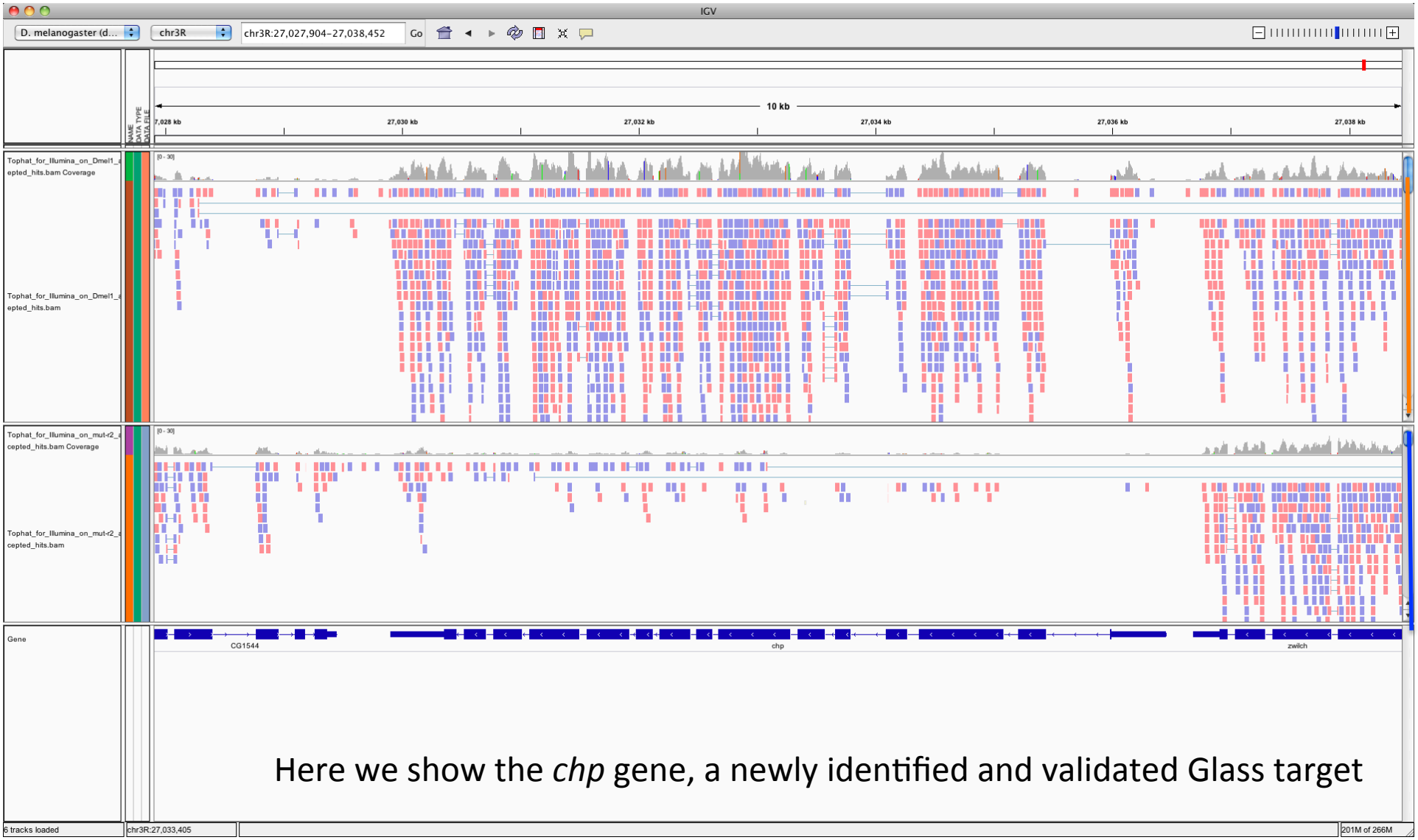
Copy BAM file URLs from http://ngsworkshop.aertslab.org/



7- Visualization in IGV

using the entire BAM file (File: Load from URL)

Copy BAM file URLs from http://ngsworkshop.aertslab.org/



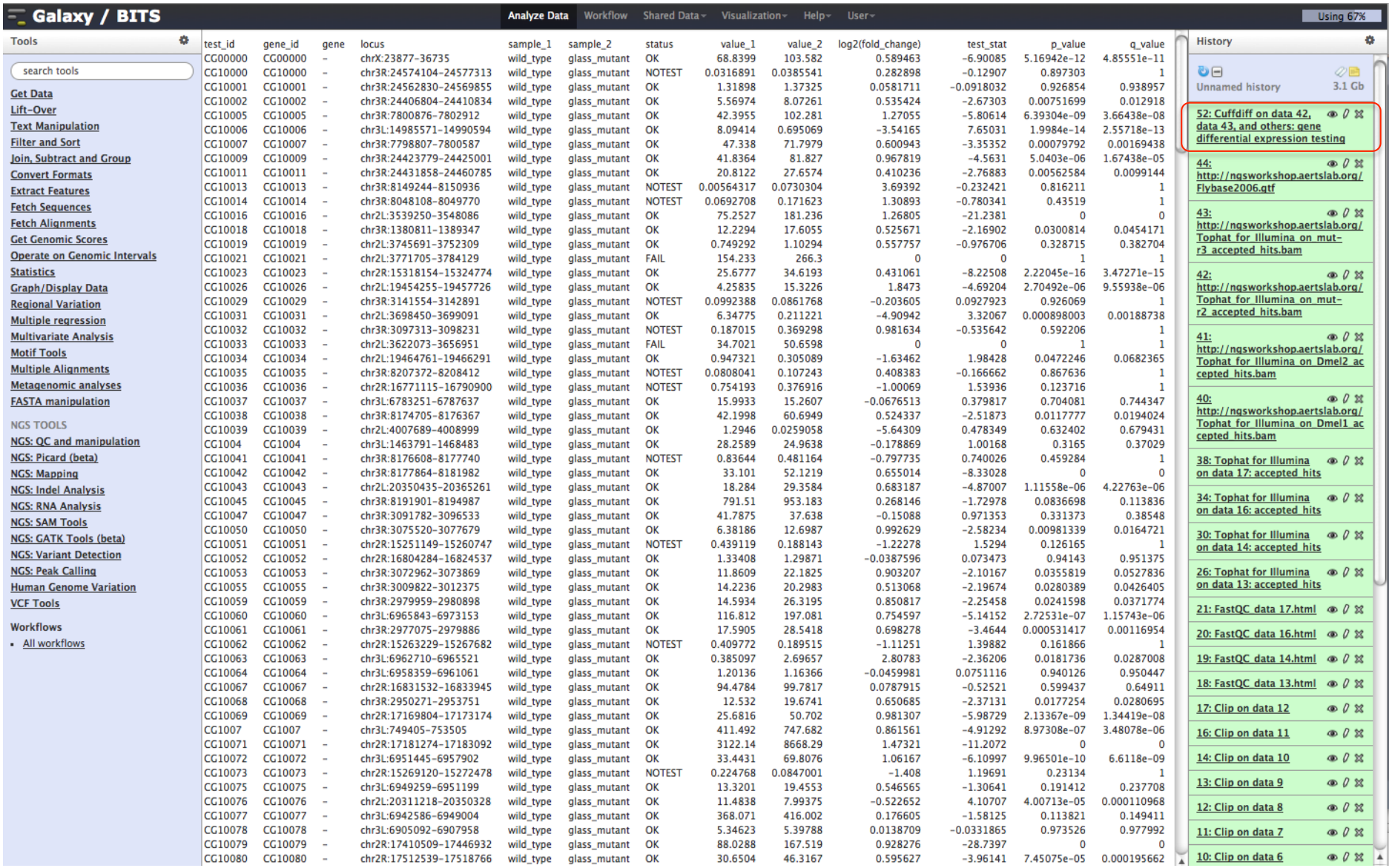
glass mutant

Wild type

Here we show the *chp* gene, a newly identified and validated Glass target

6- Cuffdiff results

Cuffdiff produce multiple files, we will now work with the gene differential expression testing



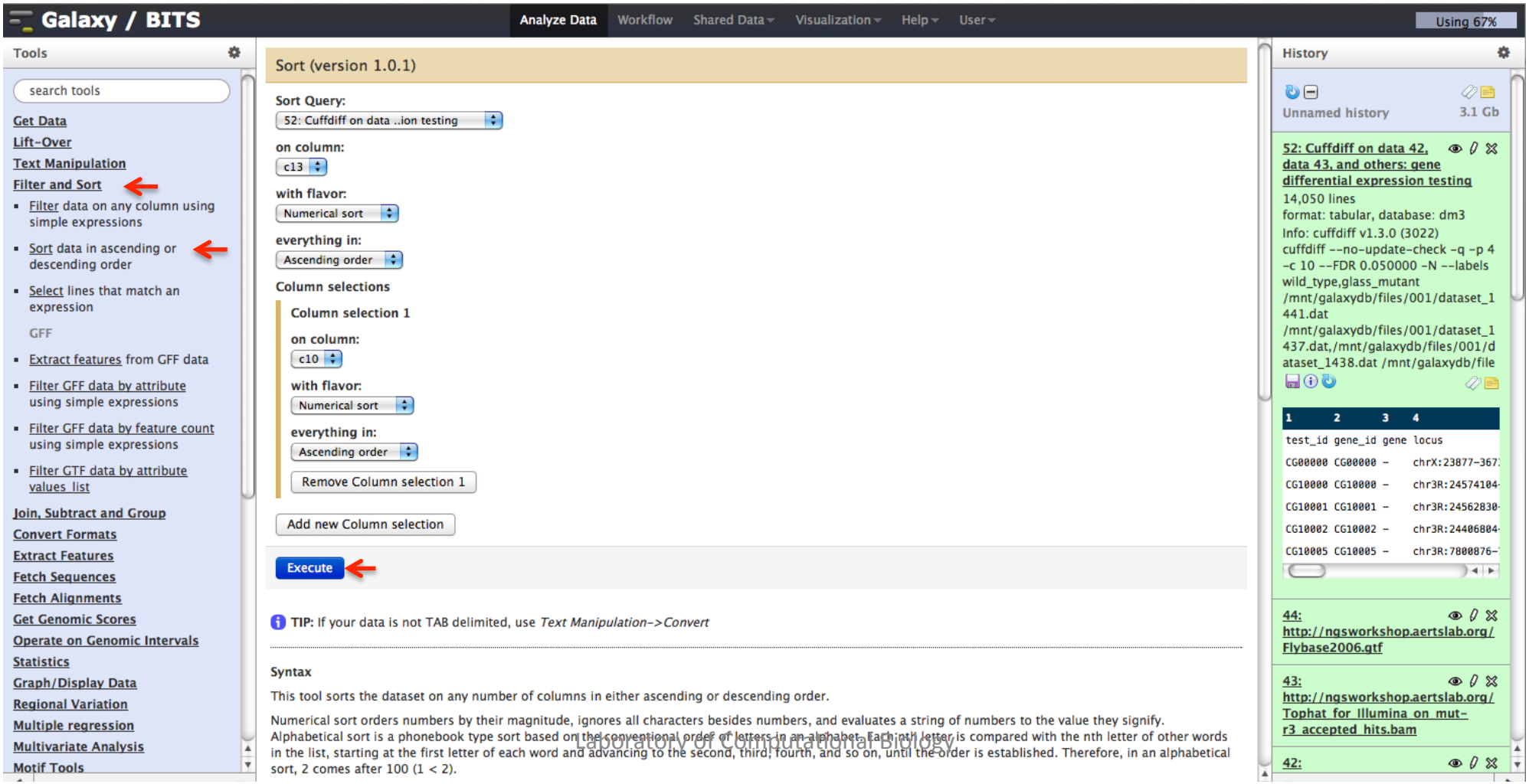
test_id	gene_id	gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value
CG00000	CG00000	-	chrX:23877-36735	wild_type	glass_mutant	OK	68.8399	103.582	0.589463	-6.90085	5.16942e-12	4.85551e-11
CG10000	CG10000	-	chr3R:24574104-24577313	wild_type	glass_mutant	NOTEST	0.0316891	0.0385541	0.282898	-0.12907	0.897303	1
CG10001	CG10001	-	chr3R:24562830-24569855	wild_type	glass_mutant	OK	1.31898	1.37325	0.0581711	-0.0918032	0.926854	0.938957
CG10002	CG10002	-	chr3R:24406804-24410834	wild_type	glass_mutant	OK	5.56974	8.07261	0.535424	-2.67303	0.00751699	0.012918
CG10005	CG10005	-	chr3R:7800876-7802912	wild_type	glass_mutant	OK	42.3955	102.281	1.27055	-5.80614	6.39304e-09	3.66438e-08
CG10006	CG10006	-	chr3L:14985571-14990594	wild_type	glass_mutant	OK	8.09414	0.695069	-3.54165	7.65031	1.9984e-14	2.55718e-13
CG10007	CG10007	-	chr3R:7798807-7800587	wild_type	glass_mutant	OK	47.338	71.7979	0.600943	-3.35352	0.00079792	0.00169438
CG10009	CG10009	-	chr3R:24423779-24425001	wild_type	glass_mutant	OK	41.8364	81.827	0.967819	-4.5631	5.0403e-06	1.67438e-05
CG10011	CG10011	-	chr3R:24431858-24460785	wild_type	glass_mutant	OK	20.8122	27.6574	0.410236	-2.76883	0.00562584	0.0099144
CG10013	CG10013	-	chr3R:8149244-8150936	wild_type	glass_mutant	NOTEST	0.00564317	0.0730304	3.69392	-0.232421	0.816211	1
CG10014	CG10014	-	chr3R:8048108-8049770	wild_type	glass_mutant	NOTEST	0.0692708	0.171623	1.30893	-0.780341	0.43519	1
CG10016	CG10016	-	chr2L:3539250-3548086	wild_type	glass_mutant	OK	75.2527	181.236	1.26805	-21.2381	0	0
CG10018	CG10018	-	chr3R:1380811-1389347	wild_type	glass_mutant	OK	12.2294	17.6055	0.525671	-2.16902	0.0300814	0.0454171
CG10019	CG10019	-	chr2L:3745691-3752309	wild_type	glass_mutant	OK	0.749292	1.10294	0.557757	-0.976706	0.328715	0.382704
CG10021	CG10021	-	chr2L:3771705-3784129	wild_type	glass_mutant	FAIL	154.233	266.3	0	0	1	1
CG10023	CG10023	-	chr2R:15318154-15324774	wild_type	glass_mutant	OK	25.6777	34.6193	0.431061	-8.22508	2.22045e-16	3.47271e-15
CG10026	CG10026	-	chr2L:19454255-19457726	wild_type	glass_mutant	OK	4.25835	15.3226	1.8473	-4.69204	2.70492e-06	9.55938e-06
CG10029	CG10029	-	chr3R:3141554-3142891	wild_type	glass_mutant	NOTEST	0.0992388	0.0861768	-0.203605	0.0927923	0.926069	1
CG10031	CG10031	-	chr2L:3698450-3699091	wild_type	glass_mutant	OK	6.34775	0.211221	-4.90942	3.32067	0.000898003	0.00188738
CG10032	CG10032	-	chr3R:3097313-3098231	wild_type	glass_mutant	NOTEST	0.187015	0.369298	0.981634	-0.535642	0.592206	1
CG10033	CG10033	-	chr2L:3622073-3656951	wild_type	glass_mutant	FAIL	34.7021	50.6598	0	0	1	1
CG10034	CG10034	-	chr2L:19464761-19466291	wild_type	glass_mutant	OK	0.947321	0.305089	-1.63462	1.98428	0.0472246	0.0682365
CG10035	CG10035	-	chr3R:8207372-8208412	wild_type	glass_mutant	NOTEST	0.0808041	0.107243	0.408383	-0.166662	0.867636	1
CG10036	CG10036	-	chr2R:16771115-16790900	wild_type	glass_mutant	NOTEST	0.754193	0.376916	-1.00069	1.53936	0.123716	1
CG10037	CG10037	-	chr3L:6783251-6787637	wild_type	glass_mutant	OK	15.9933	15.2607	-0.0676513	0.379817	0.704081	0.744347
CG10038	CG10038	-	chr3R:8174705-8176367	wild_type	glass_mutant	OK	42.1998	60.6949	0.524337	-2.51873	0.0117777	0.0194024
CG10039	CG10039	-	chr2L:4007689-4008999	wild_type	glass_mutant	OK	1.2946	0.0259058	-5.64309	0.478349	0.632402	0.679431
CG1004	CG1004	-	chr3L:1463791-1468483	wild_type	glass_mutant	OK	28.2589	24.9638	-0.178869	1.00168	0.3165	0.37029
CG10041	CG10041	-	chr3R:8176608-8177740	wild_type	glass_mutant	NOTEST	0.83644	0.481164	-0.797735	0.740026	0.459284	1
CG10042	CG10042	-	chr3R:8177864-8181982	wild_type	glass_mutant	OK	33.101	52.1219	0.655014	-8.33028	0	0
CG10043	CG10043	-	chr2L:20350435-20365261	wild_type	glass_mutant	OK	18.284	29.3584	0.683187	-4.87007	1.11558e-06	4.22763e-06
CG10045	CG10045	-	chr3R:8191901-8194987	wild_type	glass_mutant	OK	791.51	953.183	0.268146	-1.72978	0.0836698	0.113836
CG10047	CG10047	-	chr3R:3091782-3096533	wild_type	glass_mutant	OK	41.7875	37.638	-0.15088	0.971353	0.331373	0.38548
CG10050	CG10050	-	chr3R:3075520-3077679	wild_type	glass_mutant	OK	6.38186	12.6987	0.992629	-2.58234	0.00981339	0.0164721
CG10051	CG10051	-	chr2R:15251149-15260747	wild_type	glass_mutant	NOTEST	0.439119	0.188143	-1.22278	1.5294	0.126165	1
CG10052	CG10052	-	chr2R:16804284-16824537	wild_type	glass_mutant	OK	1.33408	1.29871	-0.0387596	0.073473	0.94143	0.951375
CG10053	CG10053	-	chr3R:3072962-3073869	wild_type	glass_mutant	OK	11.8609	22.1825	0.903207	-2.10167	0.0355819	0.0527836
CG10055	CG10055	-	chr3R:3009822-3012375	wild_type	glass_mutant	OK	14.2236	20.2983	0.513068	-2.19674	0.0280389	0.0426405
CG10059	CG10059	-	chr3R:2979959-2980898	wild_type	glass_mutant	OK	14.5934	26.3195	0.850817	-2.25458	0.0241598	0.0371774
CG10060	CG10060	-	chr3L:6965843-6973153	wild_type	glass_mutant	OK	116.812	197.081	0.754597	-5.14152	2.72531e-07	1.15743e-06
CG10061	CG10061	-	chr3R:2977075-2979886	wild_type	glass_mutant	OK	17.5905	28.5418	0.698278	-3.4644	0.000531417	0.00116954
CG10062	CG10062	-	chr2R:15263229-15267682	wild_type	glass_mutant	NOTEST	0.409772	0.189515	-1.11251	1.39882	0.161866	1
CG10063	CG10063	-	chr3L:6962710-6965521	wild_type	glass_mutant	OK	0.385097	2.69657	2.80783	-2.36206	0.0181736	0.0287008
CG10064	CG10064	-	chr3L:6958359-6961061	wild_type	glass_mutant	OK	1.20136	1.16366	-0.0459981	0.0751116	0.940126	0.950447
CG10067	CG10067	-	chr2R:16831532-16833945	wild_type	glass_mutant	OK	94.4784	99.7817	0.0787915	-0.52521	0.599437	0.64911
CG10068	CG10068	-	chr3R:2950271-2953751	wild_type	glass_mutant	OK	12.532	19.6741	0.650685	-2.37131	0.0177254	0.0280695
CG10069	CG10069	-	chr2R:17169804-17173174	wild_type	glass_mutant	OK	25.6816	50.702	0.981307	-5.98729	2.13367e-09	1.34419e-08
CG1007	CG1007	-	chr3L:749405-753505	wild_type	glass_mutant	OK	411.492	747.682	0.861561	-4.91292	8.97308e-07	3.48078e-06
CG10071	CG10071	-	chr2R:17181274-17183092	wild_type	glass_mutant	OK	3122.14	8668.29	1.47321	-11.2072	0	0
CG10072	CG10072	-	chr3L:6951445-6957902	wild_type	glass_mutant	OK	33.4431	69.8076	1.06167	-6.10997	9.96501e-10	6.6118e-09
CG10073	CG10073	-	chr2R:15269120-15272478	wild_type	glass_mutant	NOTEST	0.224768	0.0847001	-1.408	1.19691	0.23134	1
CG10075	CG10075	-	chr3L:6949259-6951199	wild_type	glass_mutant	OK	13.3201	19.4553	0.546565	-1.30641	0.191412	0.237708
CG10076	CG10076	-	chr2L:20311218-20350328	wild_type	glass_mutant	OK	11.4838	7.99375	-0.522652	4.10707	4.00713e-05	0.000110968
CG10077	CG10077	-	chr3L:6942586-6949004	wild_type	glass_mutant	OK	368.071	416.002	0.176605	-1.58125	0.113821	0.149411
CG10078	CG10078	-	chr3L:6905092-6907958	wild_type	glass_mutant	OK	5.34623	5.39788	0.0138709	-0.0331865	0.973526	0.977992
CG10079	CG10079	-	chr2R:17410509-17446932	wild_type	glass_mutant	OK	88.0288	167.519	0.928276	-28.7397	0	0
CG10080	CG10080	-	chr2R:17512539-17518766	wild_type	glass_mutant	OK	30.6504	46.3167	0.595627	-3.96141	7.45075e-05	0.000195662

8- GO analysis

rank the cuffdiff result file

Rank the result file with “Filter and Sort → sort” according to:

- the best Q-value
- For an equal Q-value rank according to the logFC (glass being an activator, in the ascending order)



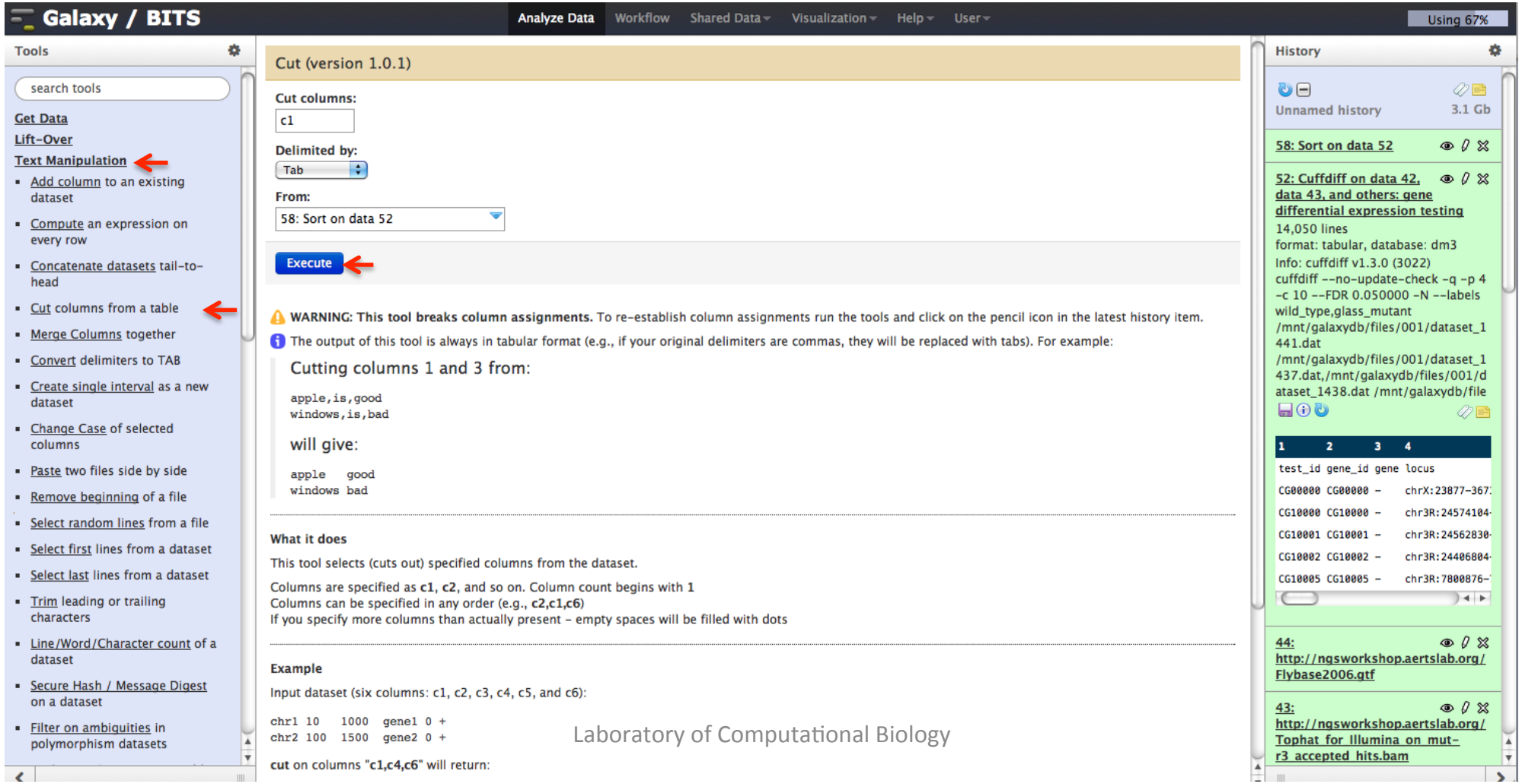
8- GO analysis

rank the cuffdiff result file

Rank the result file with “Filter and Sort → sort” according to:

- the best Q-value
- For an equal Q-value rank according to the logFC (glass being an activator, in the ascending order)

Keep only the first column with “Text manipulation → cut”



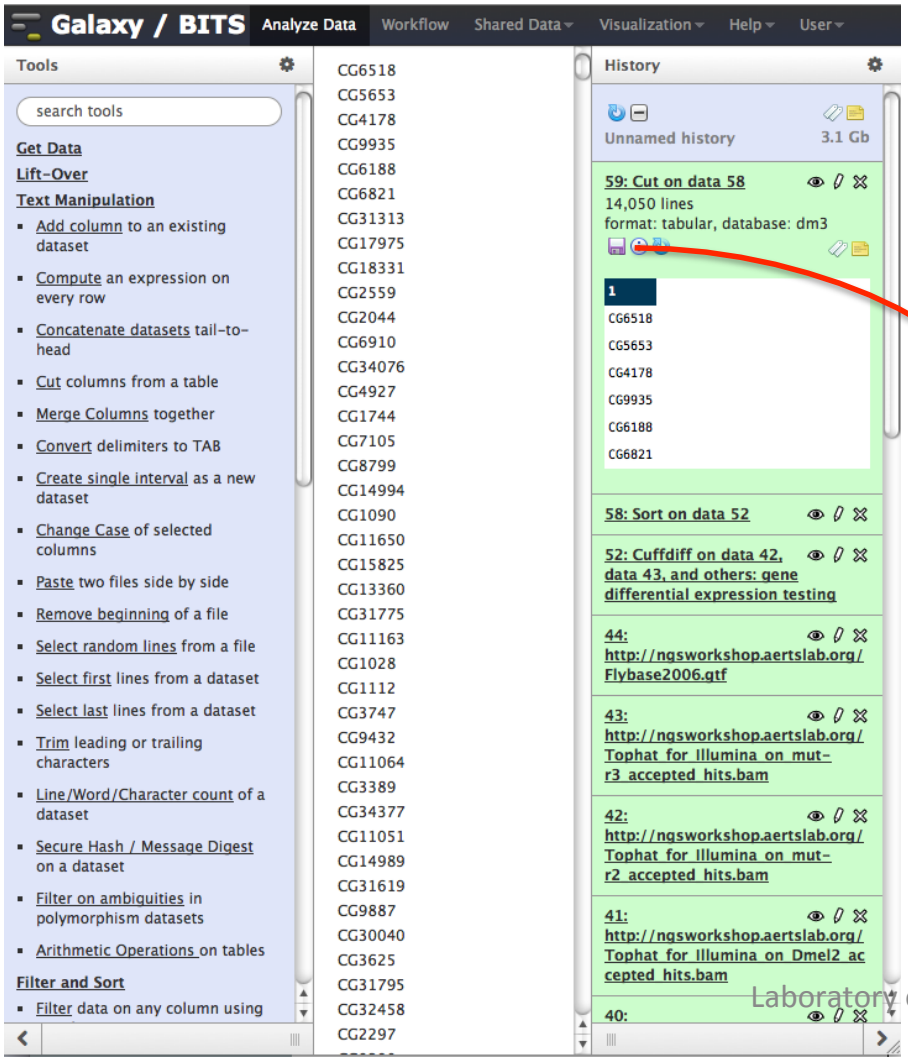
8- GO analysis

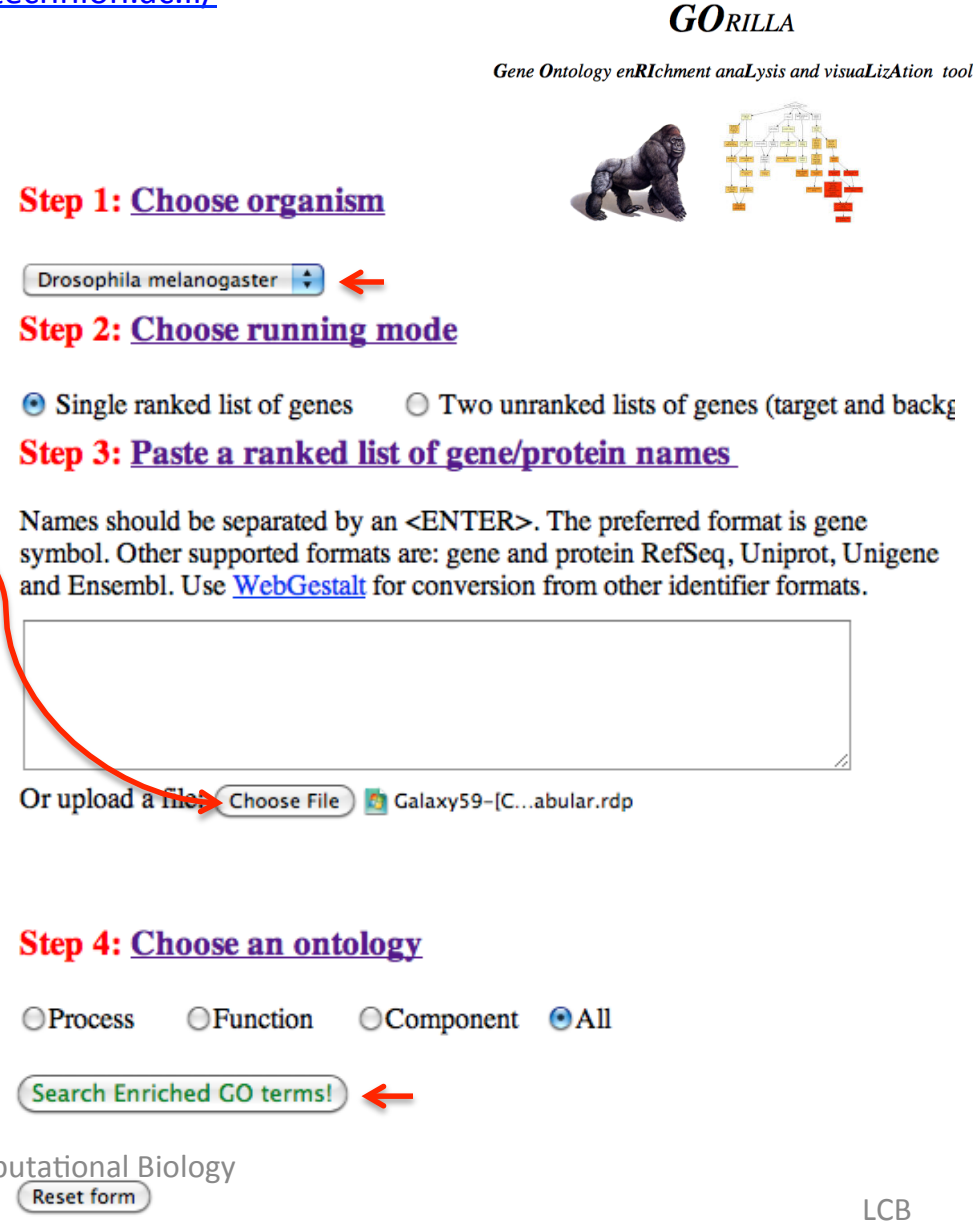
GOrilla

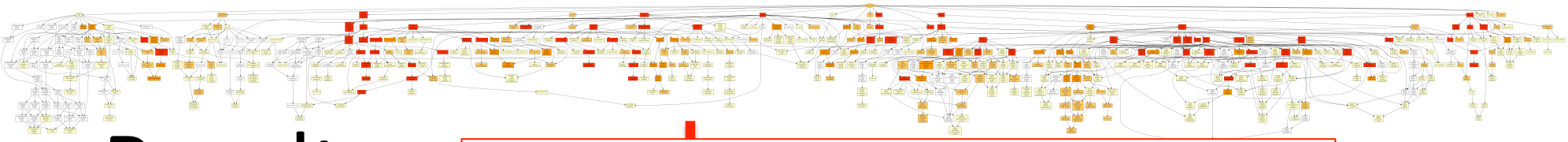
Save your ranked file to use it in Gorilla: http://cbl-gorilla.cs.technion.ac.il/

Option: - Single ranked list of genes

- All ontologies

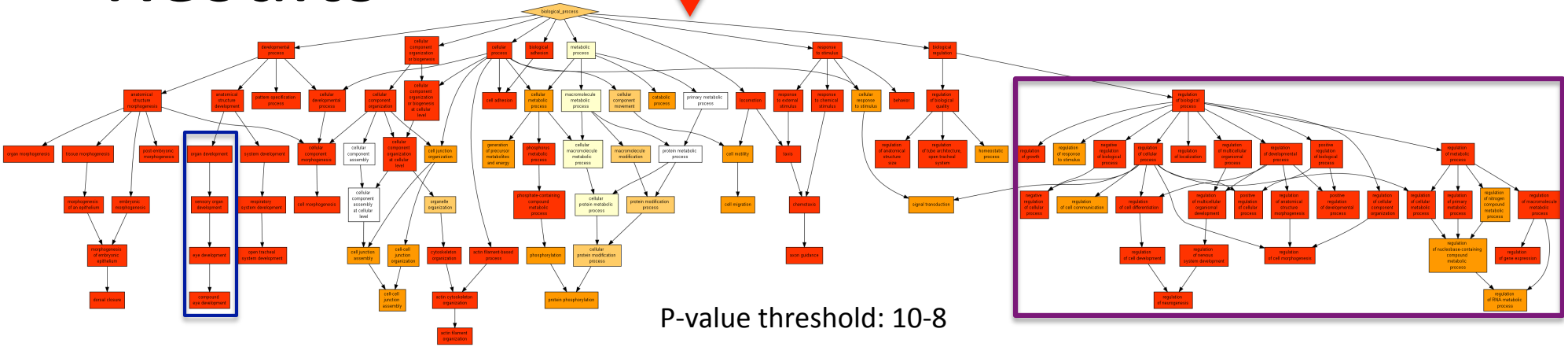




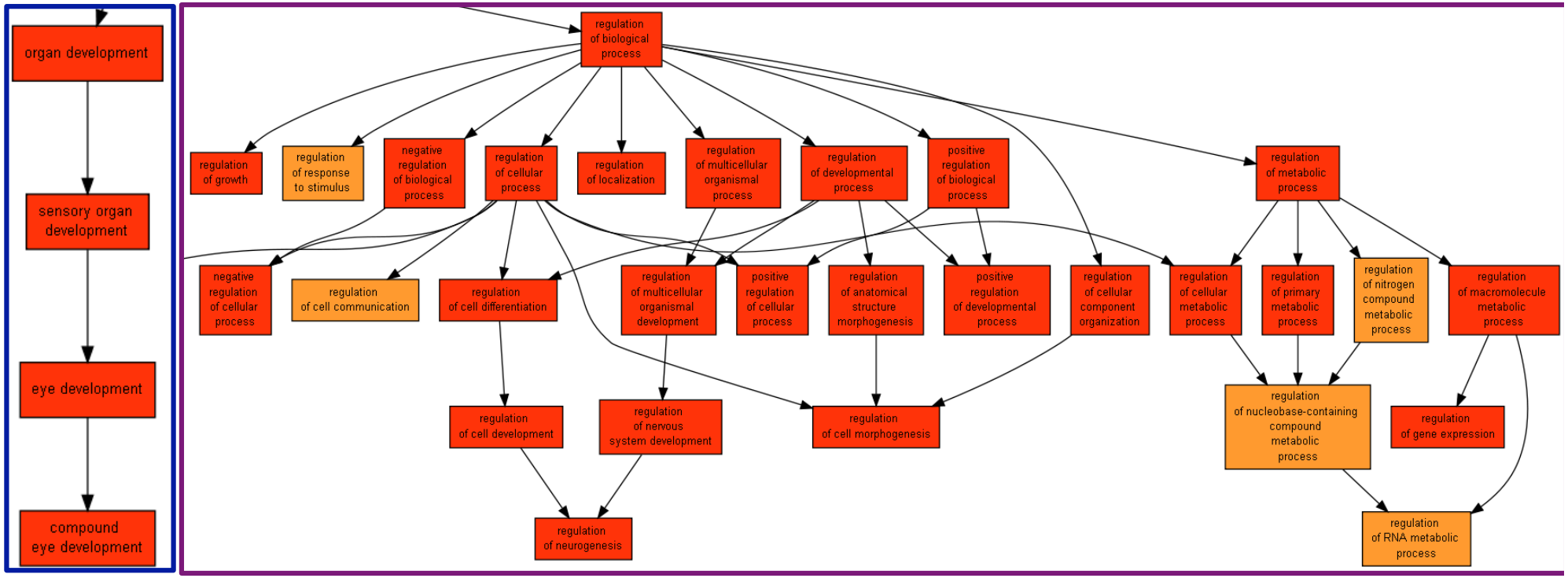


Results

To keep only the most significant results, a filter on the P-value can be applied



P-value threshold: 10-8



8- GO analysis

GOrilla

Selected parts of the Biological Process Gorilla result table:

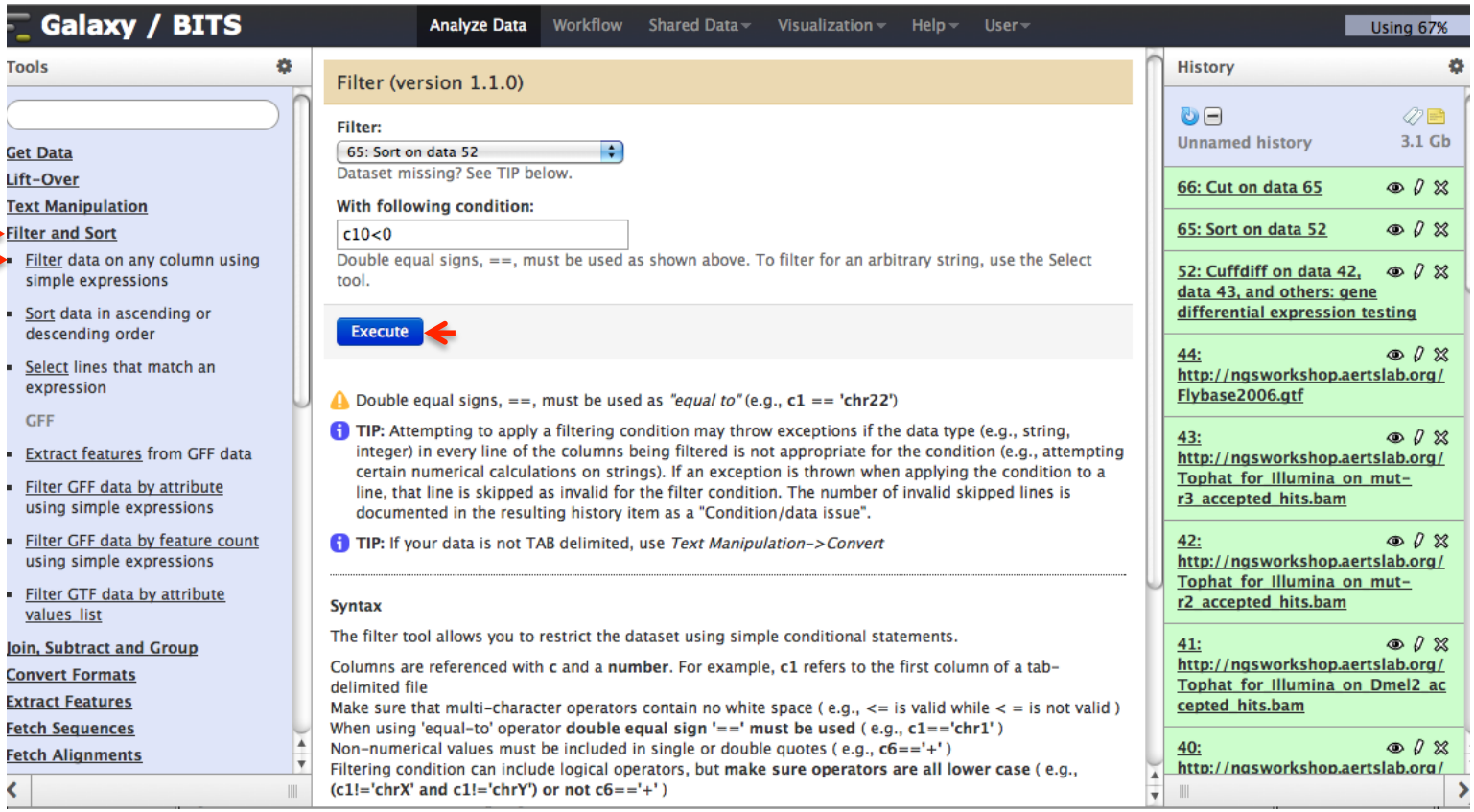
GO:0048729	tissue morphogenesis	2.78E-15	8.09E-13	3.67 (10774,178,875,53)
GO:0040008	regulation of growth	9.09E-14	1.73E-11	5.18 (10774,181,402,35)
GO:0060284	regulation of cell development	1.07E-13	1.97E-11	4.57 (10774,191,444,36)
GO:0051960	regulation of nervous system development	3.1E-13	4.8E-11	5.10 (10774,151,434,31)
GO:0007423	sensory organ development	1.37E-12	1.62E-10	2.74 (10774,217,1087,60)
GO:0007411	axon guidance	2.84E-11	2.7E-9	2.61 (10774,183,1288,57)
GO:0048749	compound eye development	2.05E-10	1.61E-8	4.02 (10774,111,748,31)
GO:0090066	regulation of anatomical structure size	2.93E-10	2.2E-8	3.25 (10774,95,1291,37)

- 10774 annotated with a Biological Process
- 111 genes annotated “compound eye development”
- 31 genes are annotated “compound eye development” in the first 748 genes of our ranking

9- Find glass direct target

i-regulon analysis

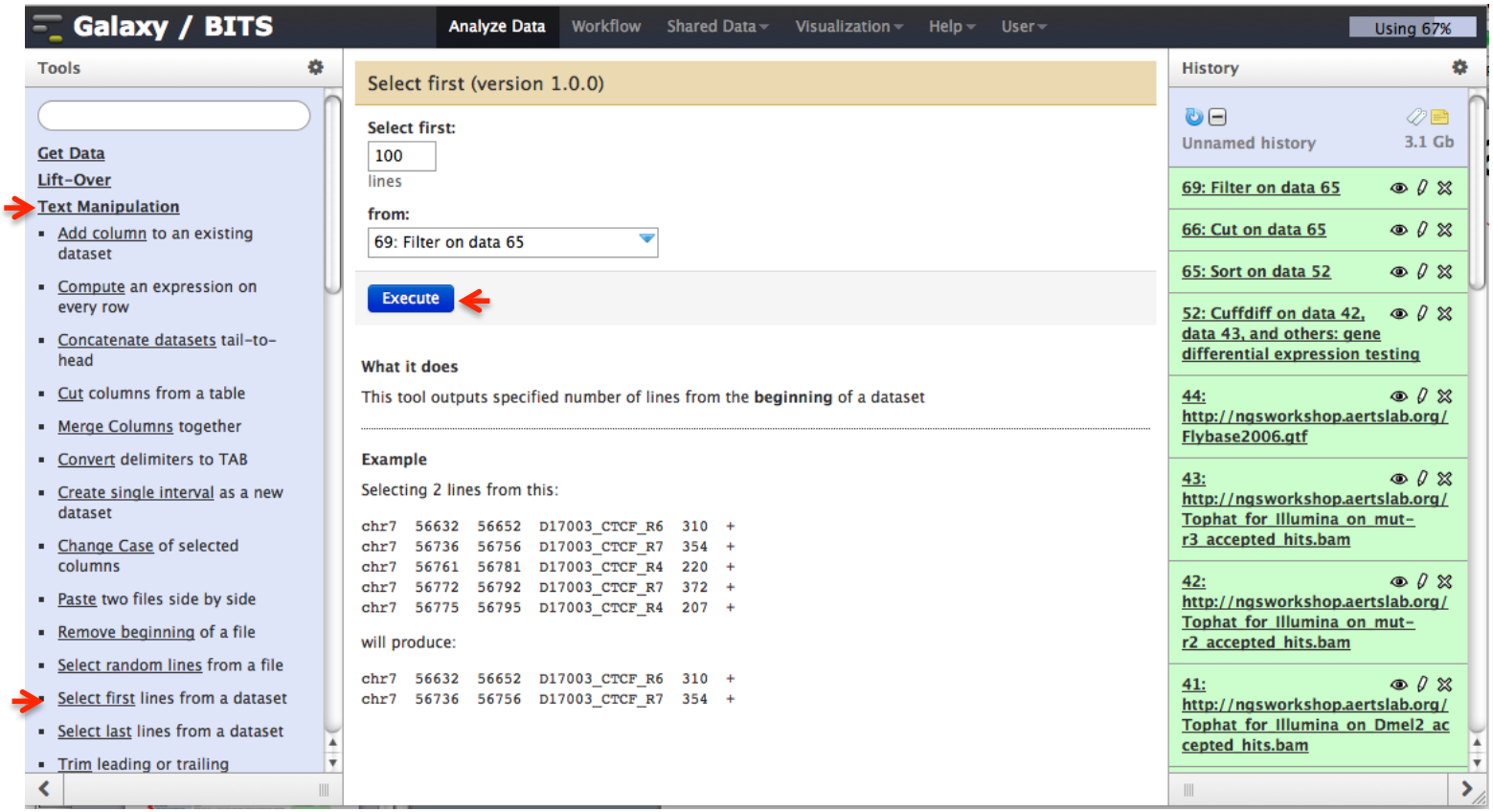
Glass being an activator, select only the down-regulated genes with a filter on the sorted data (c10<0).



9- Find glass direct target

i-regulon analysis

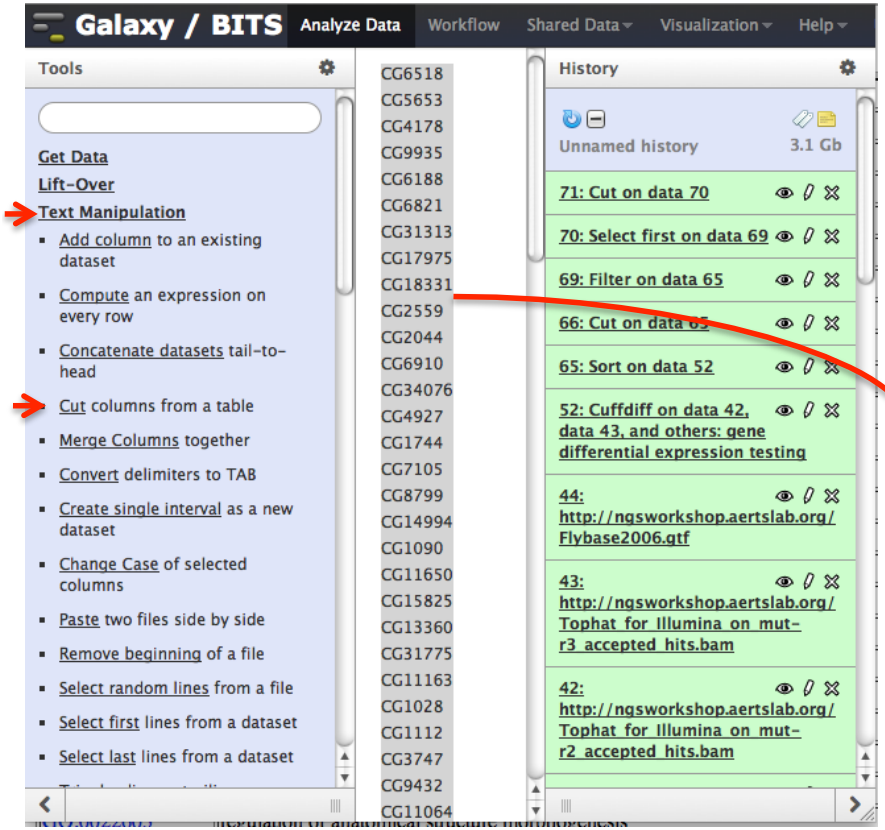
Select the top 100 down regulated genes



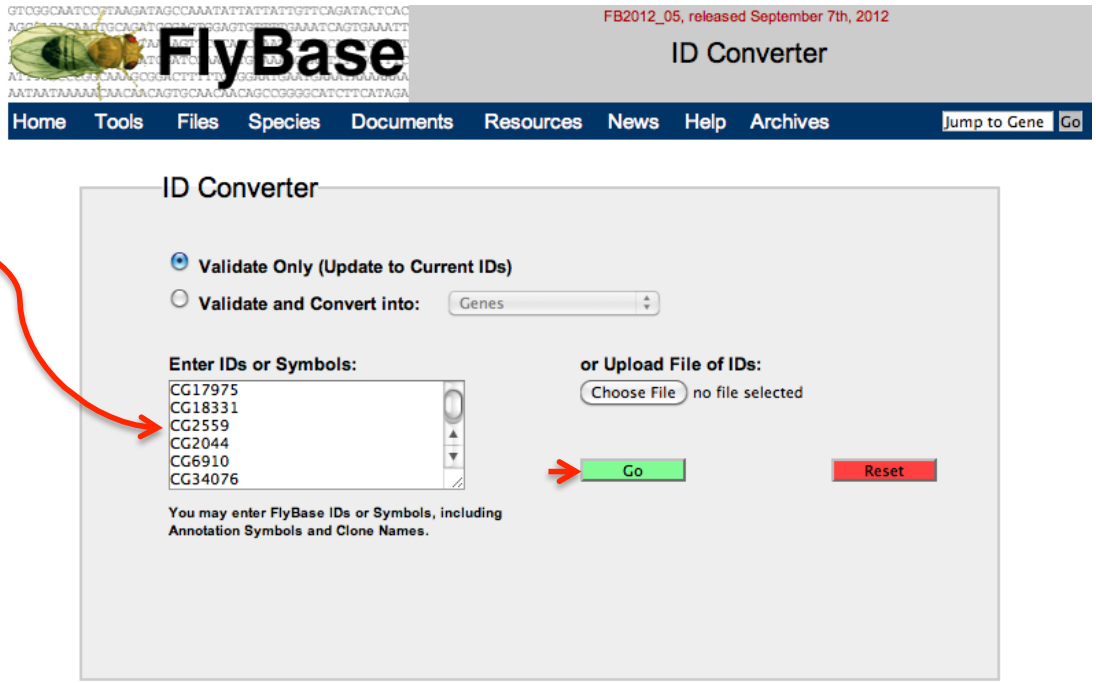
9- Find glass direct target

i-regulon analysis

Cut the 1st column and convert CG identifier to symbol in flybase



http://flybase.org/static_pages/downloads/IDConv.html



9- Select genes for an i-regulon analysis

1-click on Flybase HitList

Export converted IDs to: FlyBase HitList | file, uniq IDs only | file, conversion table

Conversion report			
Submitted ID	Current ID	Converted ID	Related record
CG6518	FBgn0004784	FBgn0004784	inaC
CG5653	FBgn0035943	FBgn0035943	CG5653
CG4178	FBgn0002563	FBgn0002563	Lsp1beta
CG9935	FBgn0039916	FBgn0039916	CG9935
CG6188	FBgn0038074	FBgn0038074	CG6188
CG6821	FBgn0002564	FBgn0002564	Lsp1gamma
CG31313	FBgn0051313	FBgn0051313	CG31313
CG17975	FBgn0028562	FBgn0028562	sut2
CG18331	FBgn0036181	FBgn0036181	Muc68Ca
CG2559	FBgn0002562	FBgn0002562	Lsp1alpha
CG2044	FBgn0002535	FBgn0002535	Lcp4

2-Removed other species corresponding genes
Click on “HitList Conversion Tools → Batch download”

Batch Download

Output Format

- FASTA Sequence — Gene region
- Database Format — Full Data Only — Chado XML
- Field Data — Selected Fields Only — As tab-separated file

Output Options

Send results to: File

You may enter FlyBase IDs or Symbols, including Annotation Symbols and Clone Names.

Enter IDs, Symbols or Sequence Coordinates:

FBgn0012034
FBgn0260642
FBgn0263111
FBgn0038247
FBgn0037238
FBgn0033484

or Upload File of IDs: Choose File — no file selected

Allow synonyms

Select fields

General Information | Check Section | Uncheck Section

- Symbol
- Name
- Feature Type
- Gene Model Status

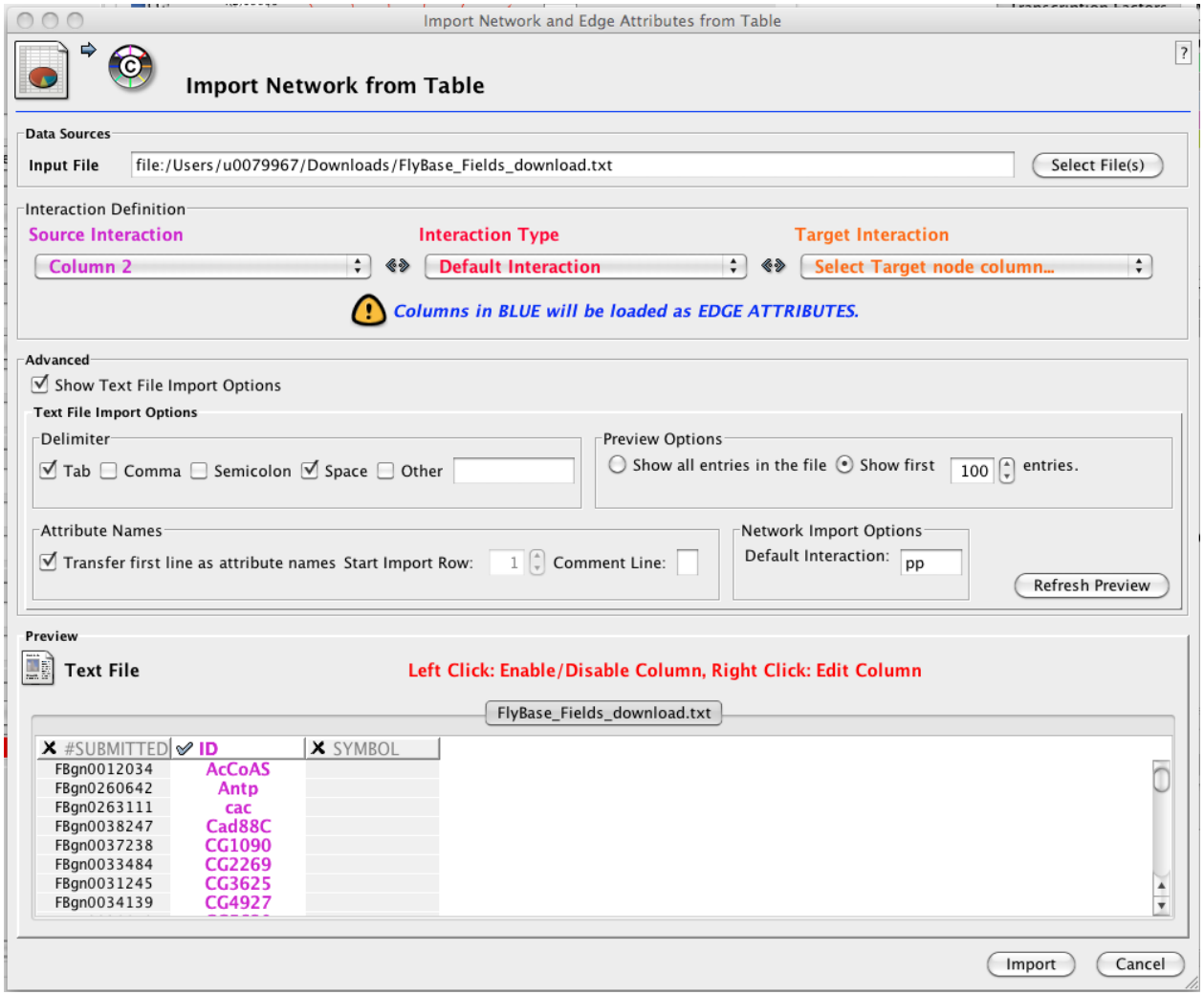
3- Get field data

61 matches — Show related: Genes — Results Analysis/Refinement — HitList Conversion Tools

#	Symbol	Name	Annotation ID	Cytology	Alleles #	Stocks #	Clones #
1	AcCoAS	Acetyl Coenzyme A synthase	CG9390	78C3-78C4	12	10	150
2	Antp	Antennapedia	CG1028	84A6-84B2	174	106	39
3	cac	cacophony	CG43368	10F7-11A1	57	13	65
4	Cad88C	Cadherin 88C	CG3389	88C10-88C10	9	5	3
5	CG1090	-	CG1090	82A4-82A4	12	10	47
6	CG2269	-	CG2269	46E4-46E4	5	2	73
7	CG3625	-	CG3625	21B7-21B7	11	7	49
8	CG4927	-	CG4927	53C8-53C8	3	3	9
9	CG5630	-	CG5630	93A3-93A4	2	3	36
10	CG5653	-	CG5653	66E5-66E5	2	3	3
11	CG6188	-	CG6188	87C5-87C5	4	3	76
12	CG6910	-	CG6910	69A1-69A1	4	5	59
13	CG9935	-	CG9935	102D1-102D1	6	4	28
14	CG11163	-	CG11163	41F10-41F11	16	13	40
15	CG13360	-	CG13360	1C4-1C4	2	2	15
16	CG14989	-	CG14989	64A5-64A5	3	3	141
17	CG31313	-	CG31313	88C6-88C6	4	2	15
18	CG31619	-	CG31619	39F1-39F3	15	12	74
19	CG31775	-	CG31775	35B5-35B5	-	-	105
20	CG32082	-	CG32082	68A8-68A9	6	3	35
21	CG34377	-	CG34377	94A5-94A5	14	13	1
22	CG42351	-	CG42351	55F1-55F1	9	7	35
23	chp	chaoptic	CG1744	100B5-100B6	16	5	78
24	Dmir\CG2269	-	-	-	-	-	-
25	Dmir\CG4821	-	-	-	-	-	-
26	Dscam3	Down syndrome cell adhesion molecule 3	CG31190	90B1-90B1	12	8	20
27	Dsim\AcCoAS	Acetyl Coenzyme A synthase	-	-	-	-	-
28	Dyak\GE16568	-	GE16568	-	-	-	-
29	Dyak\GE21891	-	GE21891	-	-	-	-
30	Dyak\pst	-	GE20418	-	-	-	-
31	Eaat1	Excitatory amino acid transporter 1	CG3747	30A8-30A8	15	4	236
32	fon	fondue	CG15825	37D3-37D3	5	1	161
33	futsch	futsch	CG34387	2A3-2A3	22	7	36
34	Gad1	Glutamic acid decarboxylase 1	CG14994	64A5-64A5	20	11	100
35	IA-2	IA-2 ortholog	CG31795	21E3-21E3	9	7	70
36	inaC	inactivation no afterpotential C	CG6518	53E1-53E1	21	11	53
37	jeb	jelly belly	CG30040	48E1-48E2	13	8	17
38	l(2)01289	lethal (2) 01289	CG9432	42C6-42C7	23	9	138
39	l(2)03659	lethal (2) 03659	CG8799	45D1-45D1	6	6	3
40	Lcp1	Larval cuticle protein 1	CG11650	44C6-44C6	4	3	30
41	Lcp4	Larval cuticle protein 4	CG2044	44D1-44D1	5	5	22
42	Lsp1α	Larval serum protein 1 α	CG2559	11A12-11A12	7	3	36
43	Lsp1β	Larval serum protein 1 β	CG4178	21E2-21E2	8	5	155
44	Lsp1γ	Larval serum protein 1 γ	CG6821	61A6-61A6	8	2	277
45	lz	lozenge	CG1689	8D5-8D6	160	62	5
46	mp	multiplexin	CG42543	65E1-65E2	18	14	48
47	mt:ND3	mitochondrial NADH-ubiquinone oxidoreductase chain 3	CG34076	-	1	-	-
48	Muc68Ca	Mucin 68Ca	CG18331	68C15-68C15	1	2	1
49	Nplp2	Neuropeptide-like precursor 2	CG11051	70A6-70A6	2	3	322
50	nrm	neuromusculin	CG43079	80C3-80C3	21	17	120
51	Obp44a	Odorant-binding protein 44a	CG2297	44B3-44B3	3	4	103
52	Proct	Proctolin	CG7105	28D2-28D2	5	5	211
53	pst	pastrel	CG8588	65F6-65F7	7	3	276
54	retn	retained	CG5403	59F5-59F5	29	8	102
55	Rfabg	Retinoid- and fatty acid-binding glycoprotein	CG11064	102F8-102F8	16	8	354
56	Scr	Sex combs reduced	CG1030	84A5-84A5	116	62	15
57	scro	scarecrow	CG17594	-	3	3	6

i-regulon analysis

Use the file from Flybase to run an i-regulon analysis



- Run cytoscape
- File

  →import

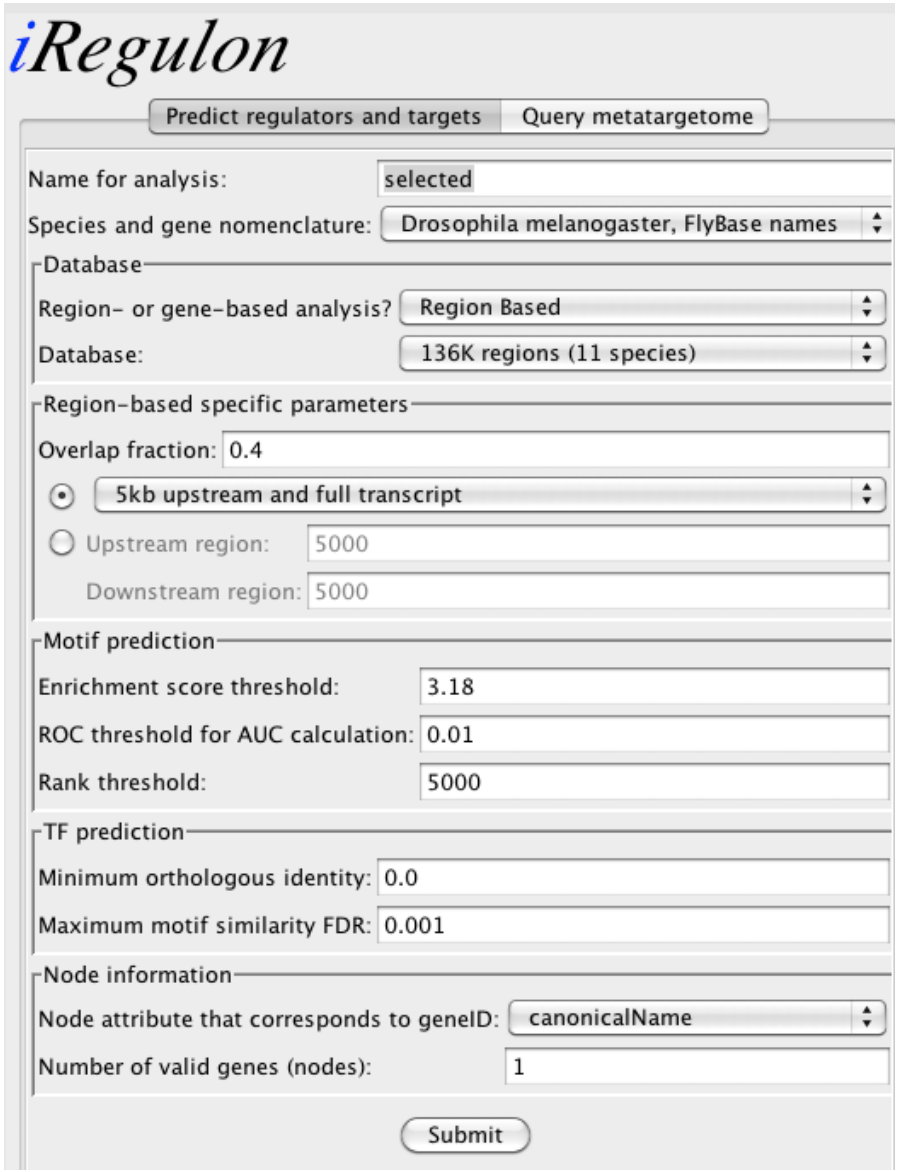
  →import network from table ....

- Select your flybase file
- Select the column2 as source interaction
- Tick "show text file import options"

  → Transfert first line as attri...

i-regulon analysis

Load i-regulon plugin ("plugins → i-regulon → add sidepanel")

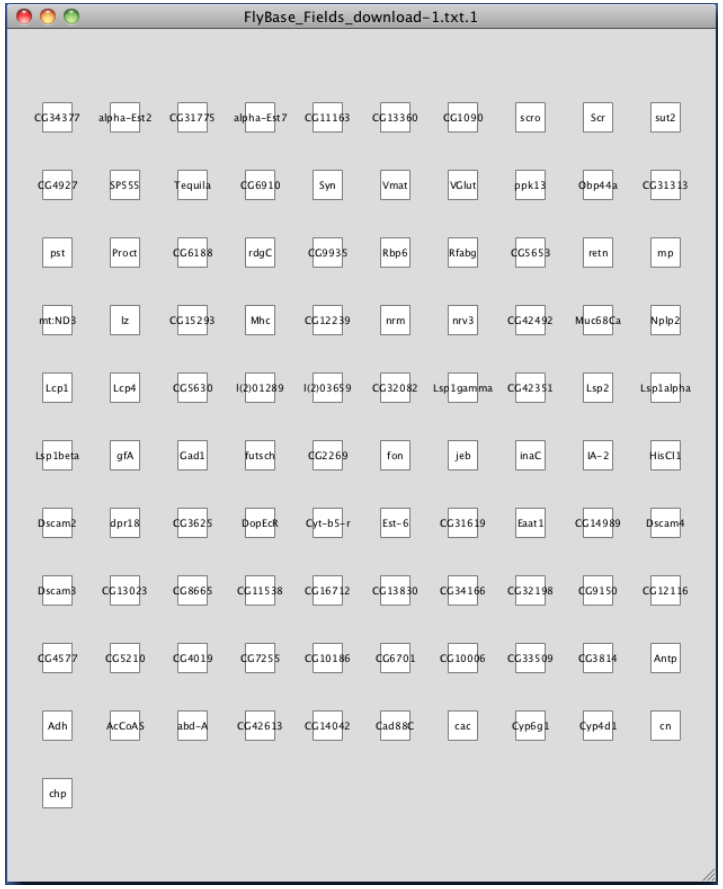


Select your putative co-regulated genes

Change default parameter to:

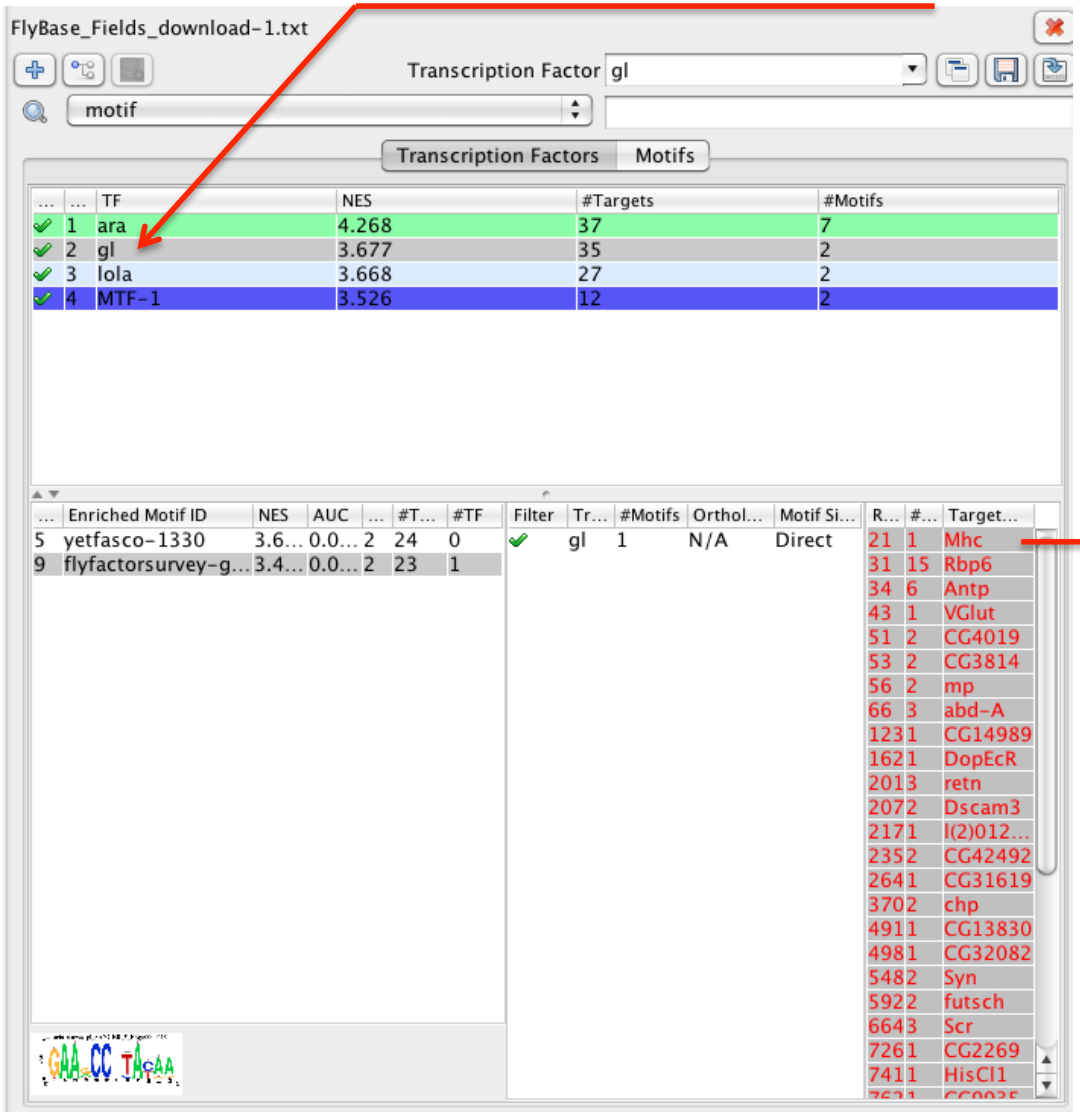
- 5kb upstream and full transcript
- ROC thresold for AUC calculation: 0.01

Submit!



i-regulon analysis

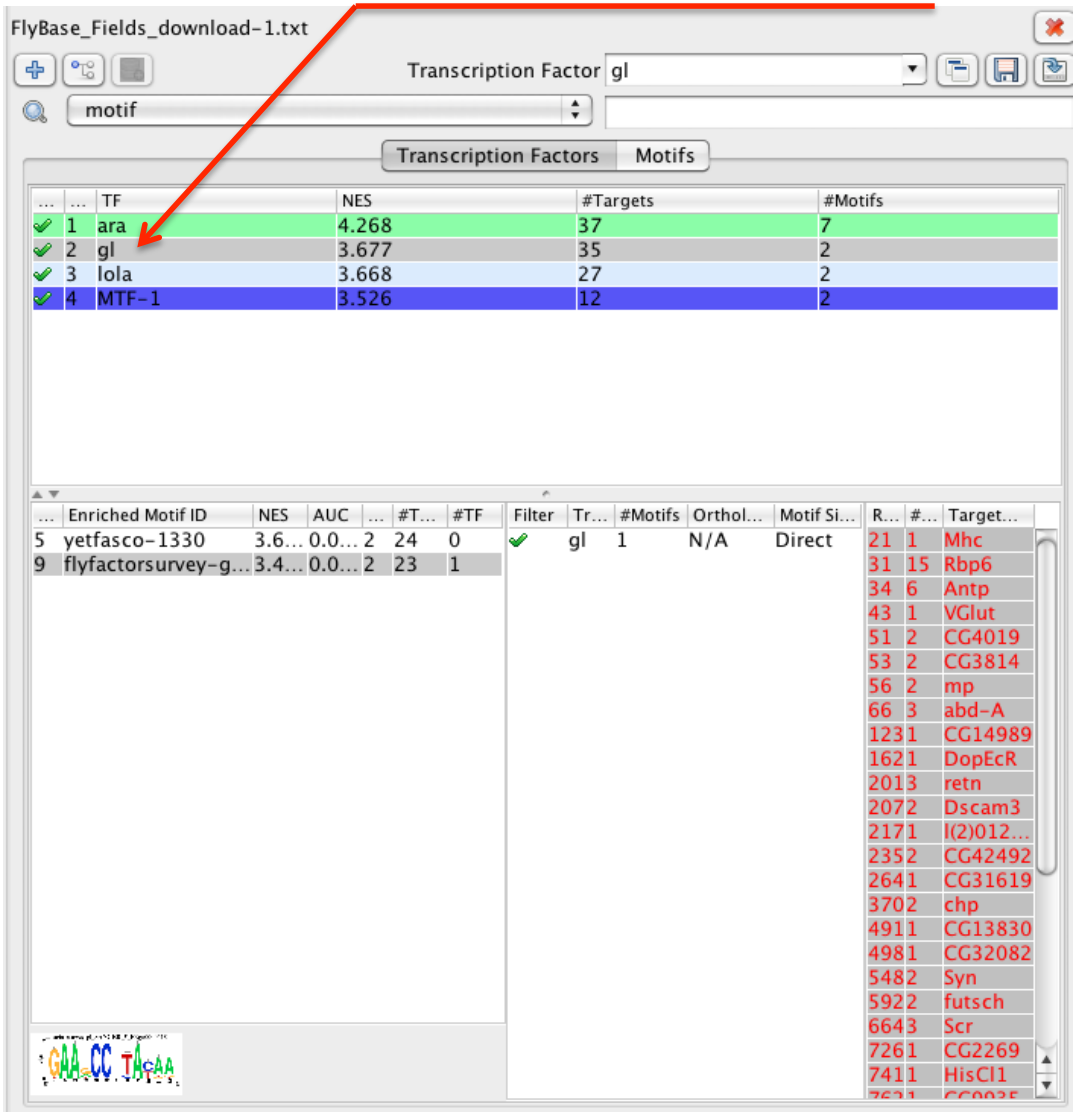
Result panel: gl appears at the 2nd position



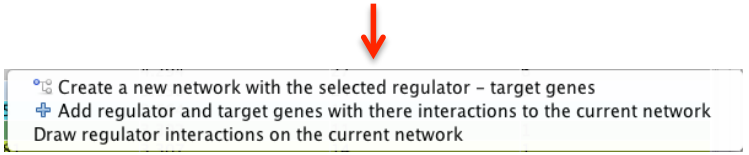
Predicted targets

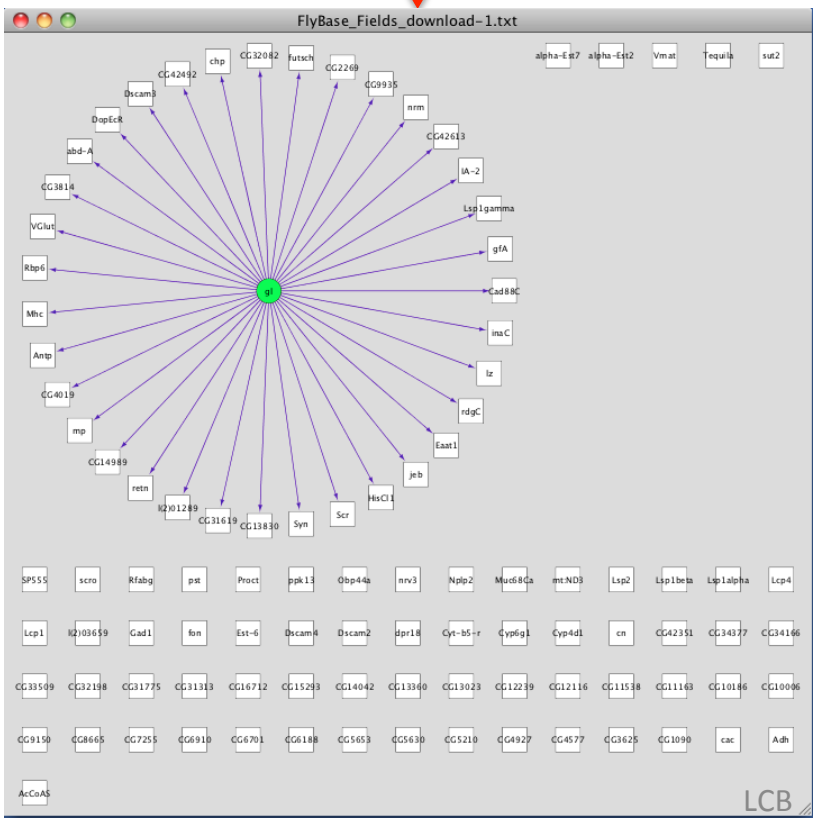
i-regulon analysis

Result panel: gl appears at the 2nd position



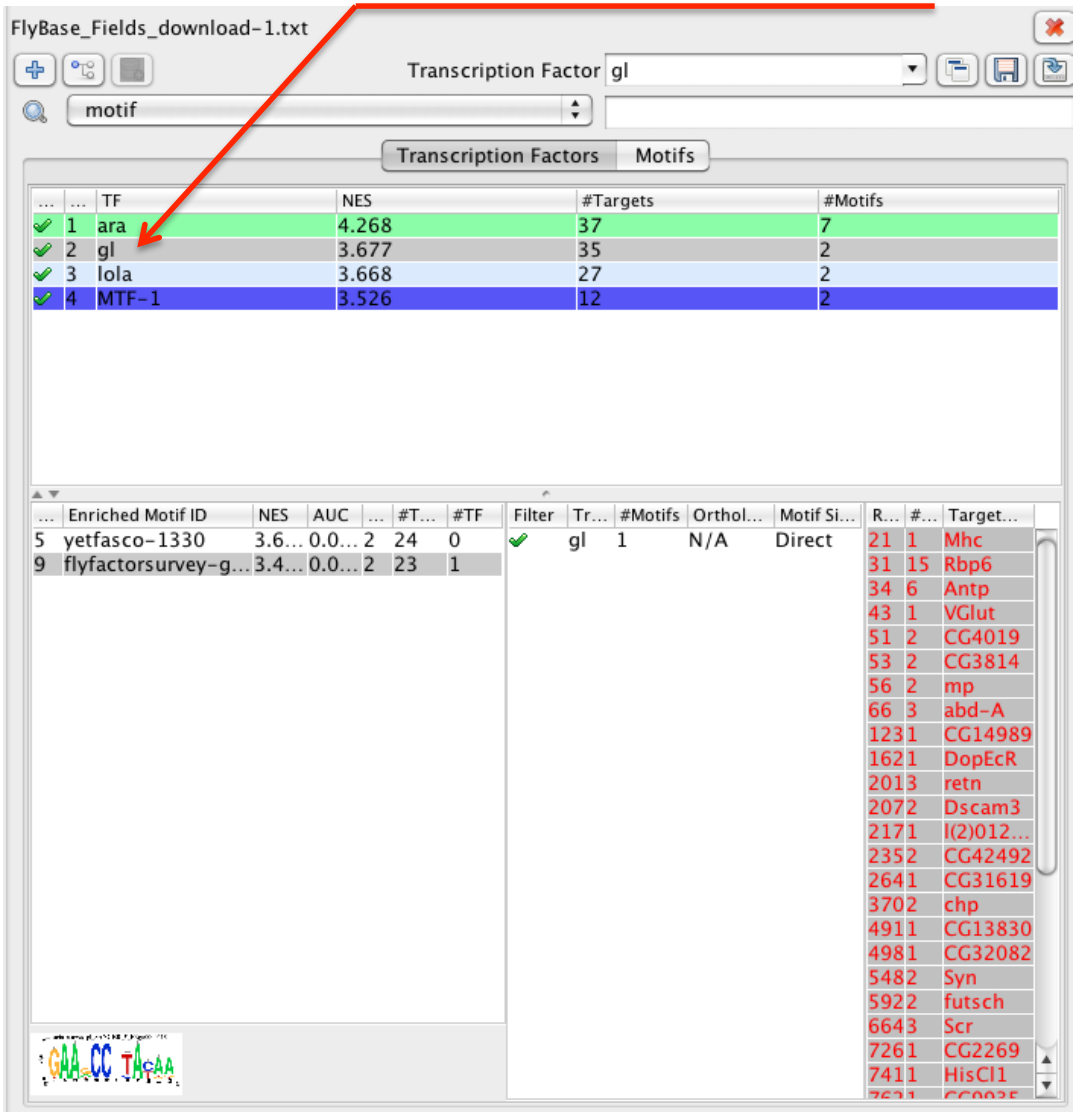
From the TF you choose (right click on it) and its predicted targets, cytoscape can draw the predicted interactions to create a network



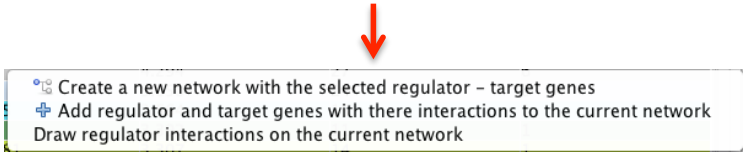


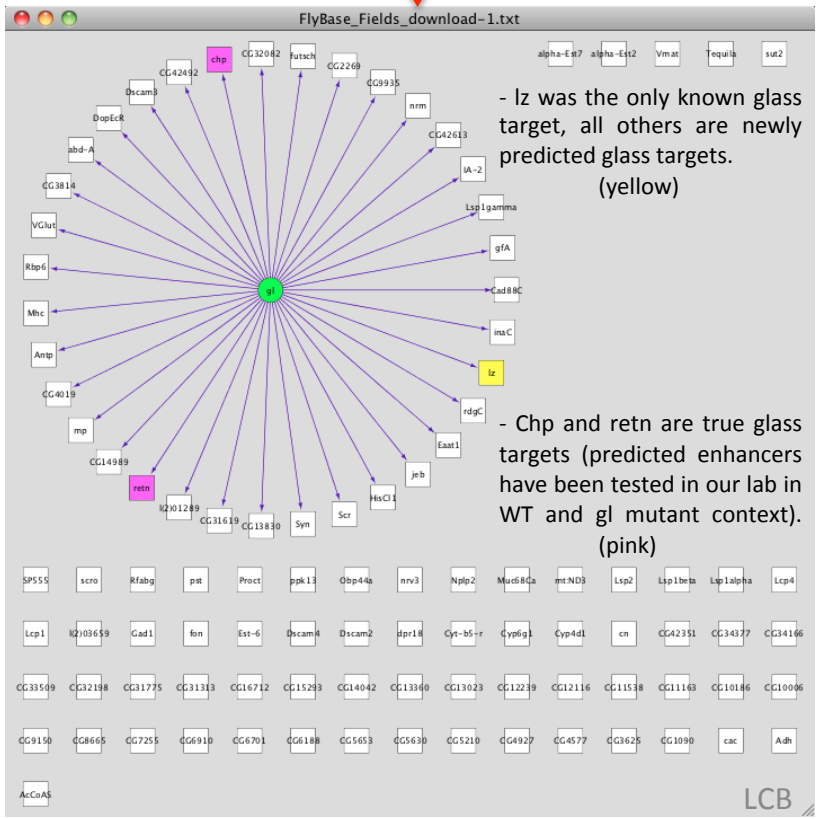
i-regulon analysis

Result panel: gl appears at the 2nd position



From the TF you choose (right click on it) and its predicted targets, cytoscape can draw the predicted interactions to create a network





- lz was the only known glass target, all others are newly predicted glass targets. (yellow)

- Chp and retn are true glass targets (predicted enhancers have been tested in our lab in WT and gl mutant context). (pink)

More galaxy training ...

- Tutorial using human adrenal and brain tissues

  → https://main.g2.bx.psu.edu/u/jeremy/p/galaxy-rna-seq-analysis-exercise

- Galaxy Screencasts and Demos

  → http://wiki.g2.bx.psu.edu/Learn/Screencasts

More advanced RNA-Seq analysis

- Use of Linux and R/Bioconductor highly recommended
- Read aggregation: HT-Seq (command-line), BEDTools/CoverageBed (command-line), sam2count (also in Galaxy) [need to map on the transcriptome]
- Normalization: EDA-Seq (GC content) (R/Bioconductor), DESeq, edgeR, …
- Differential expression: DESeq, NOISeq, edgeR, …
- Discovery of novel transcripts, update existing annotation (Cufflinks, …)
- Paired-End sequencing:
  - fusion genes (deFuse)
  - de novo transcriptome assembly (Trinity, Velvet, …)
  - Differential isoforms (DEX-Seq)
- SNP & indel detection (Samtools, …) [high coverage is needed]