Biogenouest

ifb
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

EMBRC FRANCE
CENTRE NATIONAL DE RESSOURCES BIOLOGIQUES MARINES

OCEANOMICS

IDEALG
seaweed for the future

ABiMS

15/03/2017

Galaxy

Initiation

Loraine Guéguen
Annie Lebreton

Credit to Gildas Le Corguillé – V2.3

UPMC SORBONNE UNIVERSITÉS

cnrs

Setup

- Slides available
- http://galaxy3.sb-roscoff.fr
- Login:
  - login@sb-roscoff.fr
  - **********

Plan

- Introduction
- Data import
- Tools
- Dataset
- History
- Workflow
- Share
- Dataset collections

Plan

- Introduction
- Data import 
- Tools
- Dataset 
- History
- Workflow
- Share 
- Dataset collections 

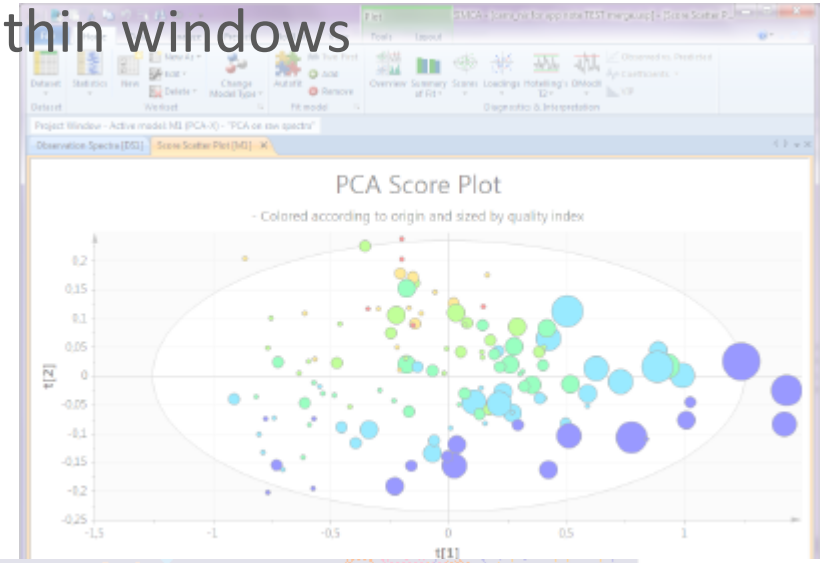
INTRODUCTION / PROBLEMATIC

Introduction / problematic

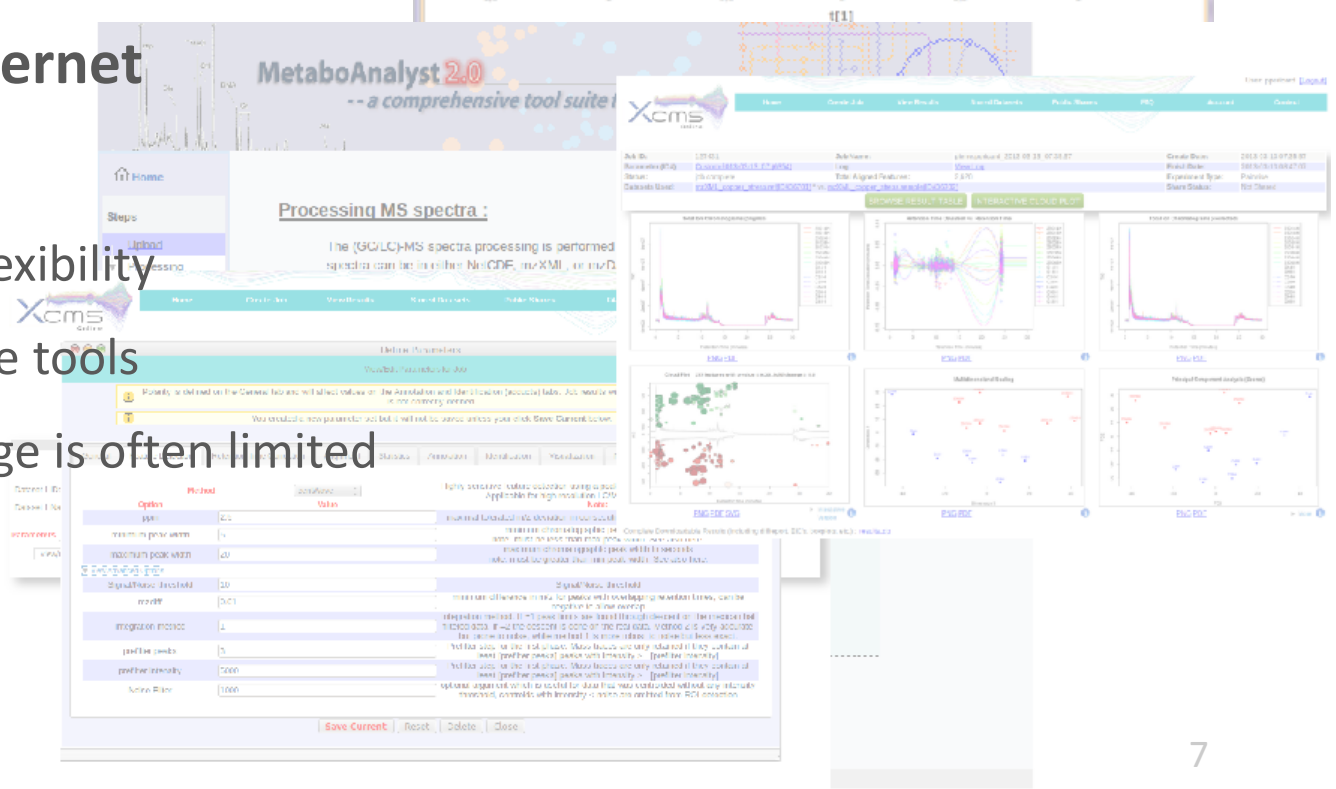
- In biomedical research, high-throughput technologies produce **large datasets**.
- How to perform analyses of these data **without bioinformatics skills** ?
    - Assemble transcript sequences de novo
    - Determine the gene expression
    - Build a phylogenetic tree
    - Predict subcellular targeting for proteins
    - Identify and quantify metabolites detected by LC-MS
    - …

Introduction / problematic

- **Graphical interface** click-button tools within windows
  - + very ergonomic
  - – too ergonomic → lack of flexibility
  - – few
  - – paying for it!



- Tools available on the **internet**
  - + very ergonomic
  - – too ergonomic → lack of flexibility
  - – A small part of the available tools
  - – the submission size / storage is often limited
  - – must not be paranoid



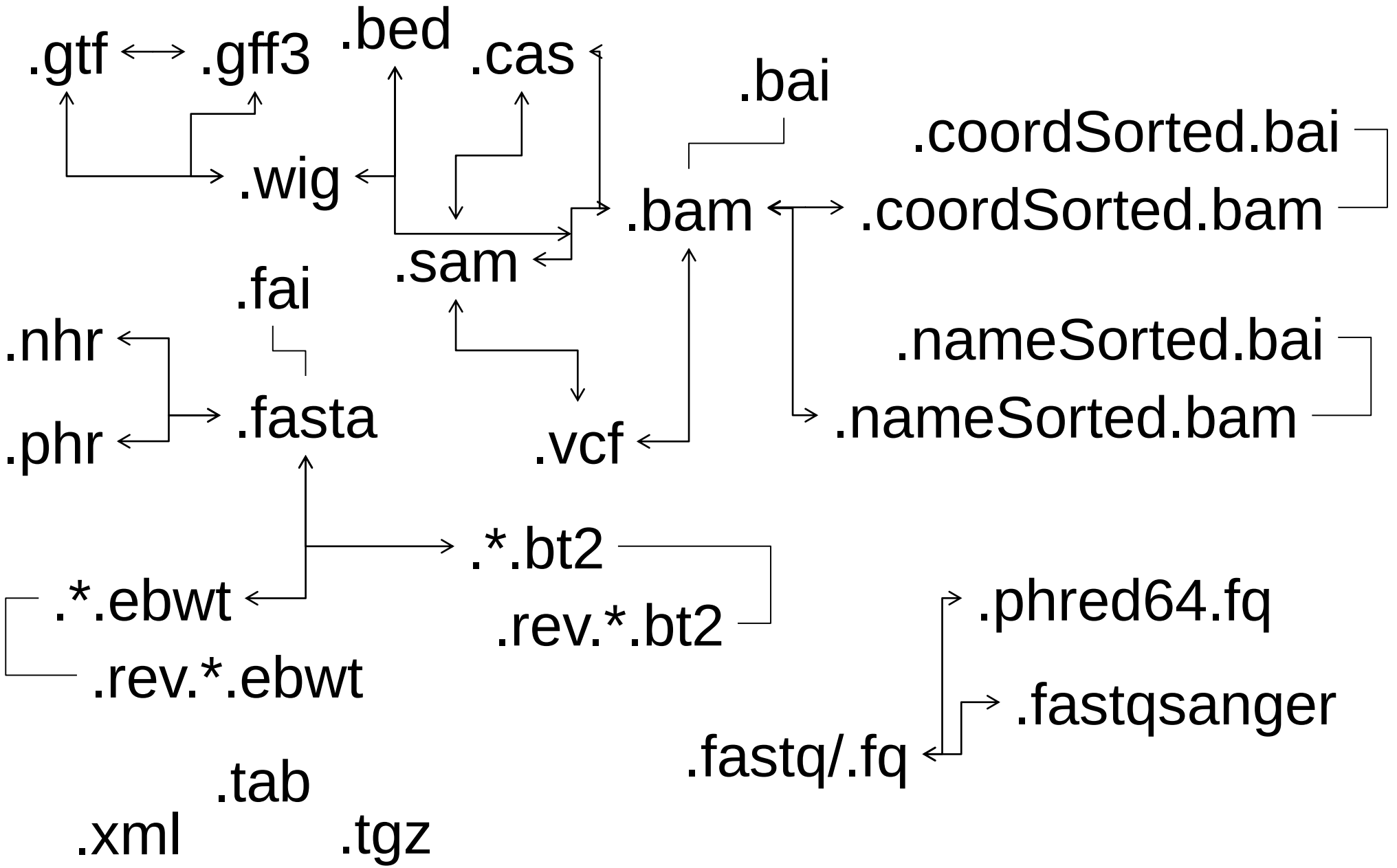
Introduction / problematic

- **Command line** tools
  - \+ represent almost the majority of scientific tools
  - \+ good parameters completeness
  - \+ can be executed on high performance computers
  - \+ automatable, workflowsable, ...
  - – minimum linux knowledge is required
  - – cruel lack of ergonomics

```
login@sbr4-1042:~$ ssh -Y login@bioinfo.sb-roscoff.fr
[...]
[login@n0 ~]$ cdprojet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v121029 ../input/IllR1-1.fq ../input/IllR1-2.fq
-GTF ../input/panda_v121029.gtf --b2-sensitive -r 100
-num-threads 8

[login@n0 mapping]$ qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
[login@n0 mapping]$ ls
accepted_hits.bam  junctions.bed    qsub.err  unmapped.bam
deletions.bed      logs             qsub.out
insertions.bed     prep_reads.info  tmp
[login@n0 mapping]$ cd ..
[login@n0 mapping]$ mkdir cufflinks
```

```
login@sbr4-1042:~$ ssh -Y login@bioinfo.sb-roscoff.fr
[...]
[login@n0 ~]$ cdprojet
[login@n0 login]$
[login@n0 mapping
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-ro
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v12                                          .R1-2.fq
-GTF ../input/par
-num-threads 8

[login@n0 mapping                                          t.qsub
Your job 5338969
[login@n0 mapping
accepted_hits.bar                                          l.bam
deletions.bed          logs                qsub.out
insertions.bed         prep_reads.info     tmp
[login@n0 mapping]$ cd ..
[login@n0 mapping]$ mkdir cufflinks
```

Introduction / problematic



Introduction / problematic

Select your level:

Introduction / problematic

Level 1

« I want to know the gene expression »

Introduction / problematic

Level 2

« I want to map my reads on a reference genome and count them »

Introduction / problematic

Level 3

« I want to launch the tools tophat2 and cufflinks.
I have fastq files and my genome in fasta and gtf. »

Introduction / problematic

Level 4

« I want 1TB for my project. I will launch tophat2 through SSH on the cluster in multi-thread mode.

Next I want to submit the bam file to my genome with cufflinks.

Except that, I will manage :P”

Introduction / problematic

Level 5

« I have a bunch of cool tools!
But I'm the only one who can launch them.

Comments? »


Galaxy / 4 / Metabolomics
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using -993344424 b
Tools
search tools
Upload File from your computer
Export Data
LC-MS
Format Conversion
Preprocessing
Normalisation
Quality Control
Statistical Analysis
Annotation
GC-MS
Preprocessing
Normalisation
Quality Control
Statistical Analysis
Annotation
NMR
Preprocessing
Normalisation
Quality Control
Statistical Analysis
COMMON TOOLS
Data Handling
Text Manipulation
Filter and Sort
Join, Subtract and Group
xcms.xcmsSet
version 2.0.1
Choose your inputs method:
Zip file from your history containing your chromatograms
Zip file:
1: sacuri.zip
Extraction method for peaks detection:
matchedFilter
[method] See the help section below
Step size to use for profile generation:
0.01
[step] The peak detection algorithm creates extracted ion base peak chromatograms (EIBPC) on a fixed step size
Full width at half maximum of matched filtration gaussian model peak:
30
[fwhm] Only used to calculate the actual sigma
Advanced options:
hide
Execute
Authors Colin A. Smith csmith@scripps.edu, Ralf Tautenhahn rtautenh@gmail.com, Steffen Neumann sneumann@ipb-halle.de, Paul Benton hpaul.benton08@imperial.ac.uk and Christopher Conley cjconley@ucdavis.edu
If you use this tool, please cite: Smith,C.A. et al.(2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal. Chem., 78, 779–787.
For details about this tool, please go to http://www.bioconductor.org/packages/release/bioc/html/xcms.html
Galaxy integration ABIMS TEAM, Station biologique de Roscoff.
Contact support@workflow4metabolomics.org for any questions or concerns about the Galaxy implementation of this tool.
History
search datasets
Sacuri Zip
19 shown
289.7 MB
19: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv (Xdiffreport)
18: xset.group.retcor.group.fillPeaks.annotate.negative.Rdata
17: xset.group.retcor.group.fillPeaks.annotate.dataMatrix.tsv
16: xset.group.retcor.group.fillPeaks.annotate.variableMetadata.tsv
15: xset.group.retcor.group.fillPeaks.RData
14: xset.group.retcor.group.Rplots.pdf
13: xset.group.retcor.group.RData
12: xset.group.retcor.BPCs_corrected.pdf
11:

INTRODUCTION / GALAXY

Why Galaxy ?

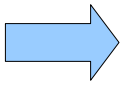
- Accessibility
- Reproductibility
- Transparency

Introduction / Galaxy

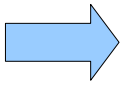
- Galaxy it's …
  - A web-based interface
  - No need to execute a command line through a terminal
  - Programming or scripting skills are not required
  - Submission of jobs is transparent through a high performance computer cluster
  - Secure histories and data manager
  - A data and protocols sharing system
  - Tool-boxes of several bioinformatics fields
    - NGS
    - Metabolomics
    - Statistics
    - Chemistry
    - Image analysis
    - Etc …

Introduction / Galaxy

MR. GEEK

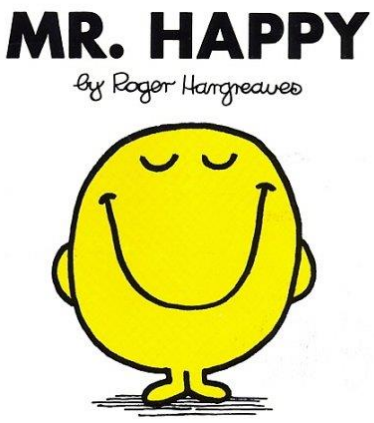


```
[login@n0 ~]$ cdprojet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v121029 ../input/IllR1-1.fq ../input/IllR1-2.fq
-GTF ../input/panda_v121029.gtf --b2-sensitive -r 100
-num-threads 8

[login@n0 mapping]$ qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
```



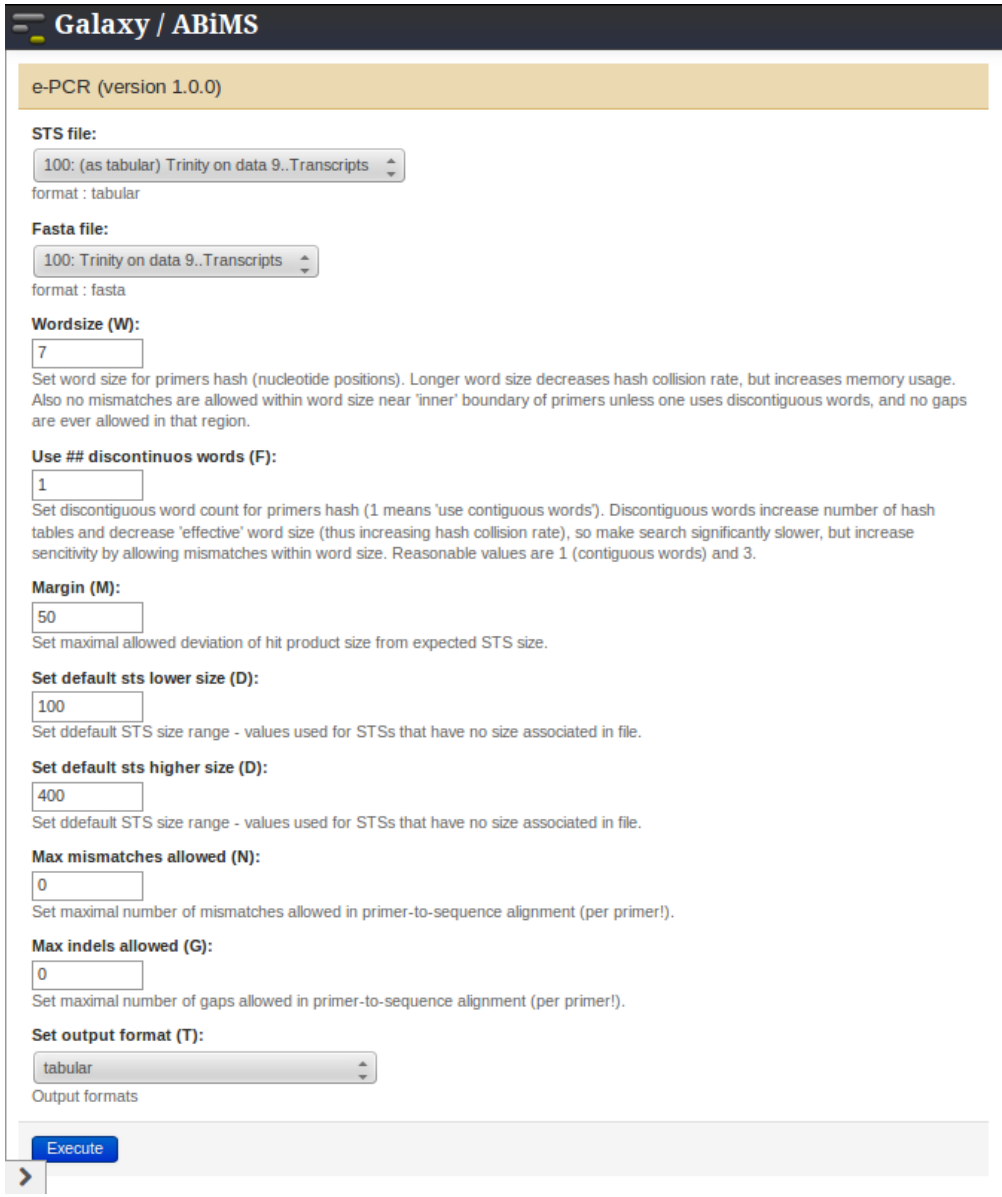
Introduction / Galaxy

```
[login@n0 ~]$ cdprojet
[login@n0 login]$ cd 13-07-29-panda/tmp/mapping
[login@n0 mapping]$ cat tophat.qsub
#!/bin/bash
#$ -S /bin/bash
#$ -M login@sb-roscoff.fr
#$ -m bea
#$ -V
#$ -cwd
#$ -o qsub.out
#$ -e qsub.err

tophat2 panda_v121029 ../input/Ill1R1-1.fq ../input/Ill1R1-2.fq
-GTF ../input/panda_v121029.gtf --b2-sensitive -r 100
-num-threads 8

[login@n0 mapping]$ qsub -q long.q -pe thread 8 tophat.qsub
Your job 5338969 ("tophat.qsub") has been submitted
```

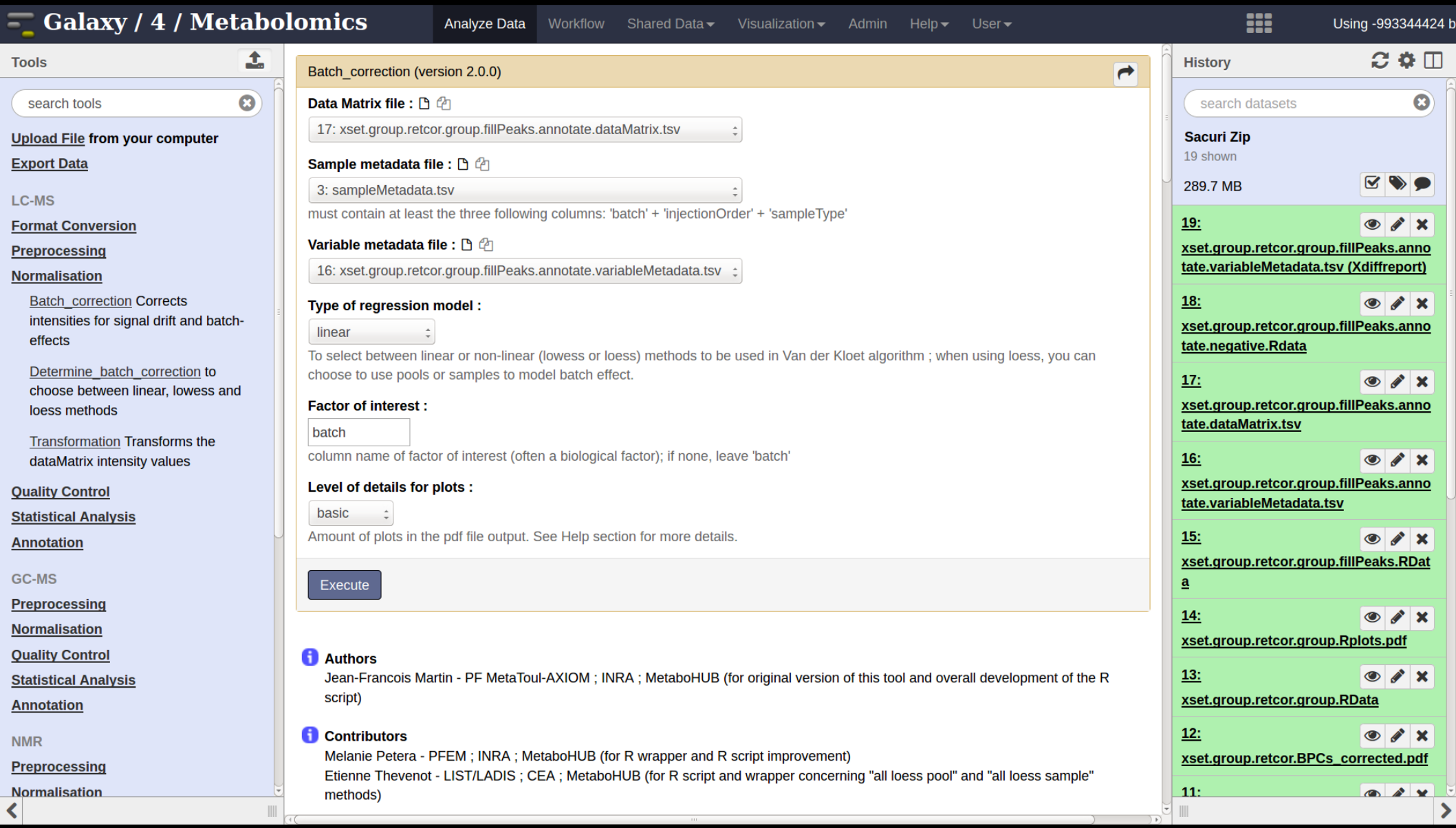


Introduction / Galaxy

```
[lecorguille@n0 ~]$ e-PCR --help
e-PCR: invalid option -- -
usage: [-hV] [posix-options] stsfile [fasta ...]
[compat-options]
where posix-options are:
          -m ##     Margin (default 50)
          -w ##     Wordsize  (default 7)
          -n ##     Max mismatches allowed (default 0)
          -g ##     Max indels allowed (default 0)
          -f ##     Use ## discontiguos words, slow if
                              ##>1
          -o ##     Set output file
          -t ##     Set output format:
                    1 - classic, range (pos1..pos2)
                    2 - classic, midpoint
                    3 - tabular
                    4 - tabular with alignment in
comments
                              (slow)
          -d##-##   Set default size range
                              (default 100-350)
          -p +-     Turn hits postprocess on/off
          -v ##     Verbosity flags
          -a a|f    Use presize alignmens (only if
                             gaps>0), slow
                     a - Allways or f - as Fallback
          -x +-     Use 5'-end lowercase masking of
                             primers (default -)
          -u +-     Uppercase all primers (default -)

[...]
```

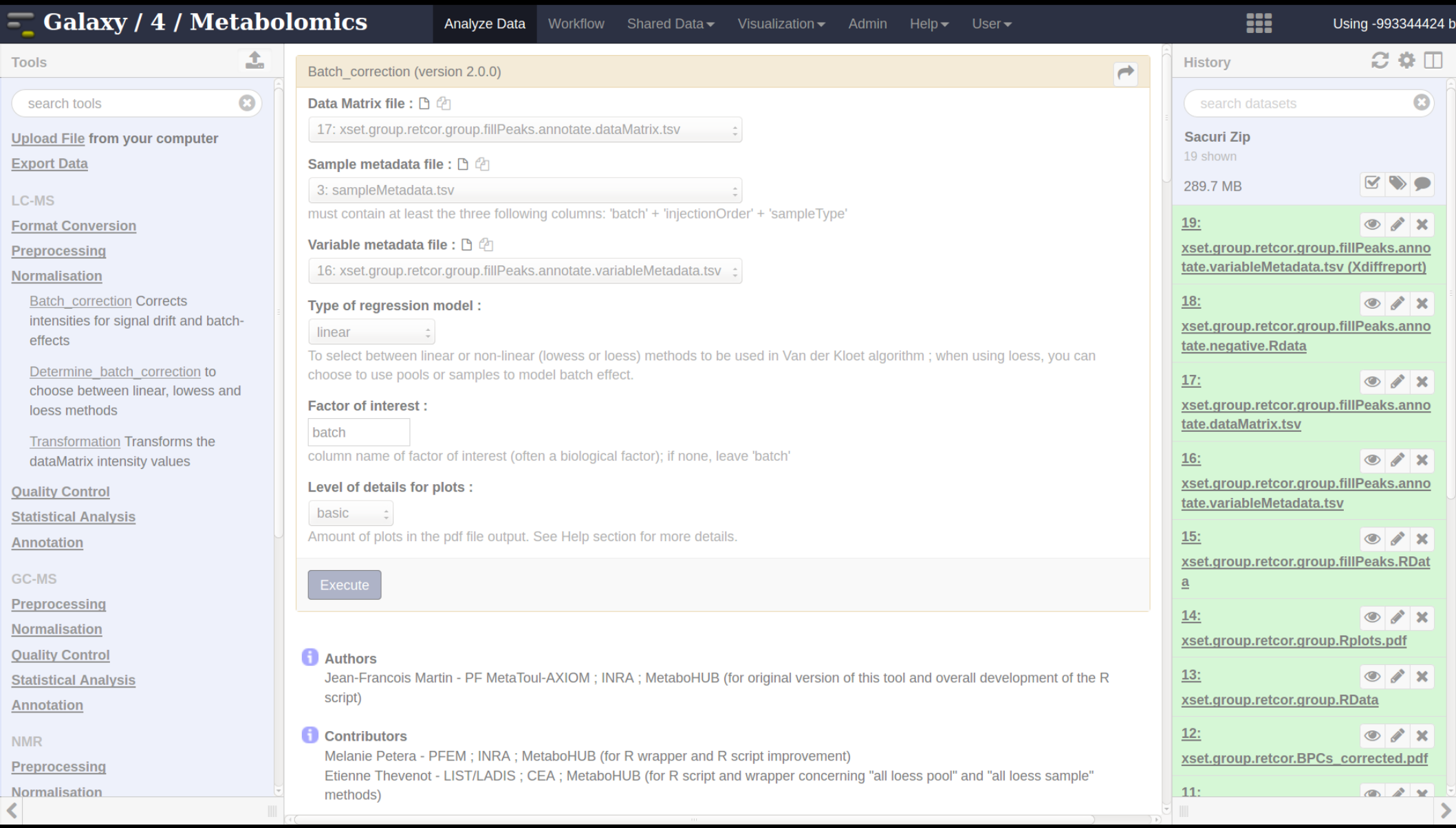


Galaxy interface



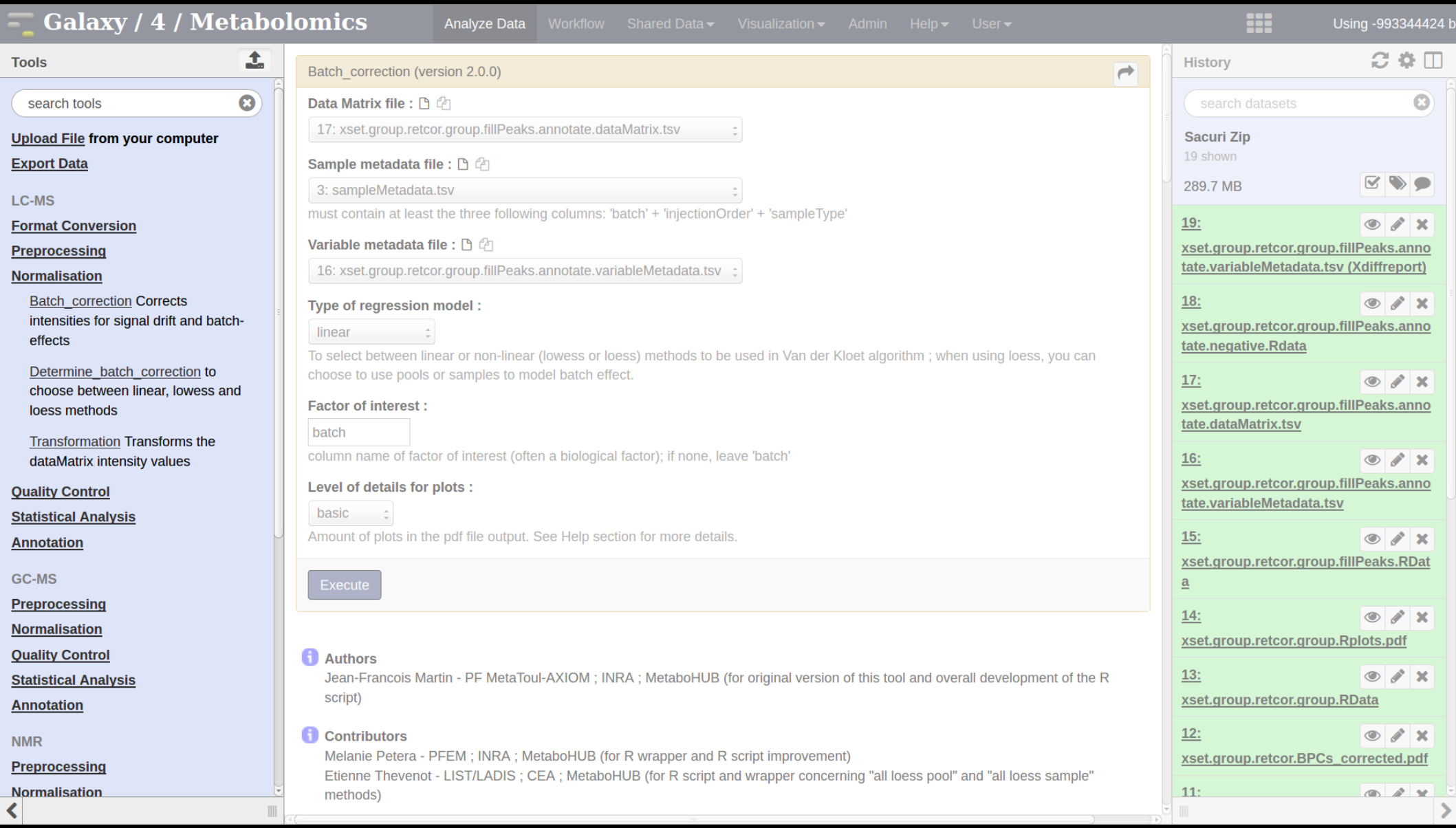
Galaxy interface

Menu



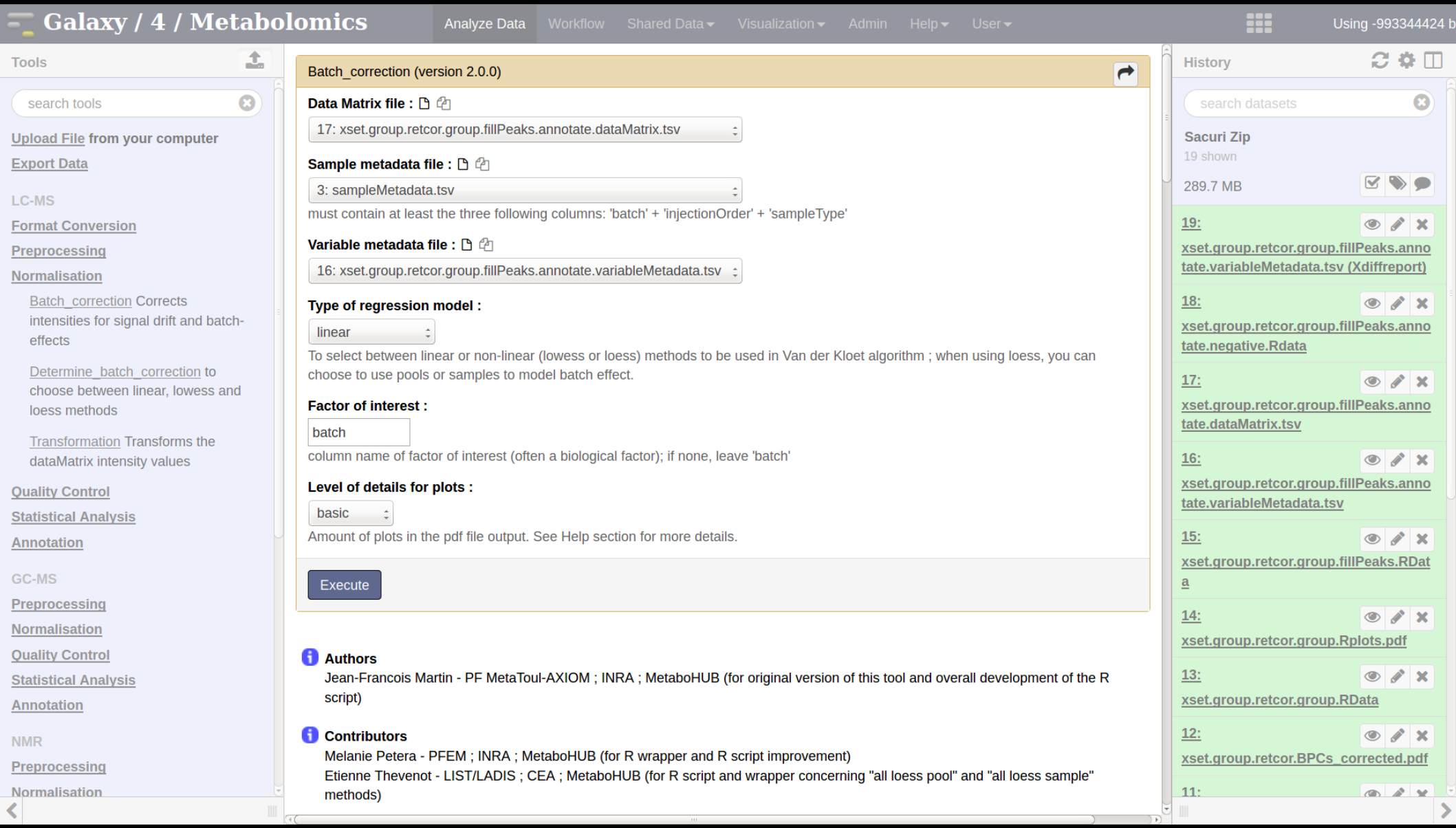
Galaxy interface

Tool list



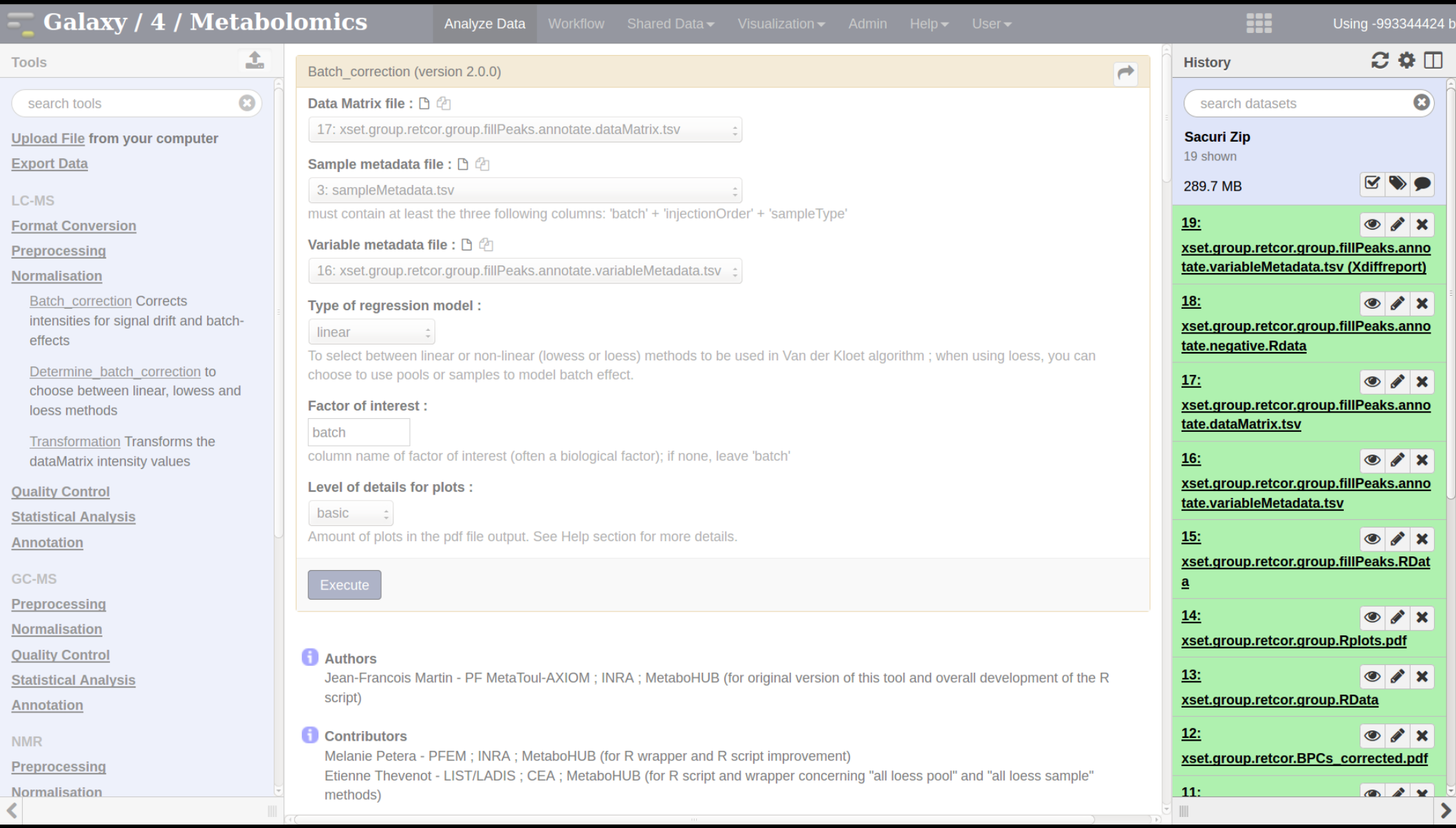
Galaxy interface

Web forms / dataset visualization / diverse information



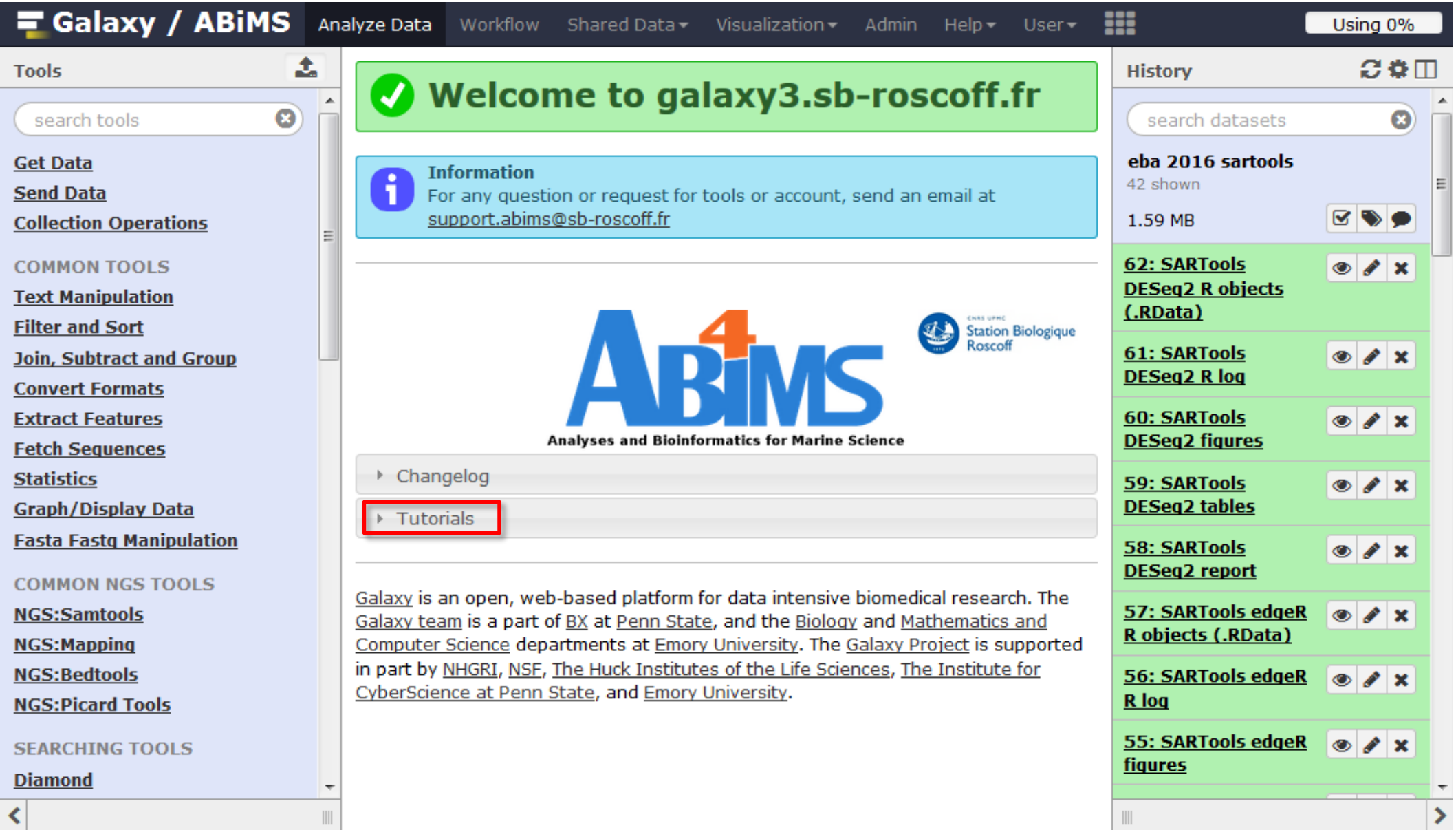
Galaxy interface

History

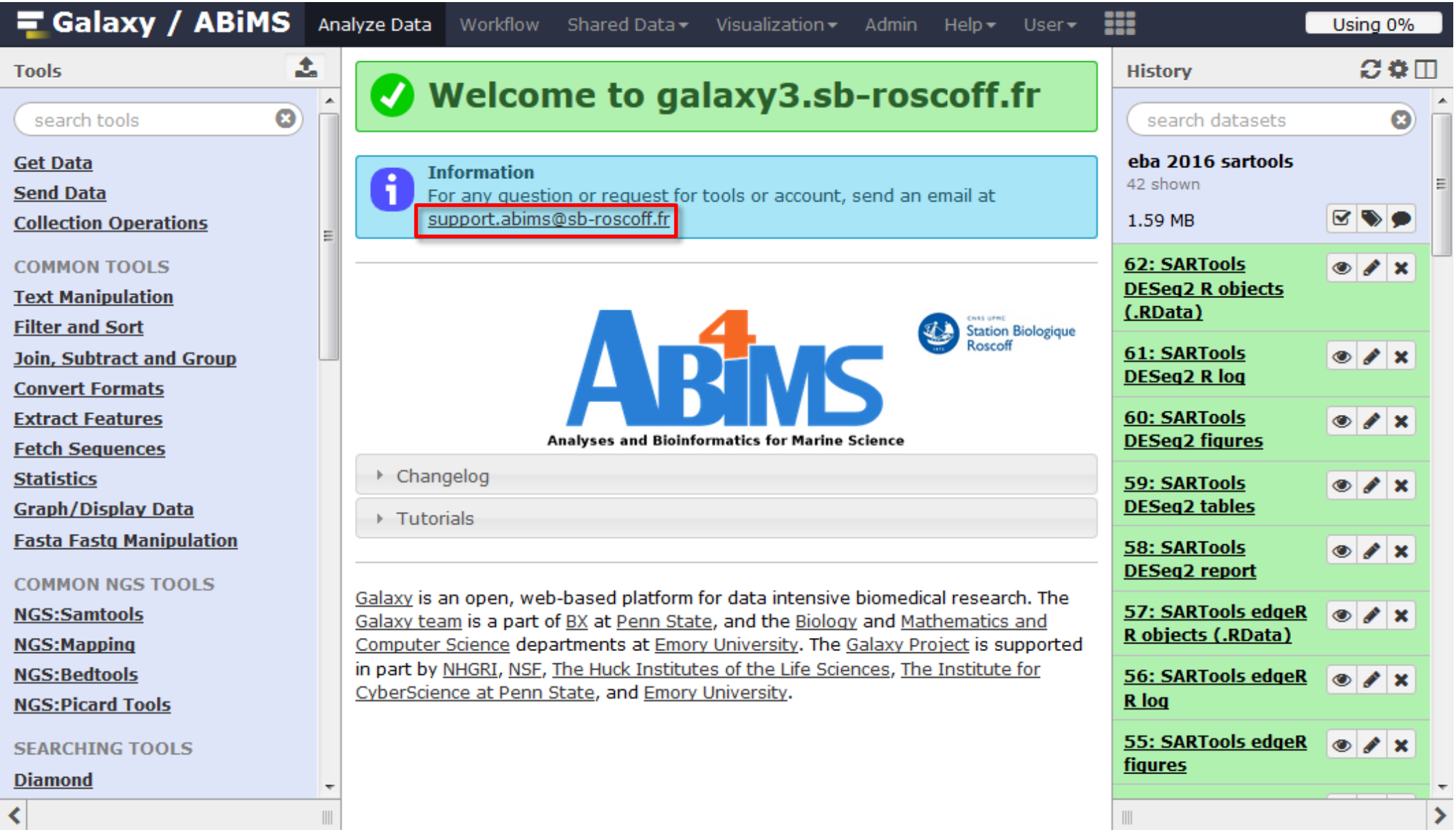


GET HELP

Get help


Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 0%
Tools
search tools
Get Data
Send Data
Collection Operations
COMMON TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Statistics
Graph/Display Data
Fasta Fastq Manipulation
COMMON NGS TOOLS
NGS:Samtools
NGS:Mapping
NGS:Bedtools
NGS:Picard Tools
SEARCHING TOOLS
Diamond
Welcome to galaxy3.sb-roscoff.fr
Information
For any question or request for tools or account, send an email at support.abims@sb-roscoff.fr
Analyses and Bioinformatics for Marine Science
Station Biologique Roscoff
Changelog
Tutorials
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
eba 2016 sartools
42 shown
1.59 MB
62: SARTools DESeq2 R objects (.RData)
61: SARTools DESeq2 R log
60: SARTools DESeq2 figures
59: SARTools DESeq2 tables
58: SARTools DESeq2 report
57: SARTools edgeR R objects (.RData)
56: SARTools edgeR R log
55: SARTools edgeR figures

Get help

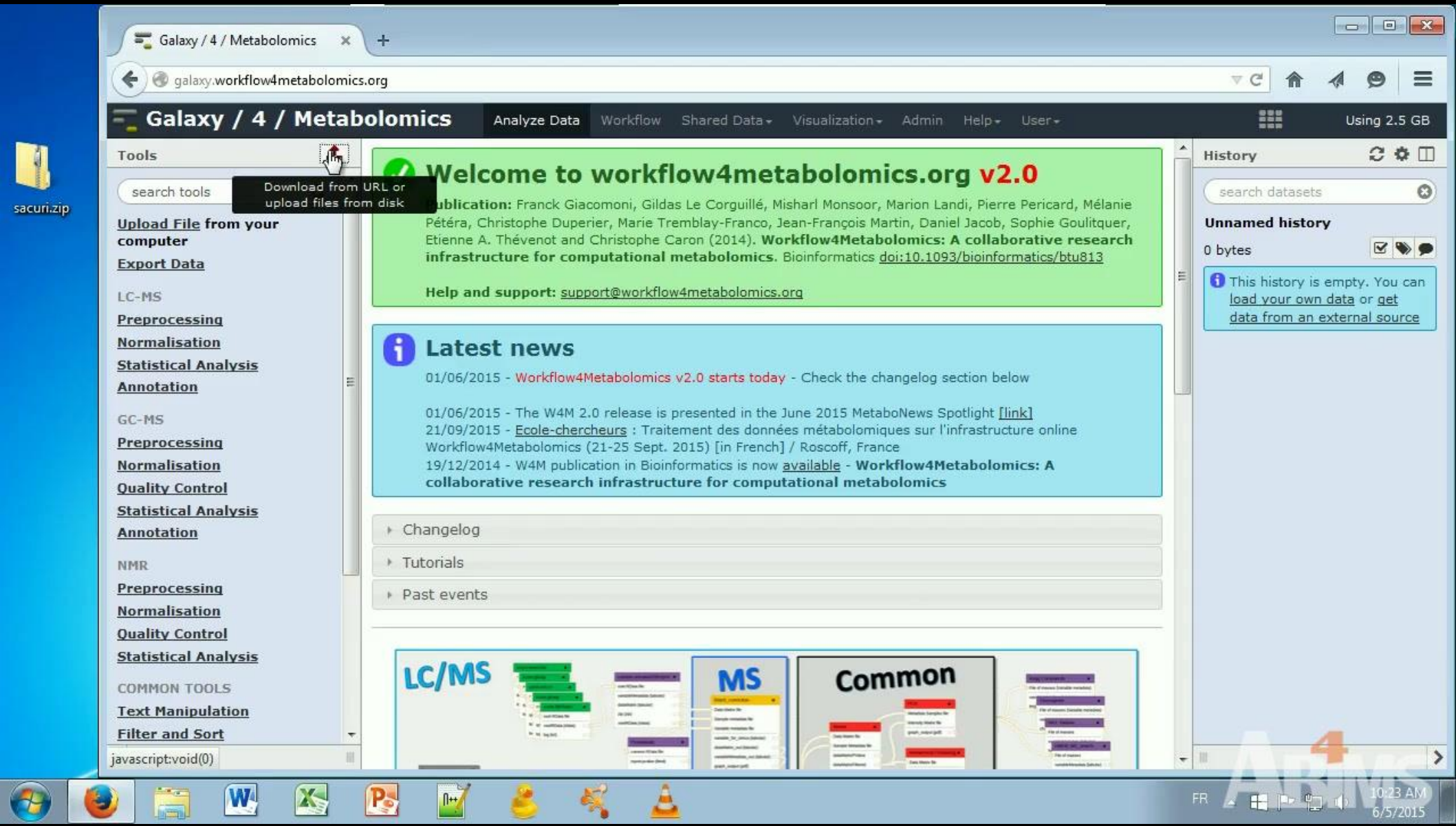

Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 0%
Tools
search tools
Get Data
Send Data
Collection Operations
COMMON TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Statistics
Graph/Display Data
Fasta Fastq Manipulation
COMMON NGS TOOLS
NGS:Samtools
NGS:Mapping
NGS:Bedtools
NGS:Picard Tools
SEARCHING TOOLS
Diamond
Welcome to galaxy3.sb-roscoff.fr
Information
For any question or request for tools or account, send an email at
support.abims@sb-roscoff.fr
Analyses and Bioinformatics for Marine Science
Station Biologique Roscoff
Changelog
Tutorials
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
eba 2016 sartools
42 shown
1.59 MB
62: SARTools DESeq2 R objects (.RData)
61: SARTools DESeq2 R log
60: SARTools DESeq2 figures
59: SARTools DESeq2 tables
58: SARTools DESeq2 report
57: SARTools edgeR R objects (.RData)
56: SARTools edgeR R log
55: SARTools edgeR figures

DATA IMPORT

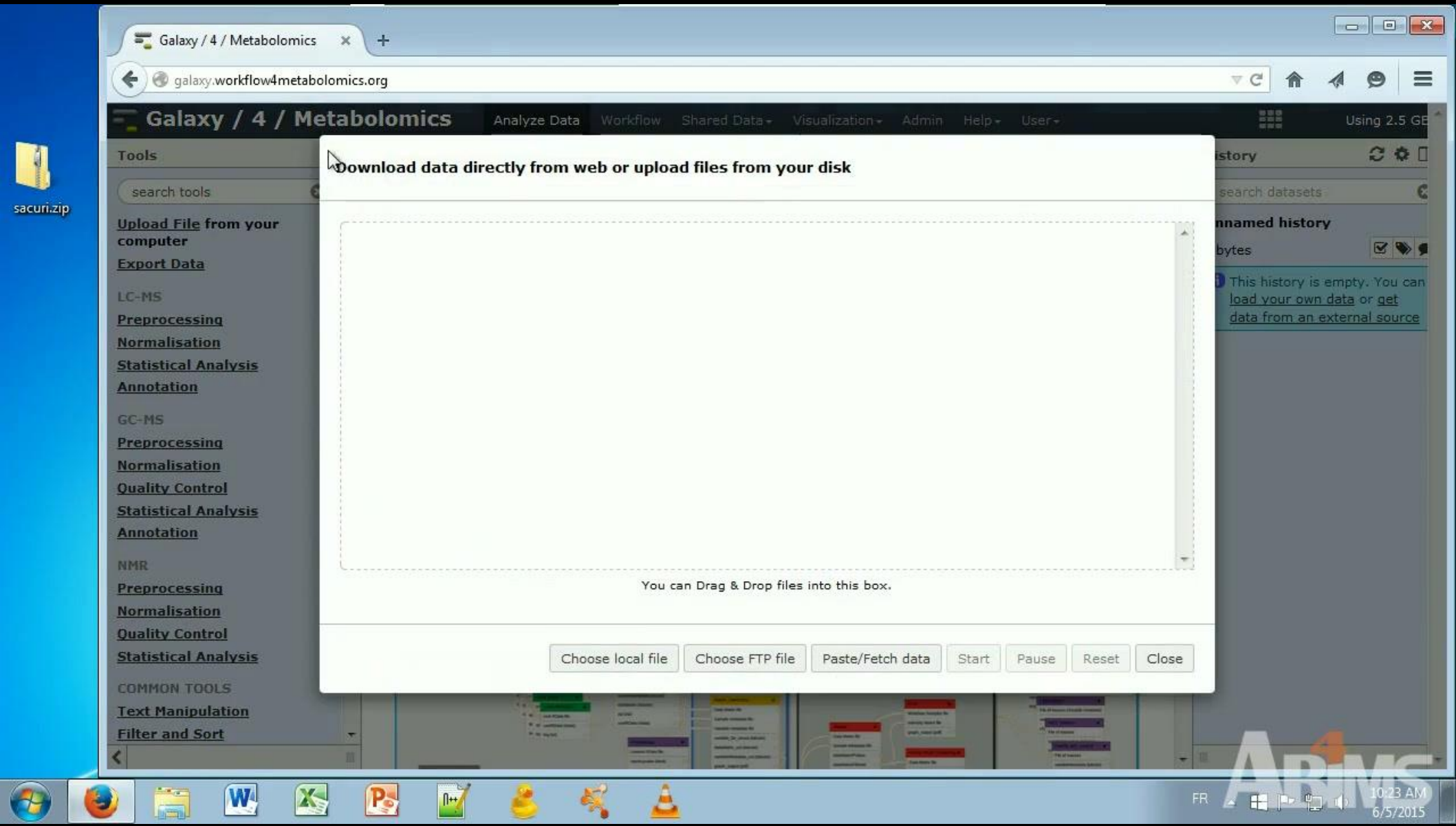
DATA IMPORT

< 2 GO

Data import < 2 Go

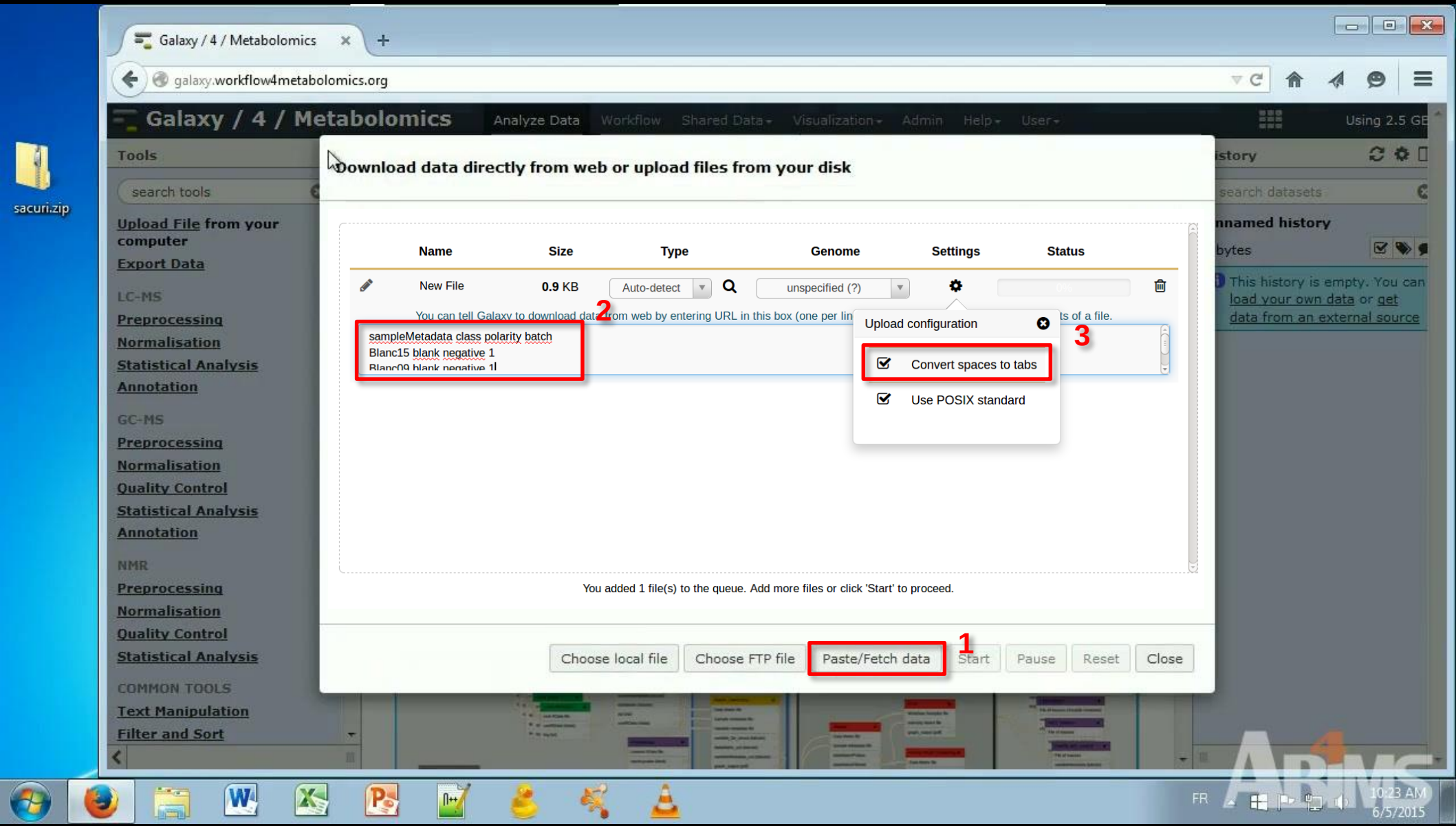


Data import < 2 Go



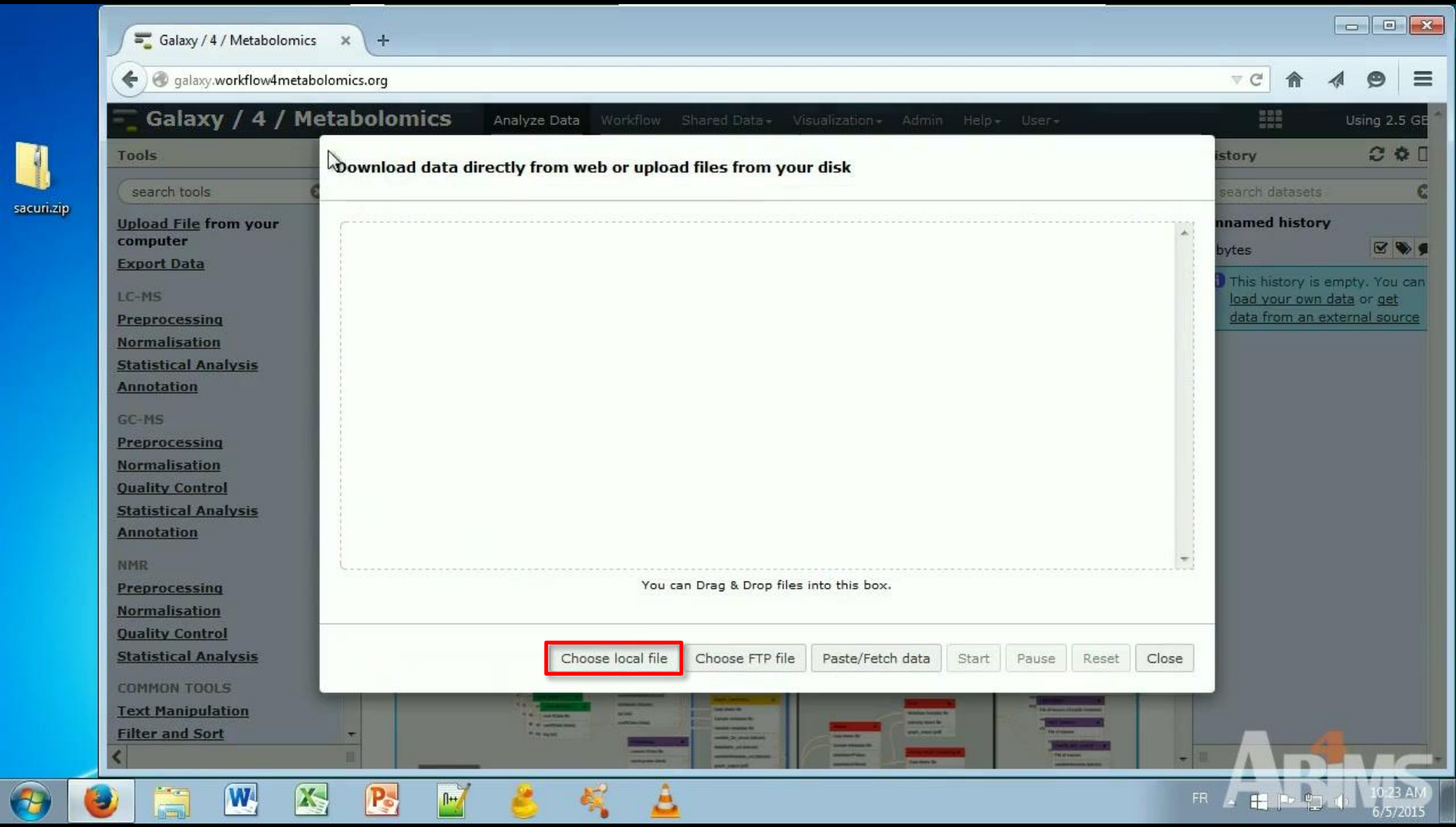
Data import < 2 Go

Copy / Paste data



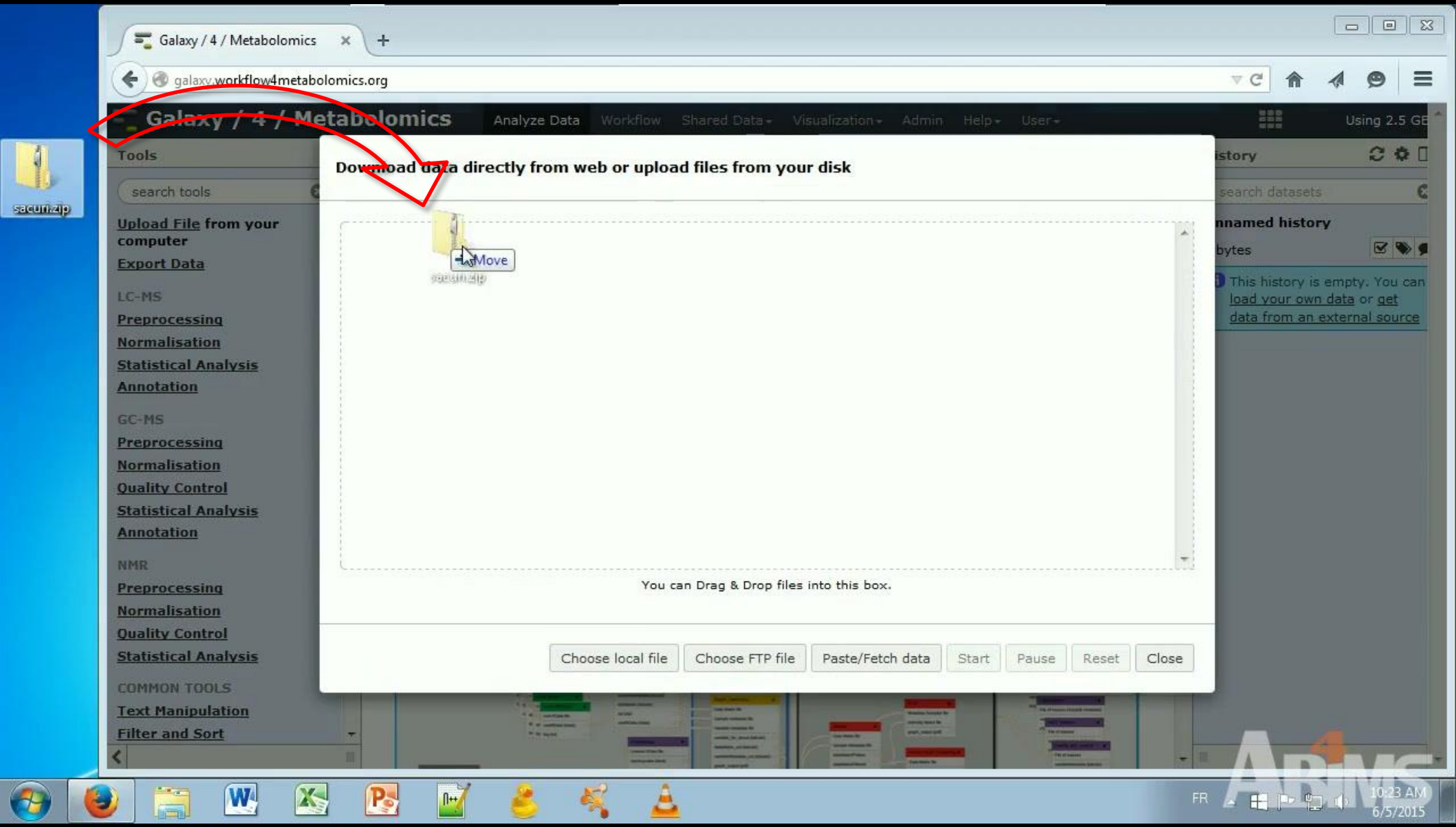
Data import < 2 Go

From local files



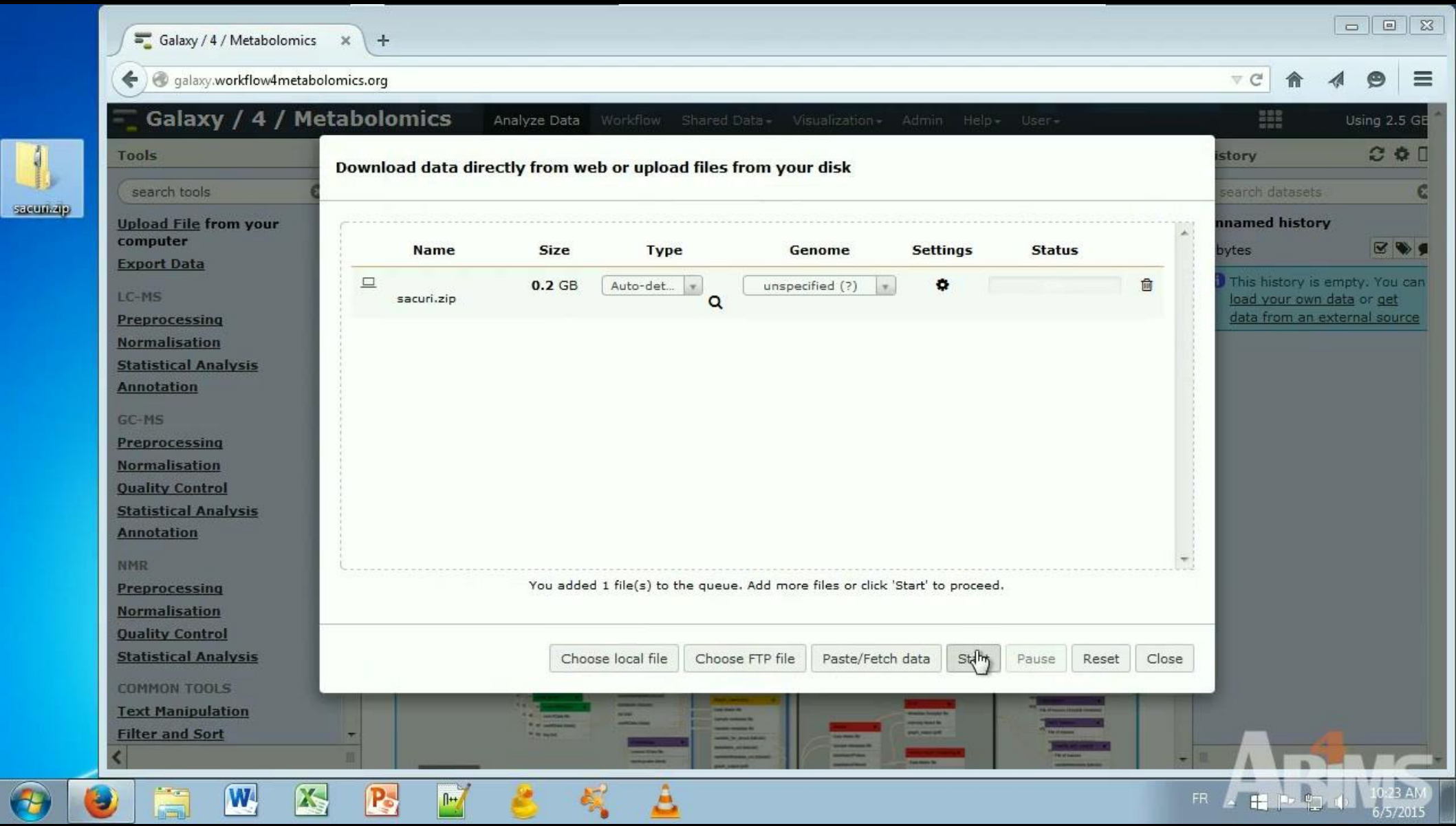
Data import < 2 Go

From local files



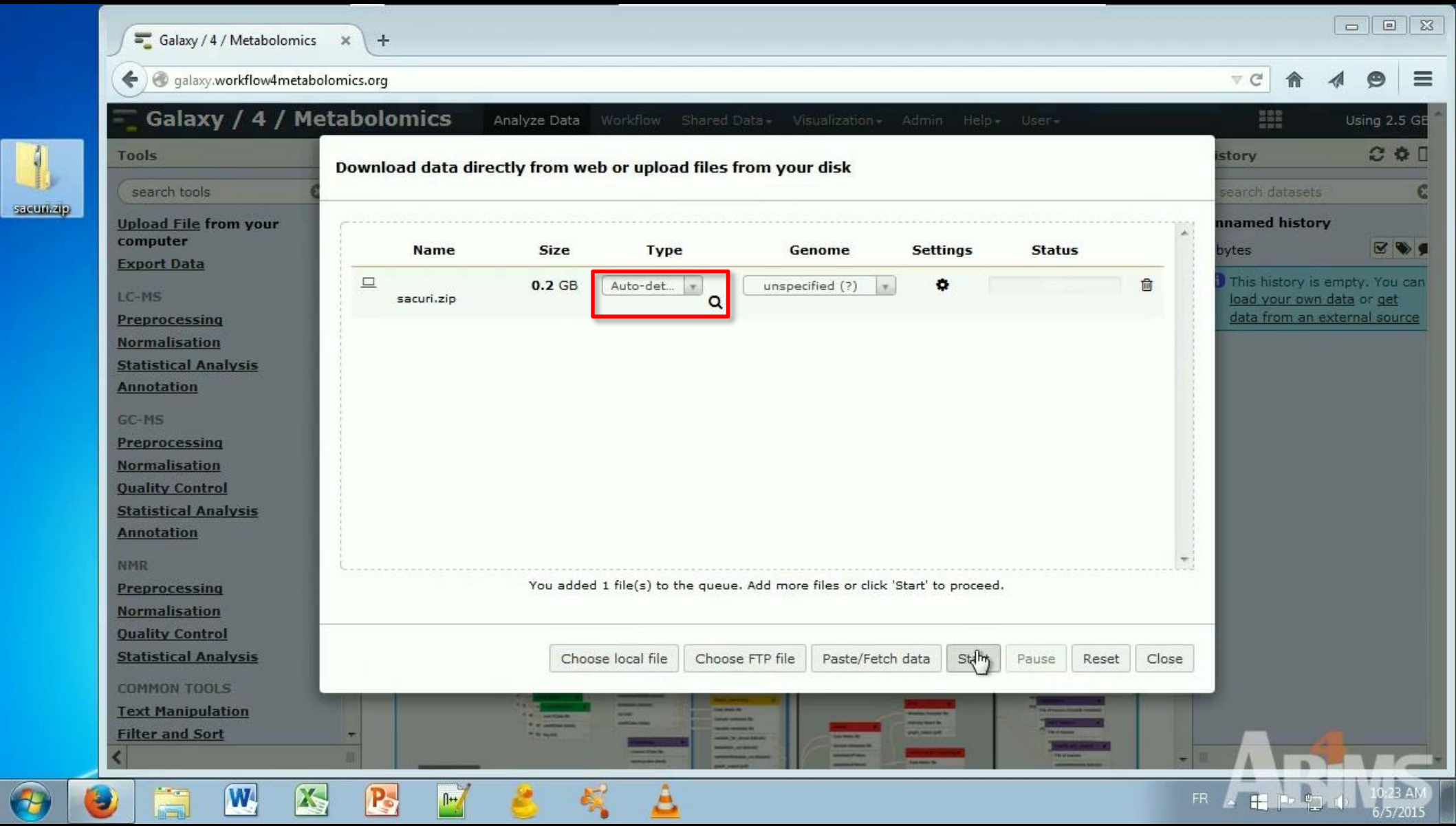
Data import < 2 Go

From local files



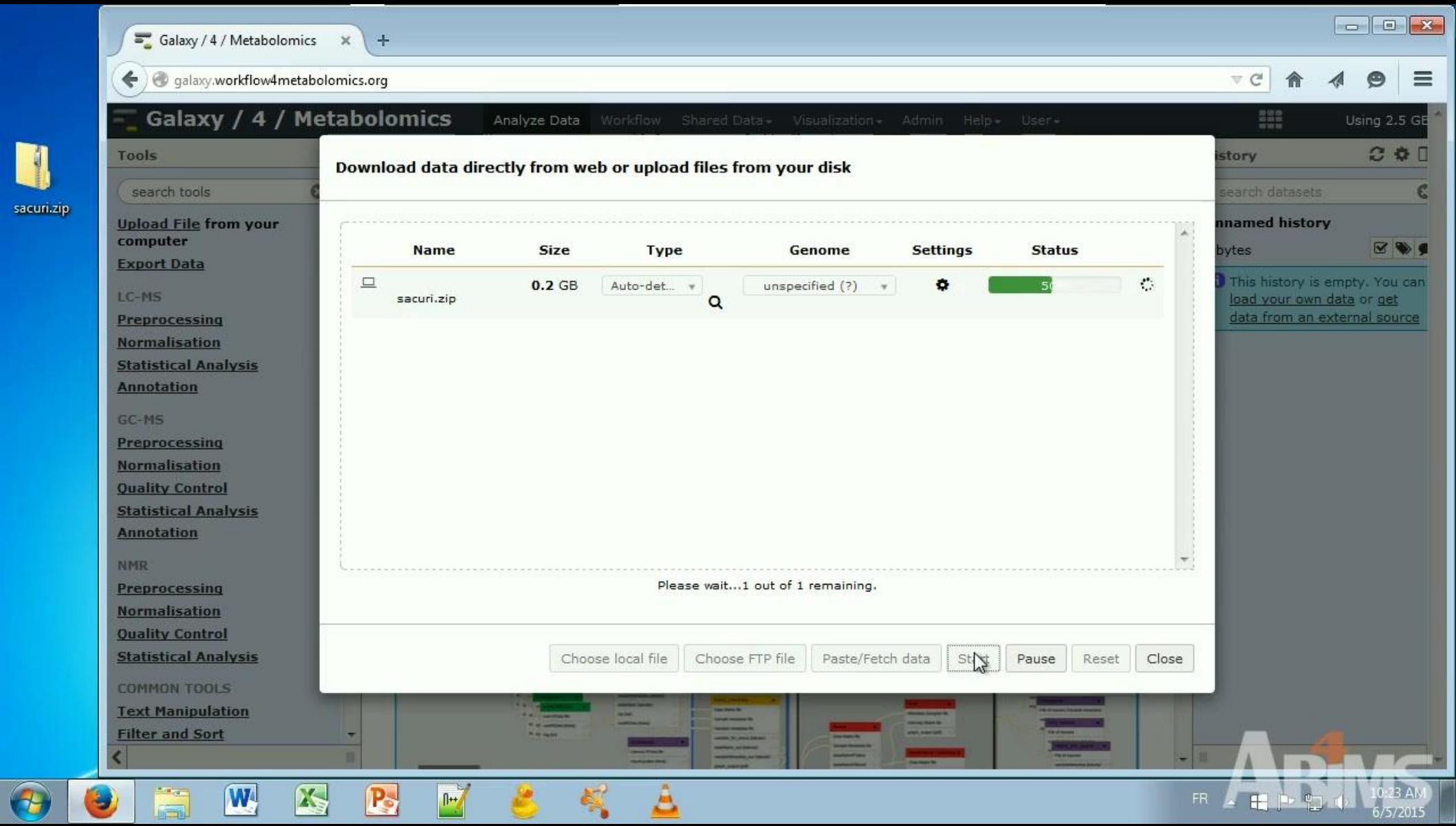
Data import < 2 Go

From local files



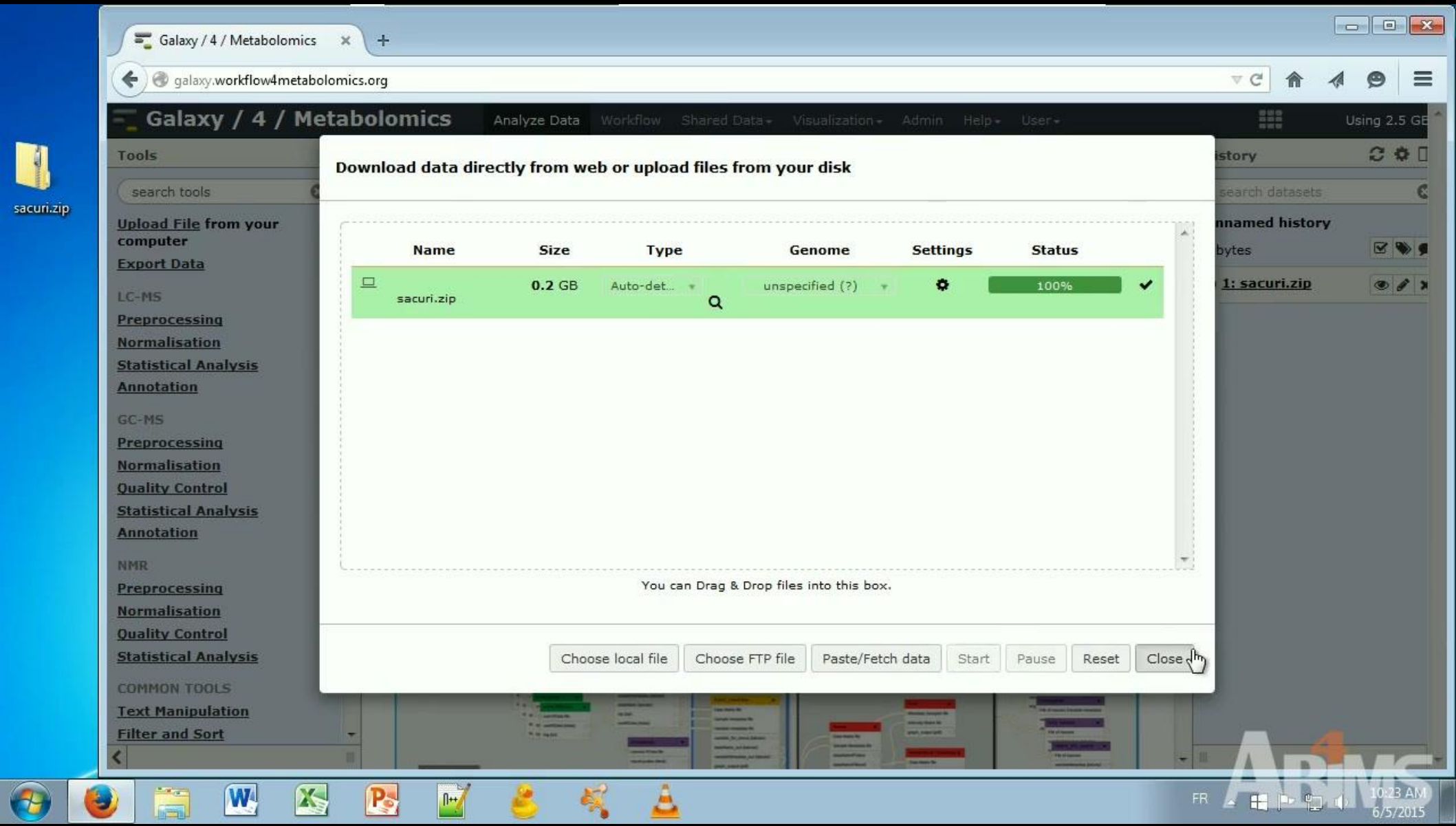
Data import < 2 Go

From local files



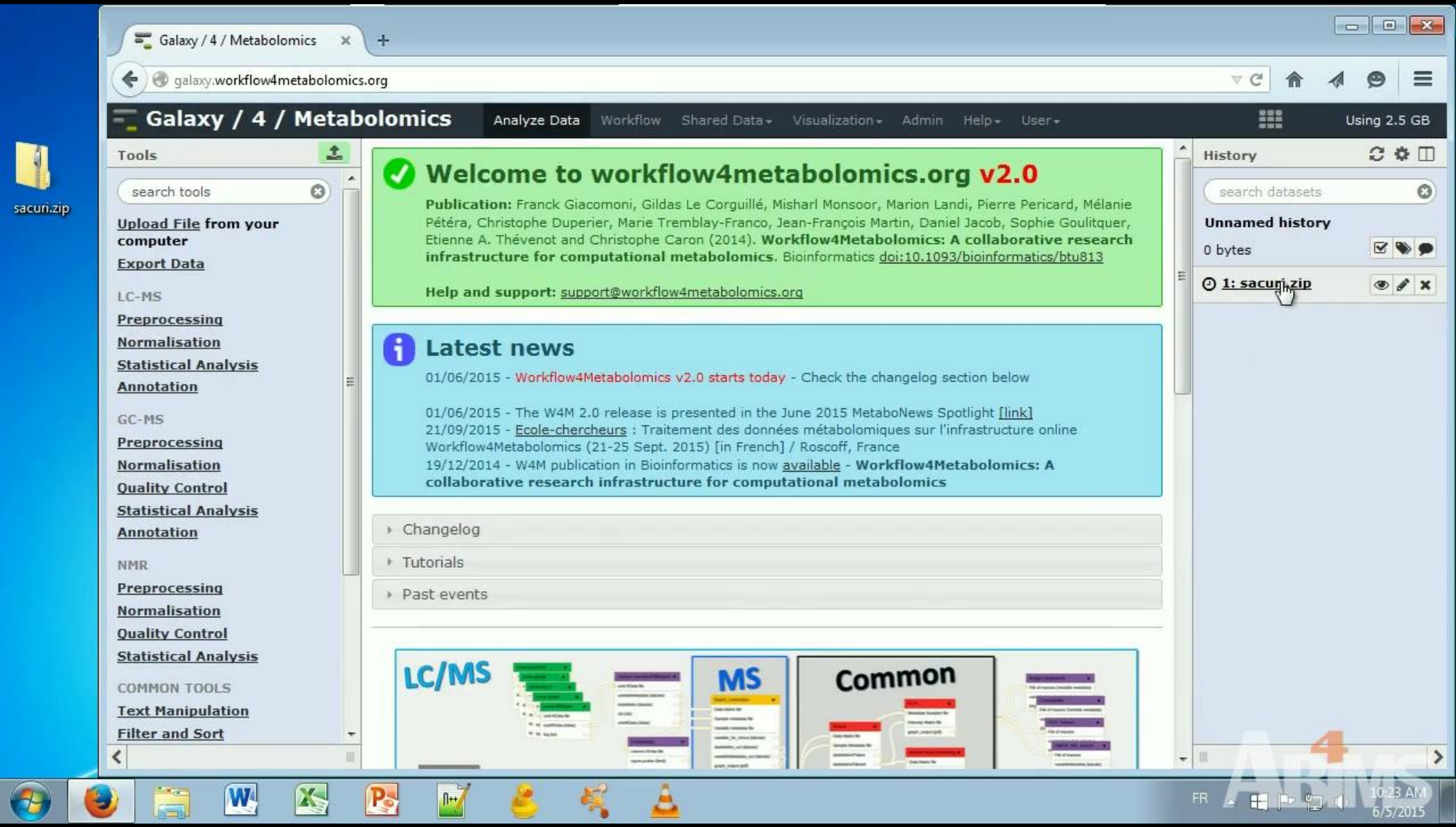
Data import < 2 Go

From local files



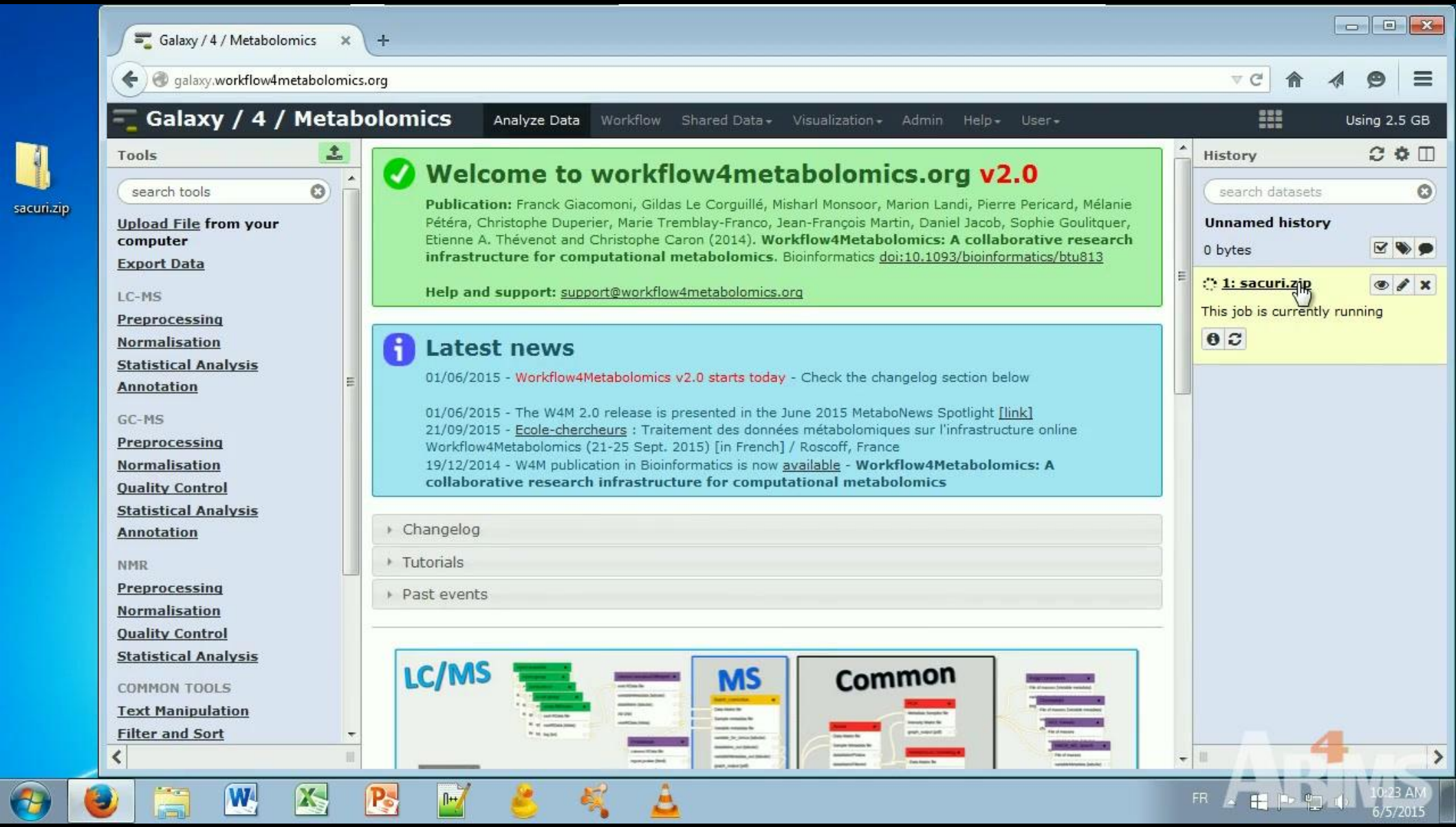
Data import < 2 Go

From local files



Data import < 2 Go

From local files



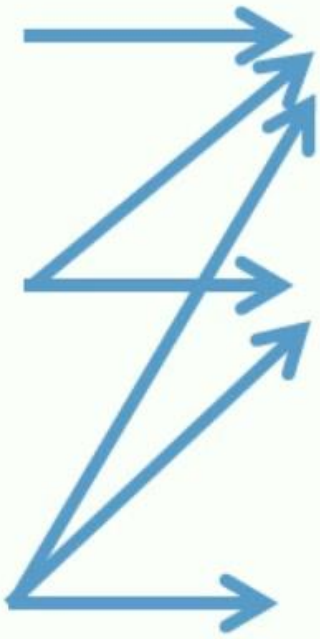
Step 1: Choose a FTP Client

DATA IMPORT

> 2 GO

STEP 1: CHOOSE A FTP CLIENT

STEP 1: CHOOSE A FTP CLIENT



Avoid:
Malwares inside

FileZilla

Cyberduck

WinSCP

STEP 1: CHOOSE A FTP CLIENT

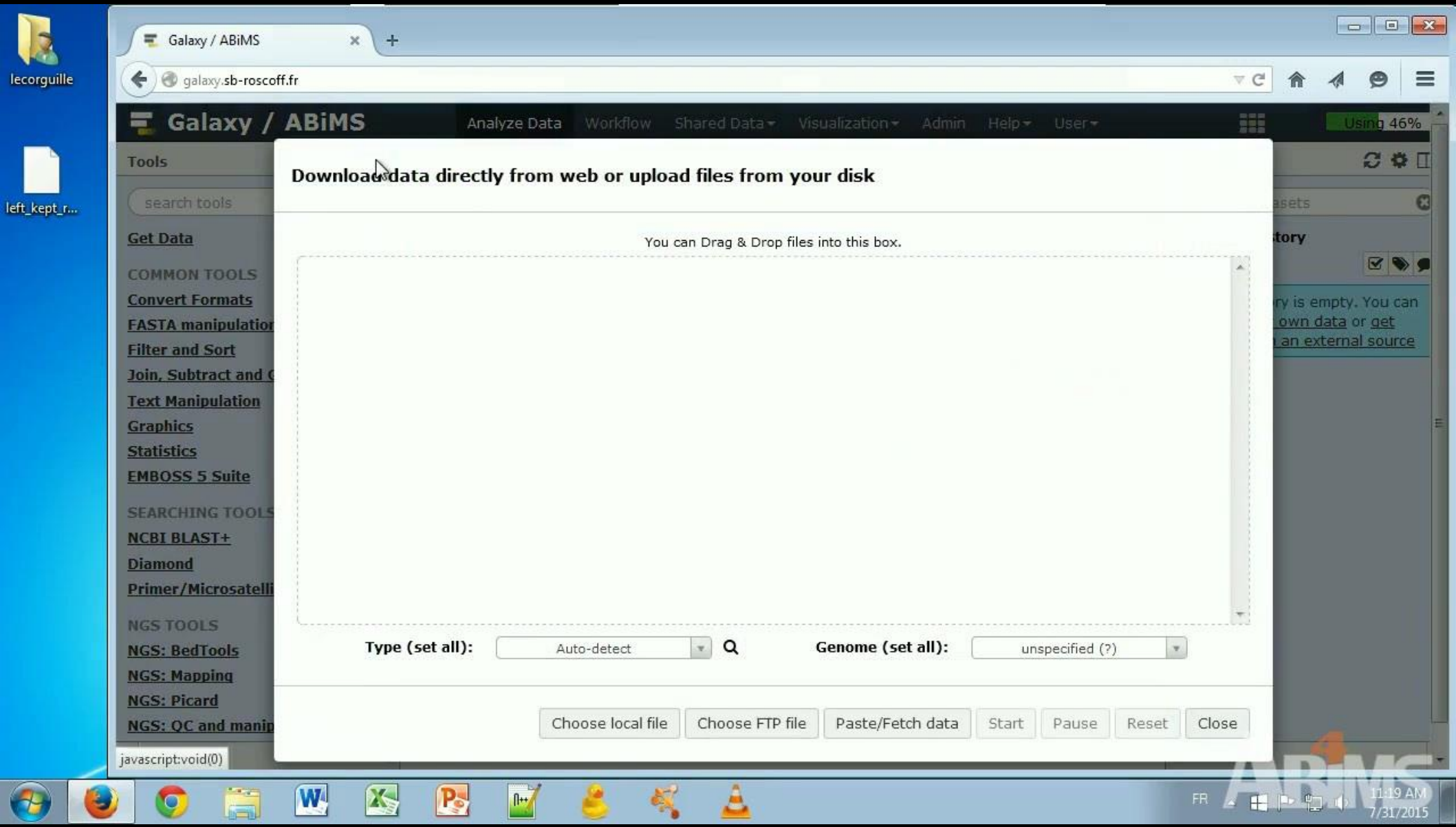
Step 2: Easy!

DATA IMPORT

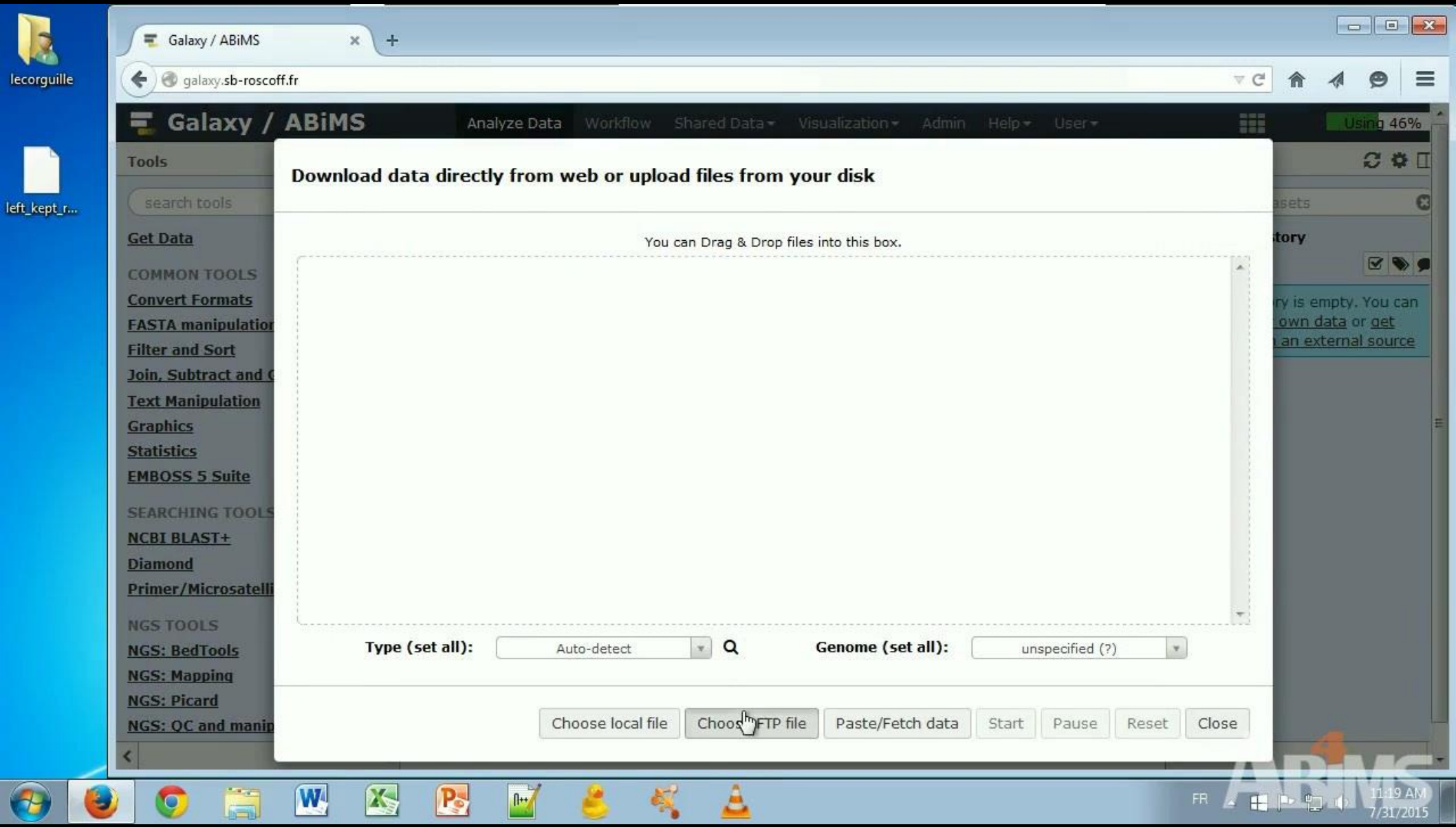
> 2 GO

Data import > 2 Go

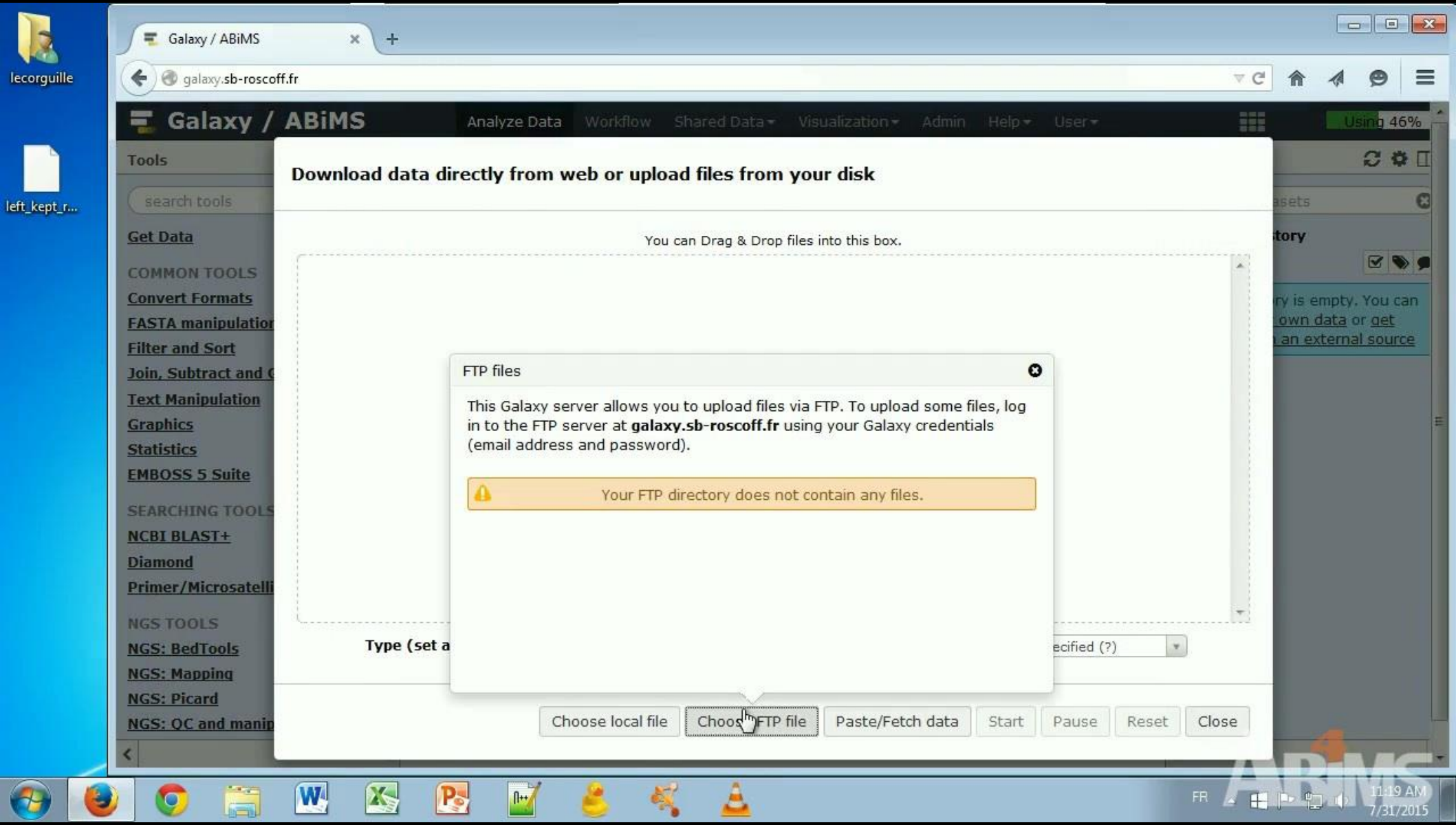
Data import > 2 Go



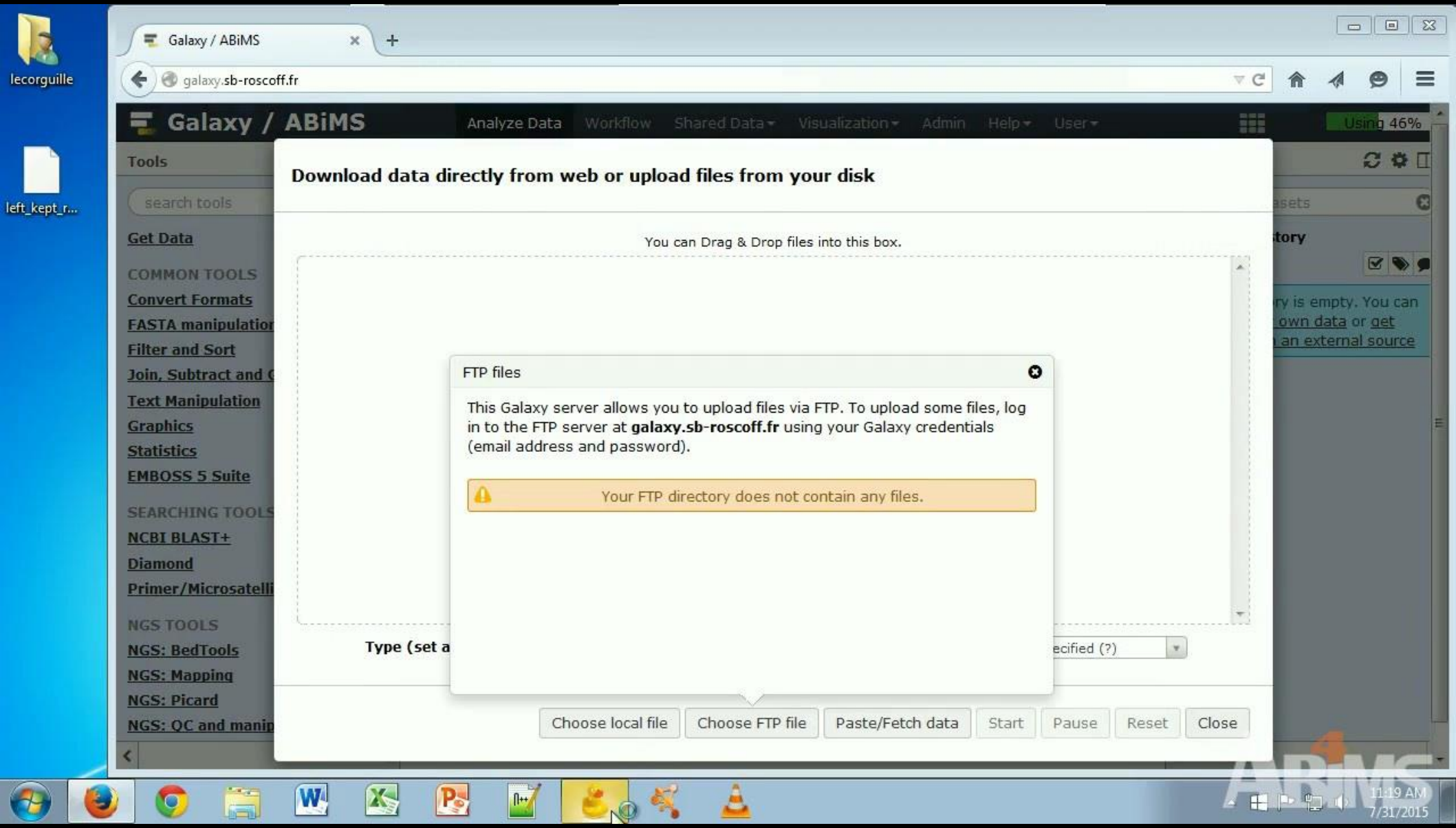
Data import > 2 Go



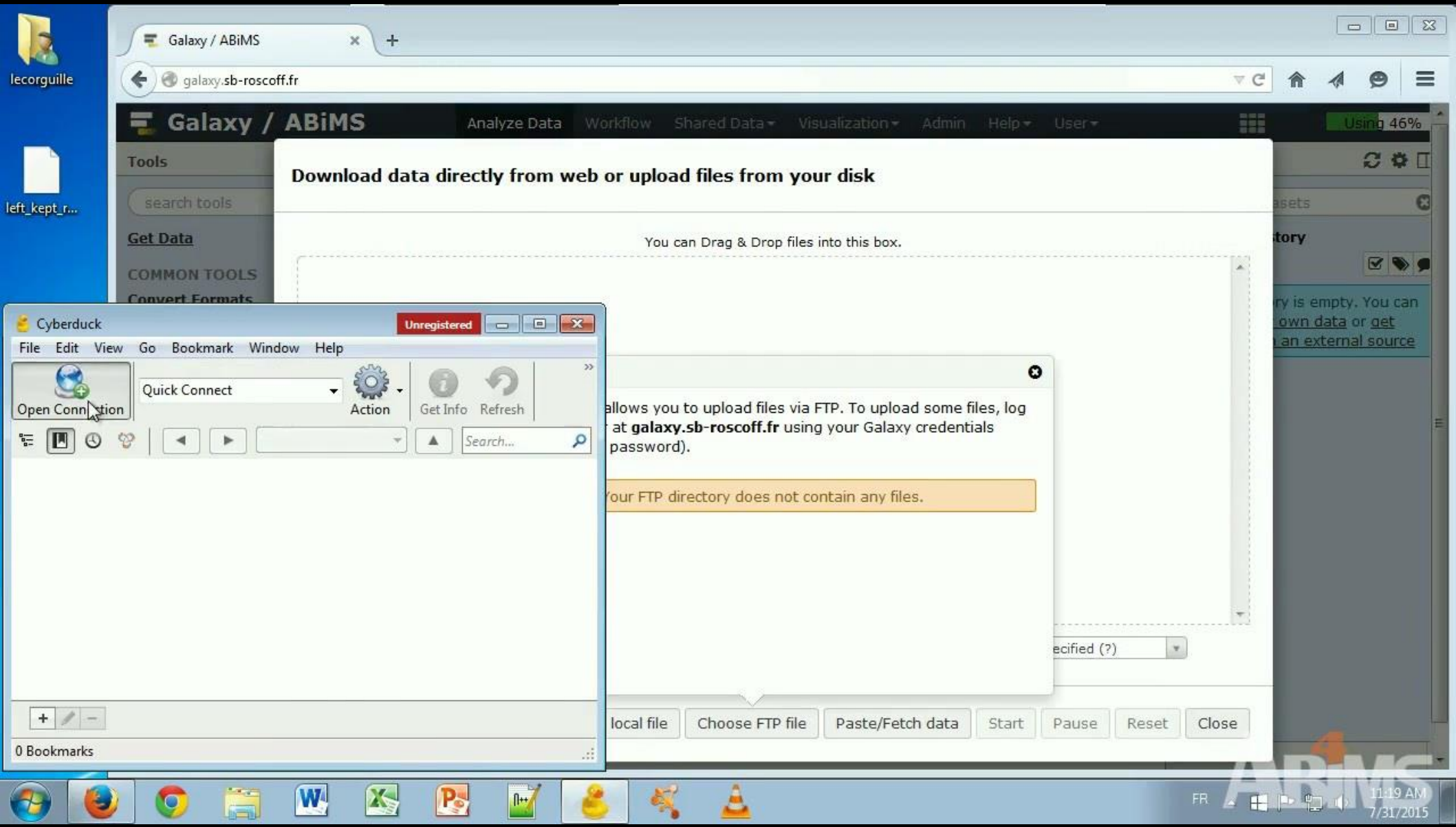
Data import > 2 Go



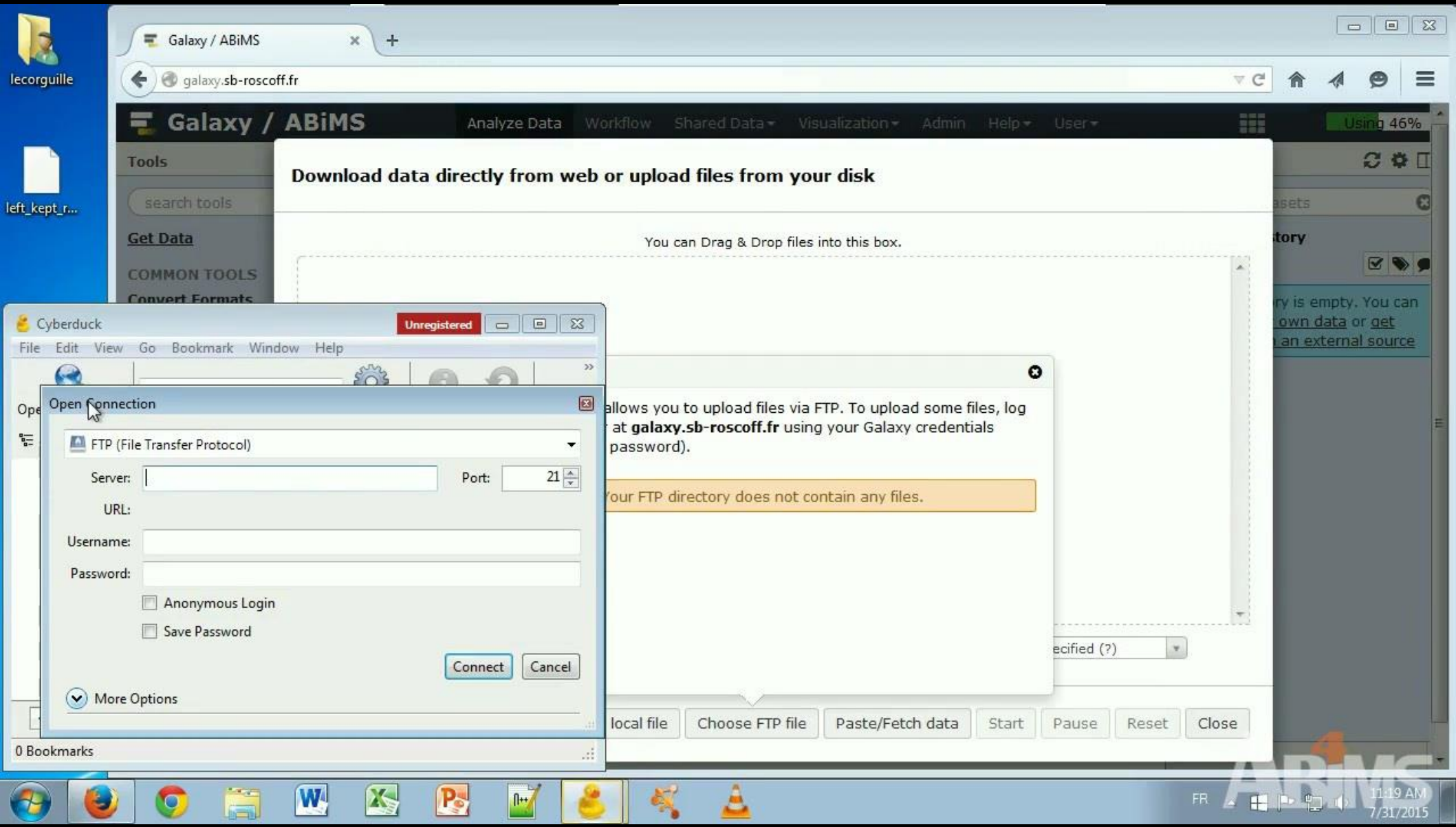
Data import > 2 Go



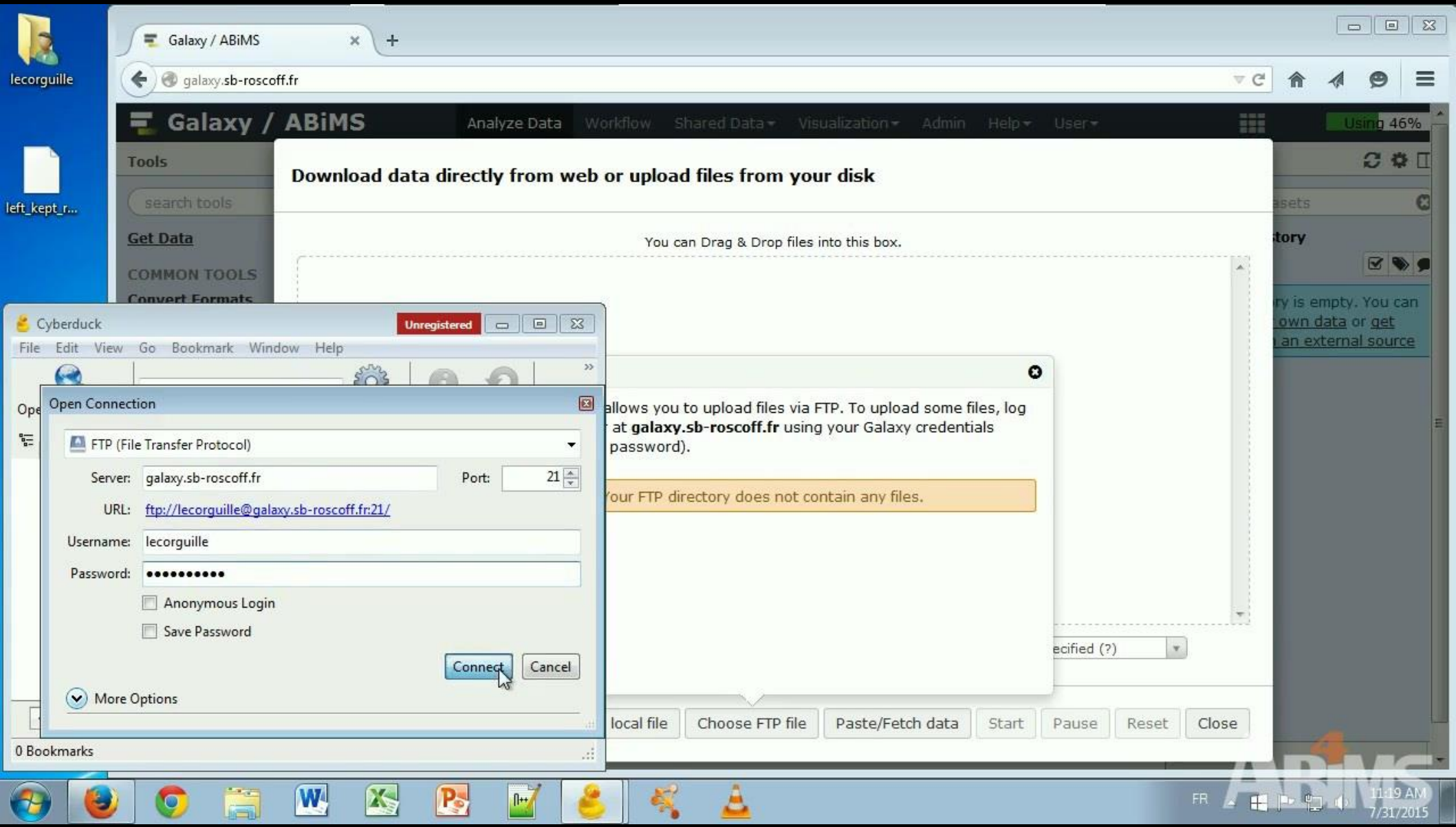
Data import > 2 Go



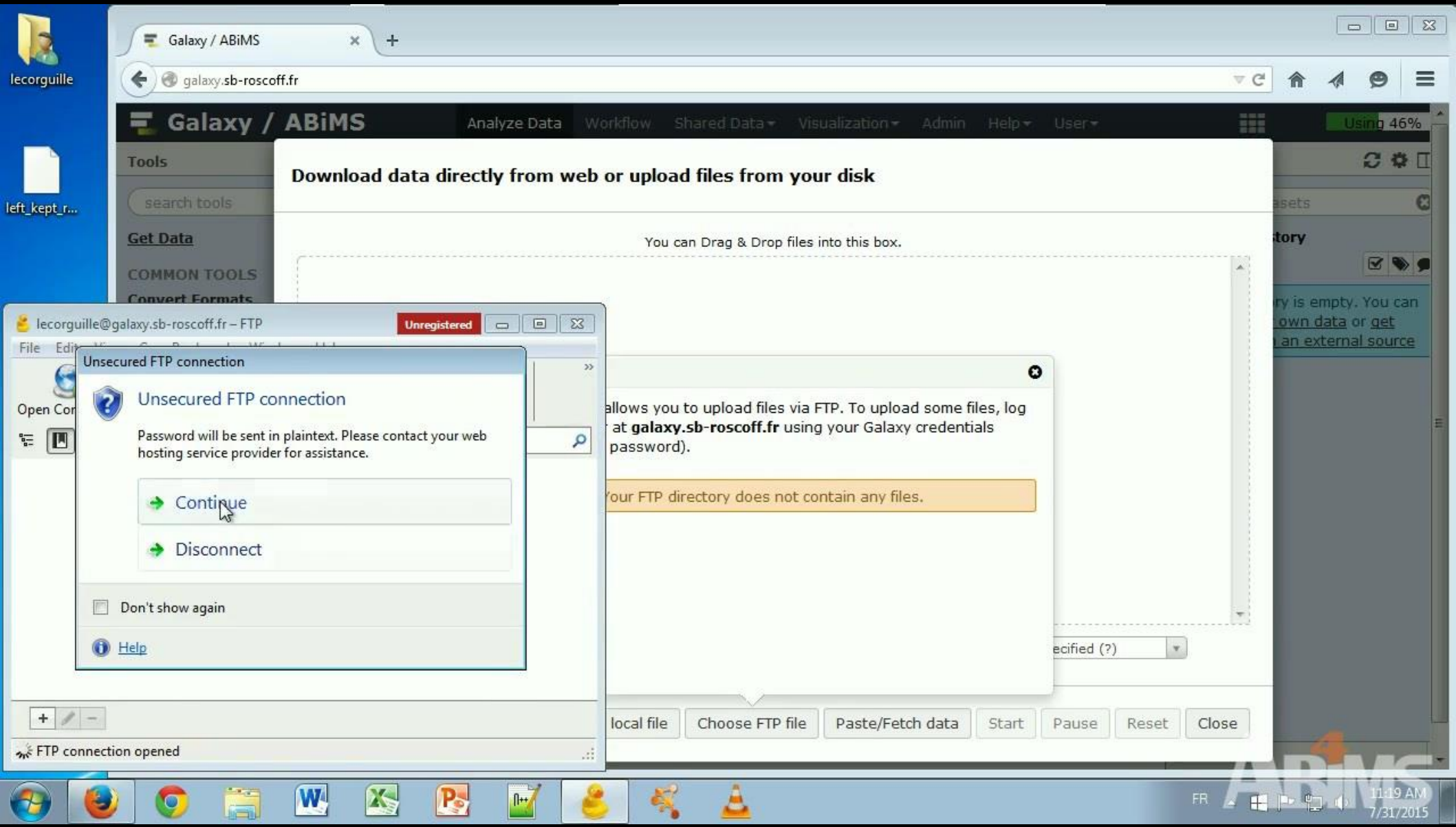
Data import > 2 Go



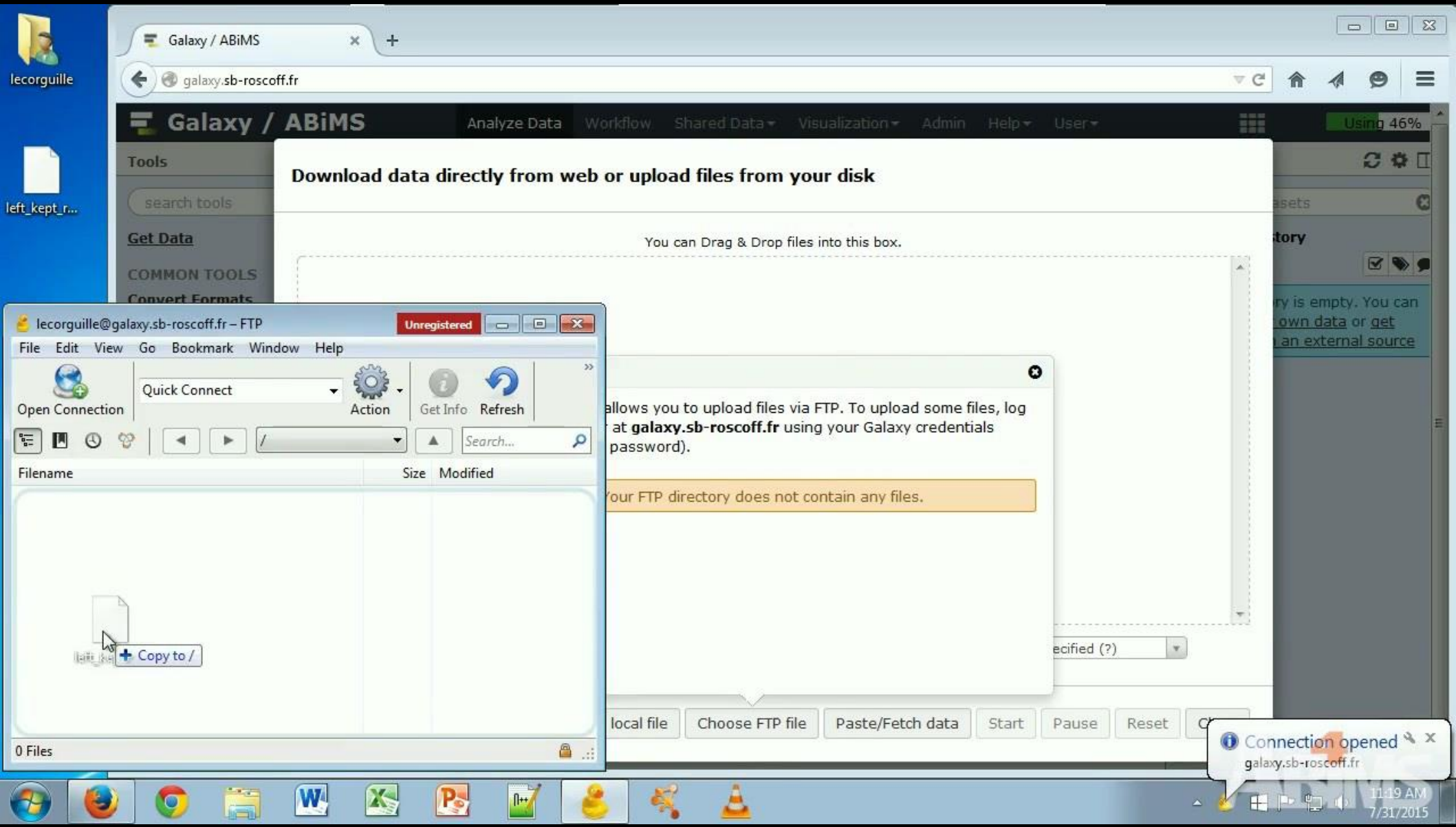
Data import > 2 Go



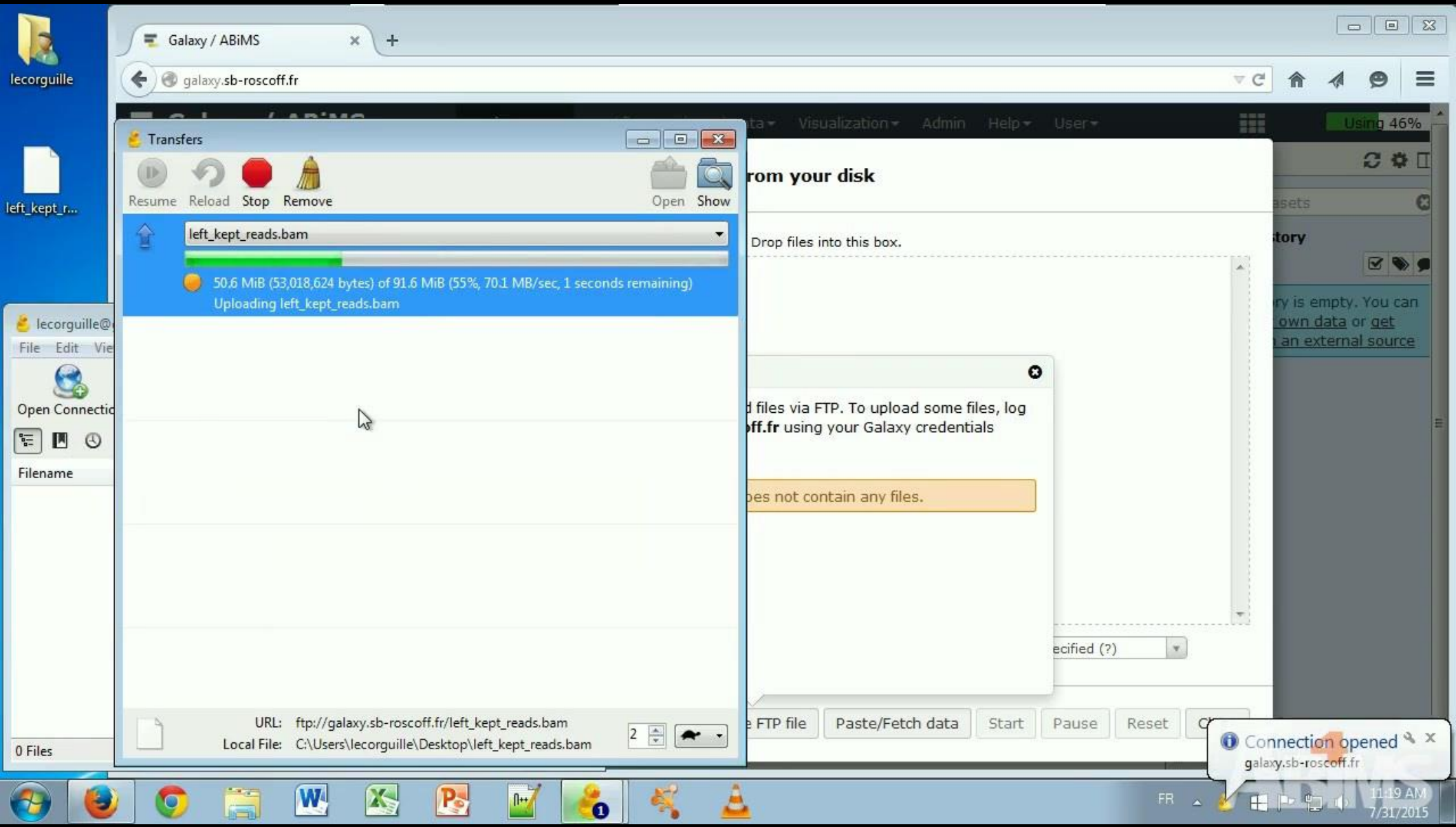
Data import > 2 Go



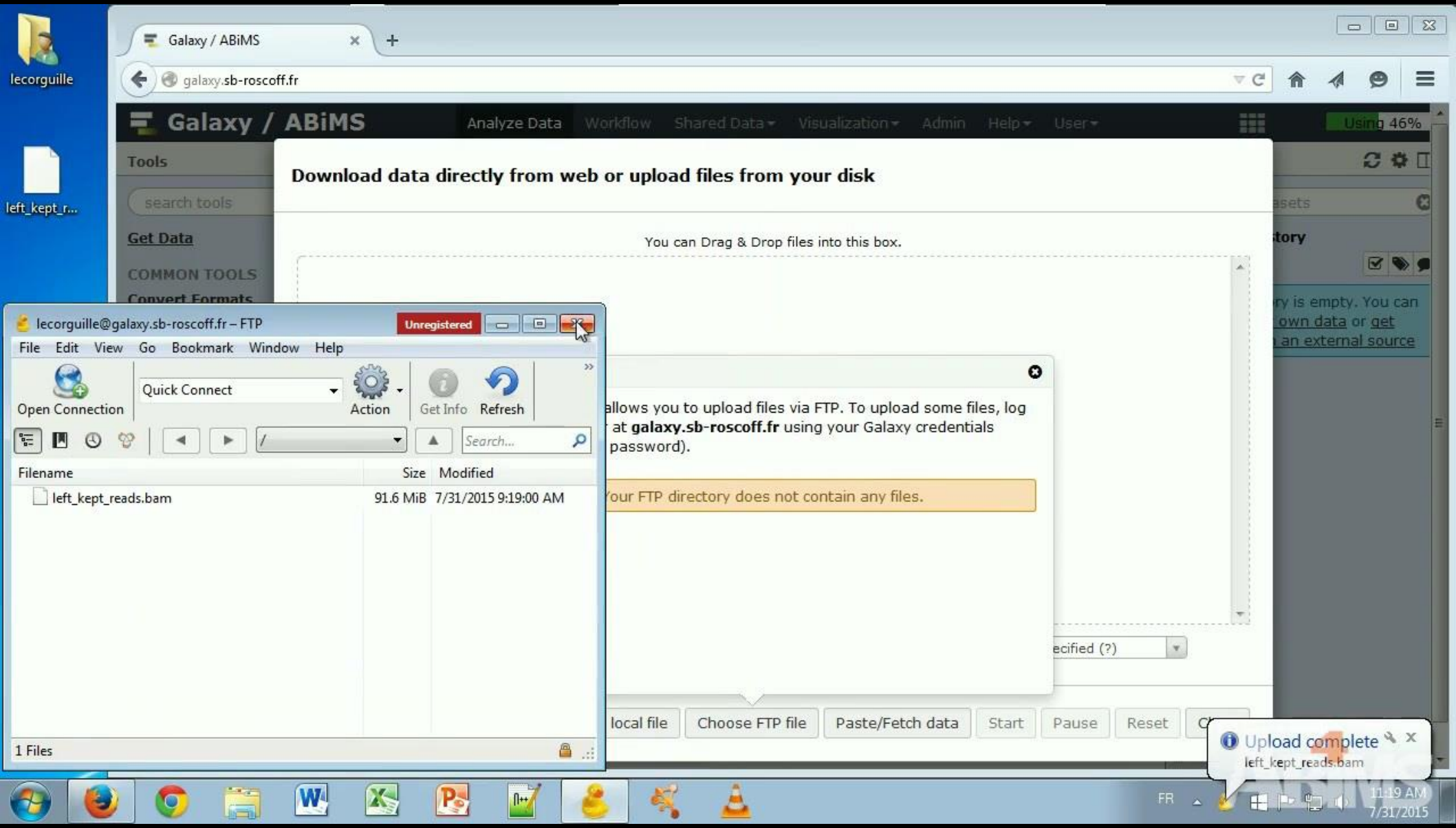
Data import > 2 Go



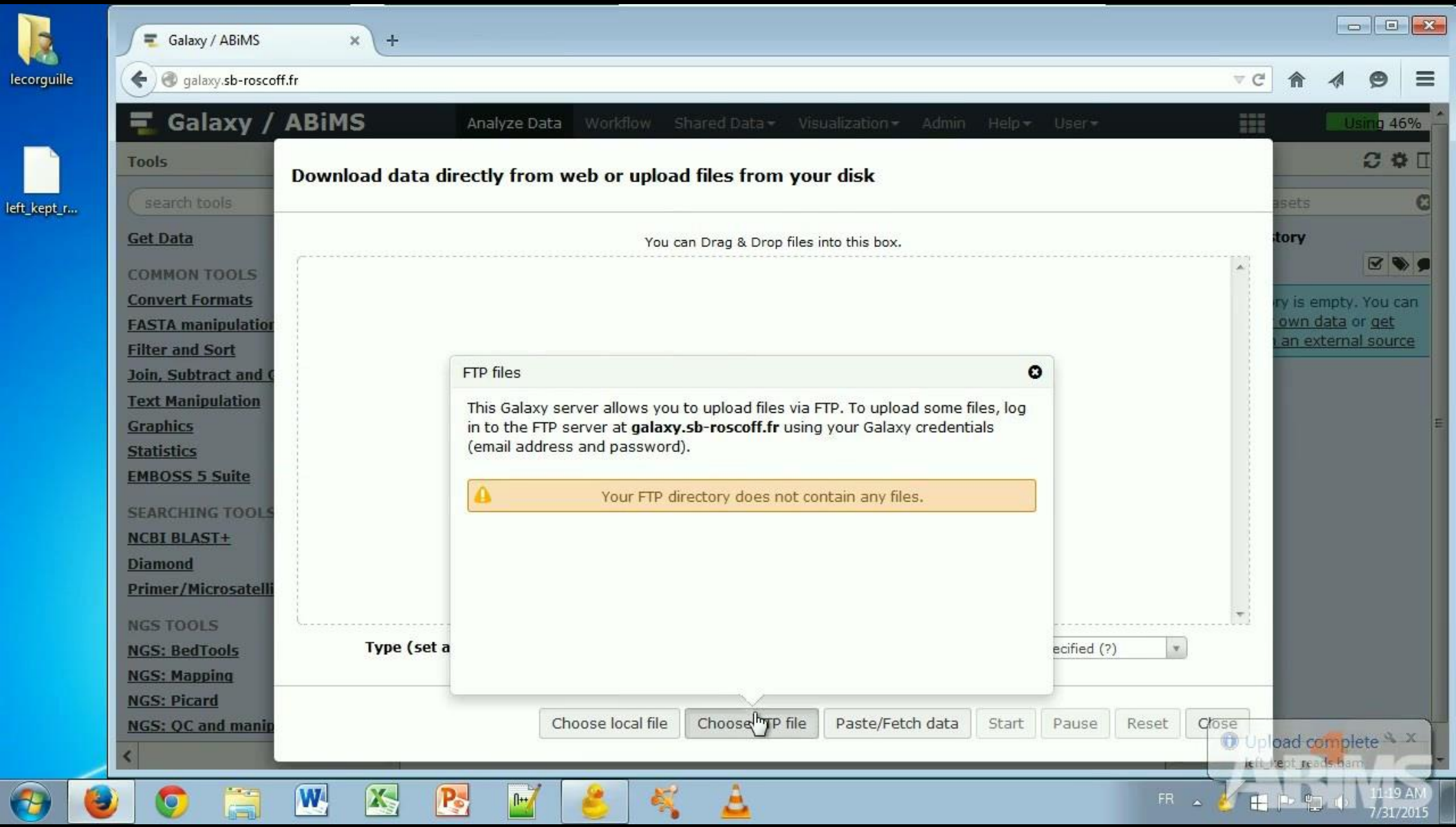
Data import > 2 Go



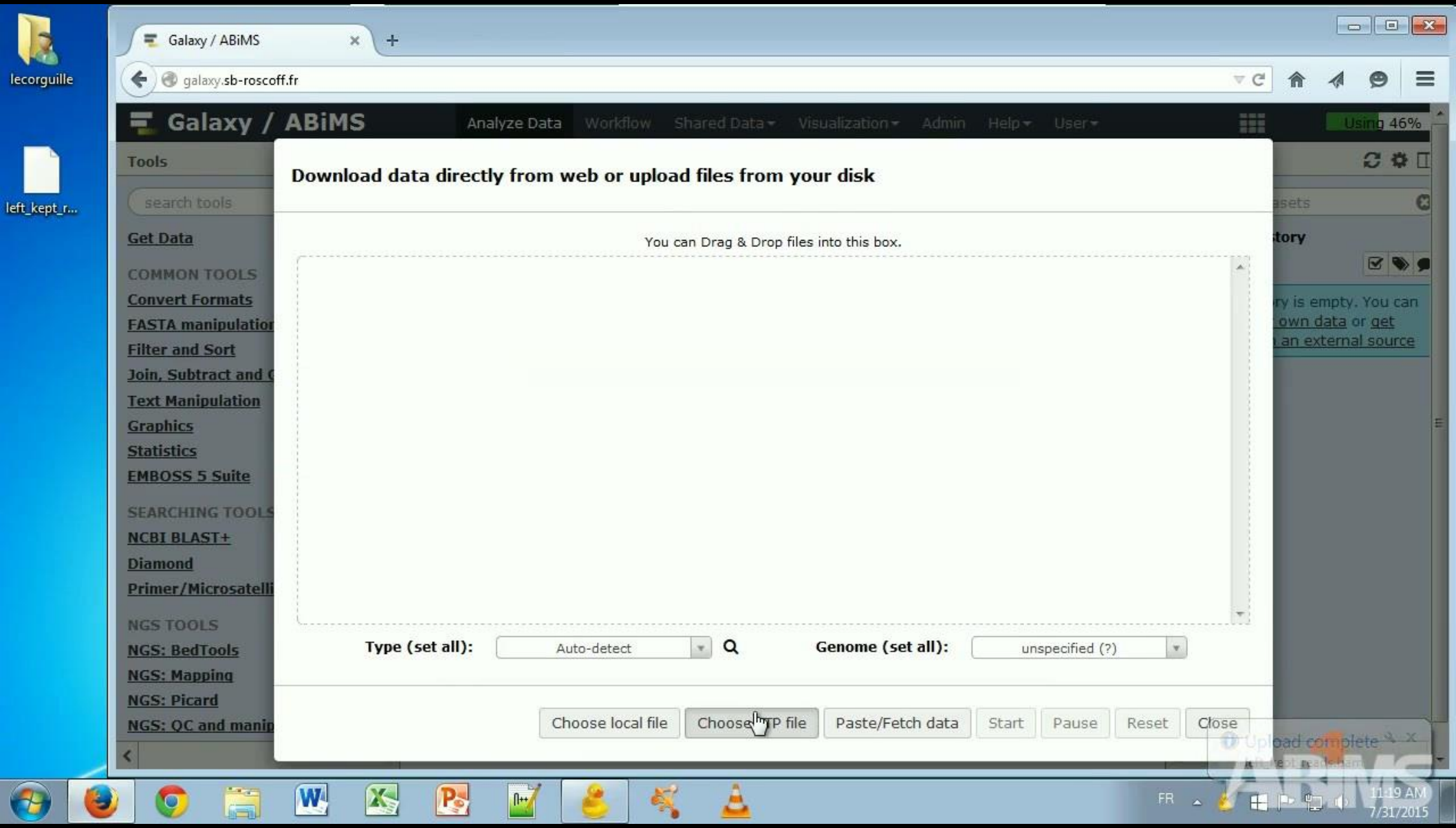
Data import > 2 Go



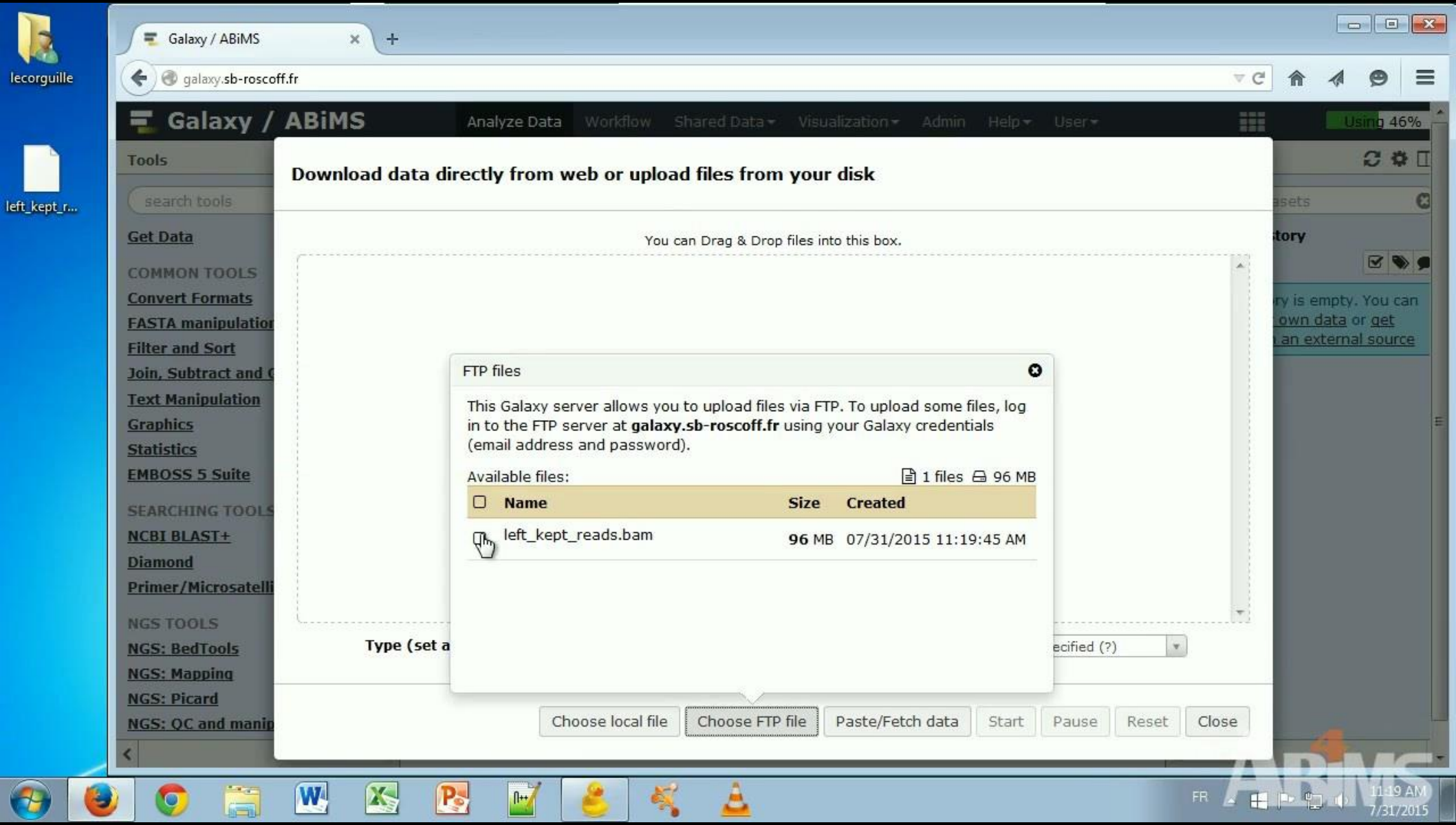
Data import > 2 Go



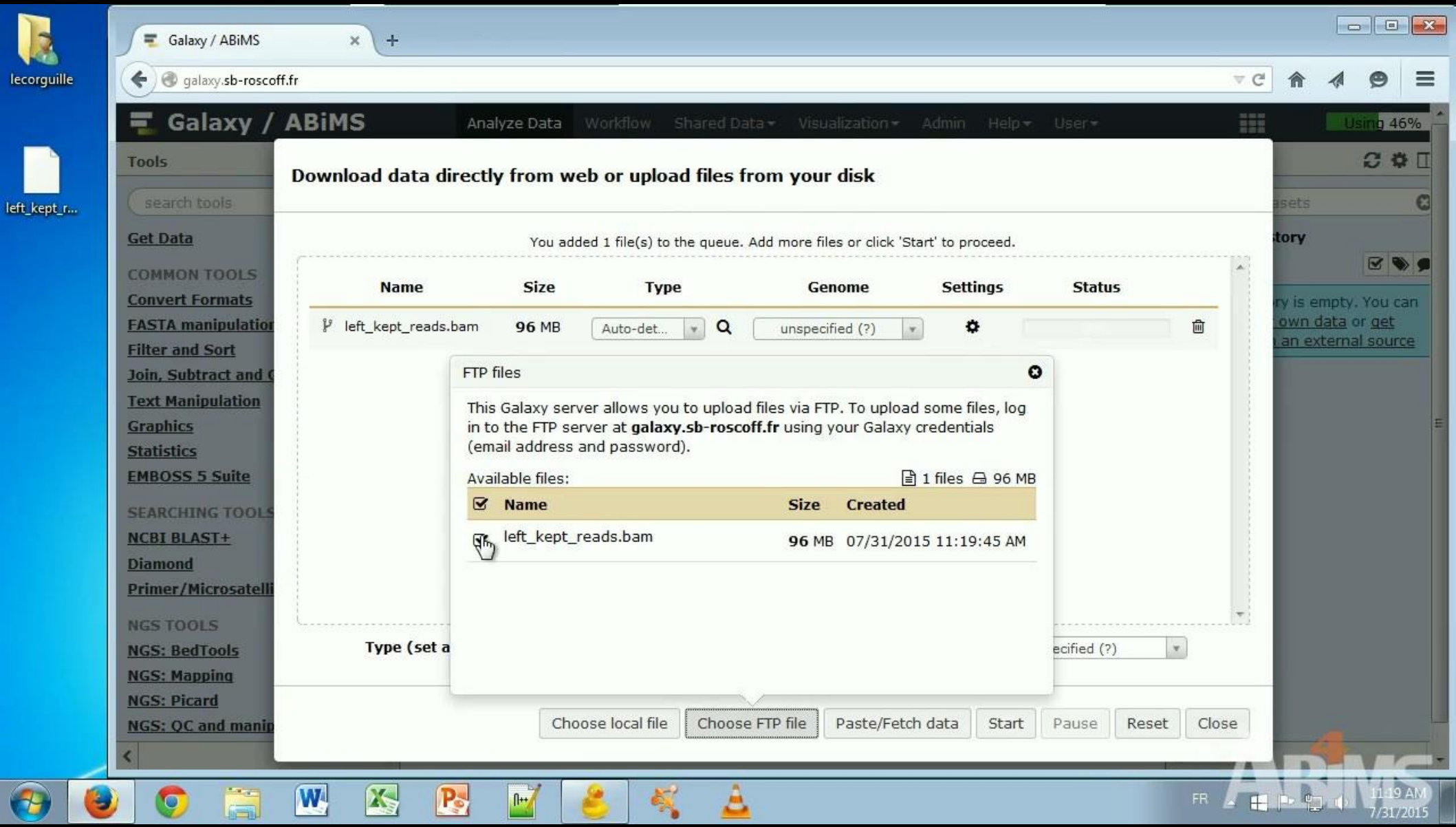
Data import > 2 Go



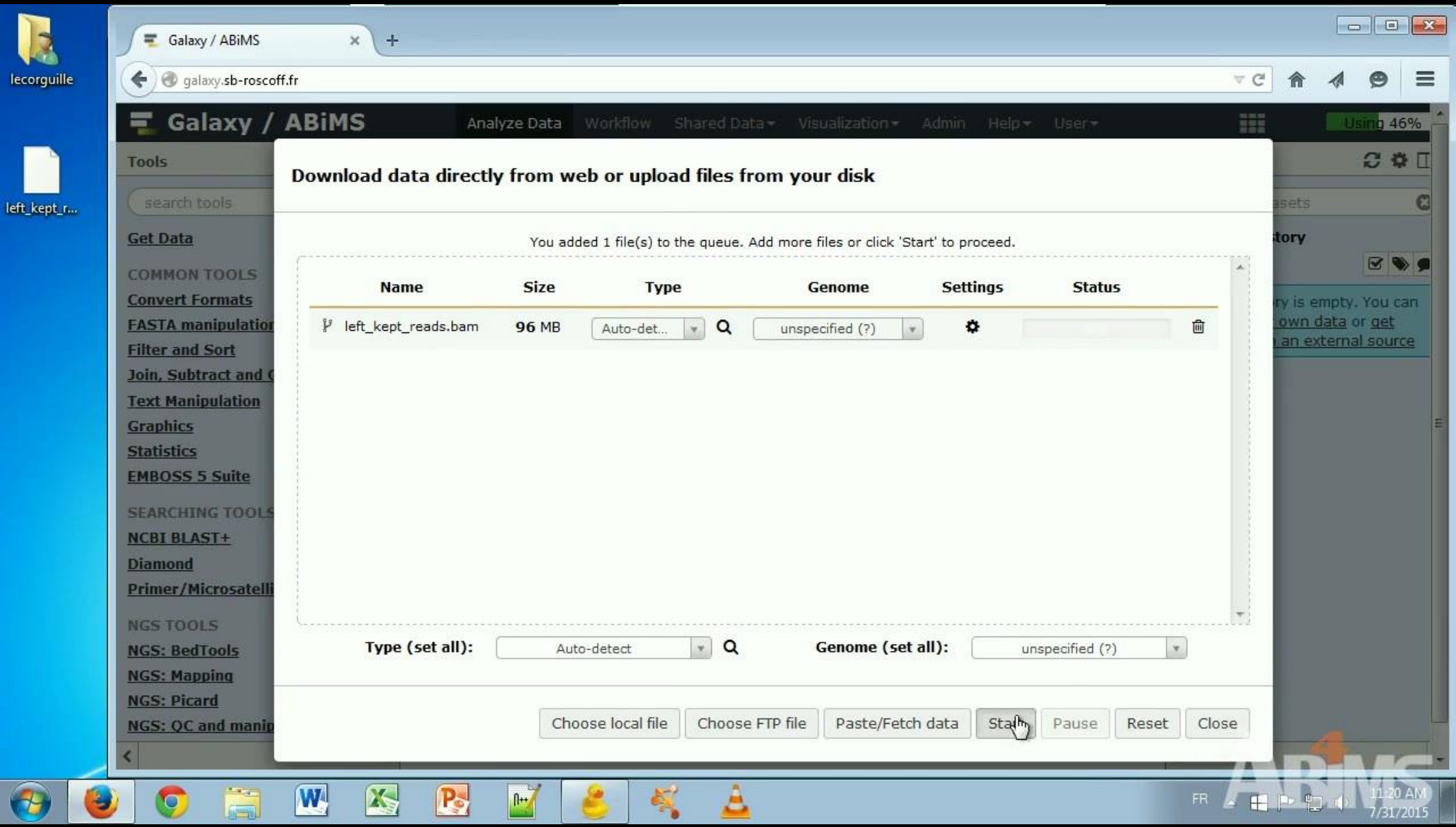
Data import > 2 Go



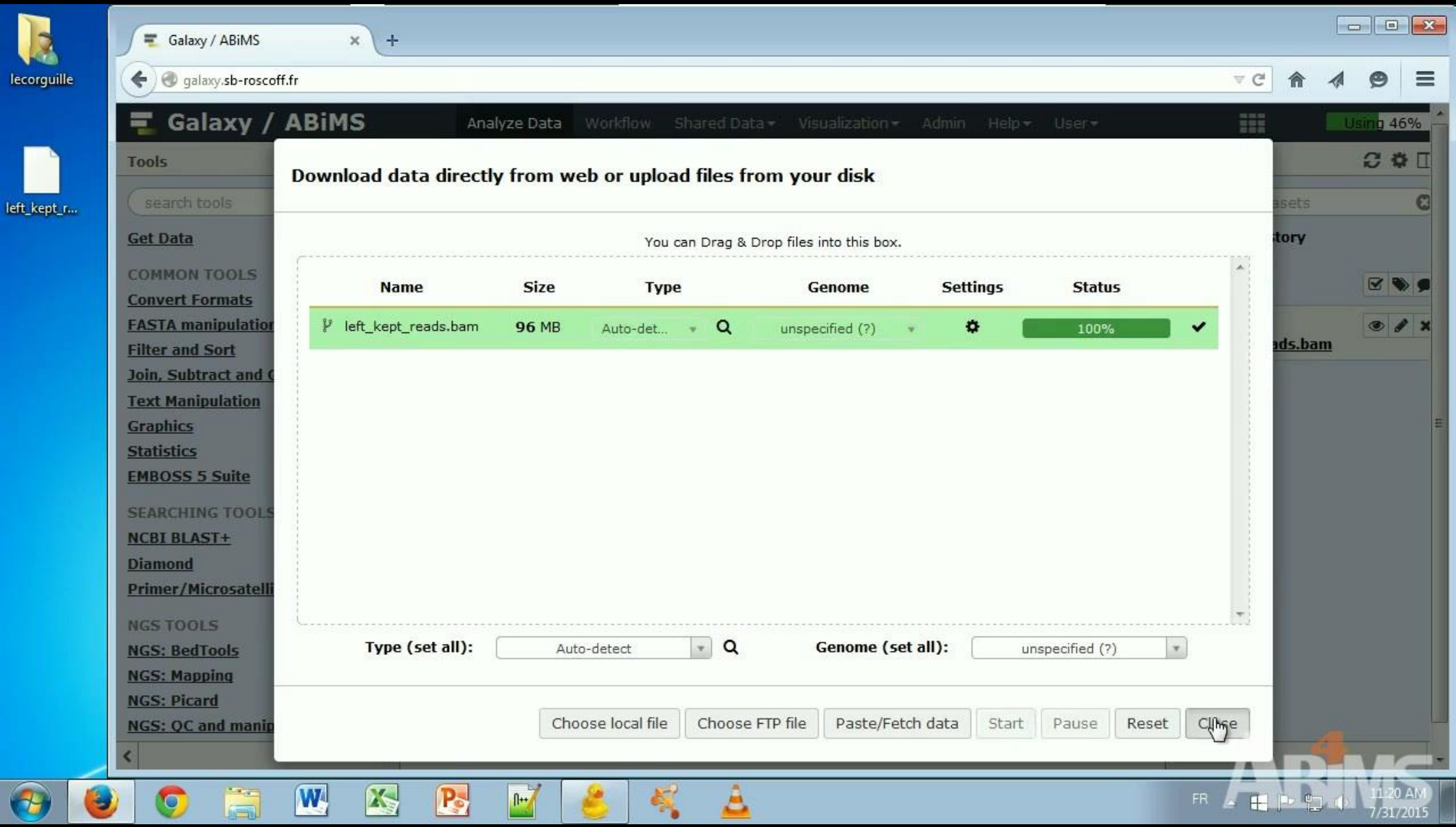
Data import > 2 Go



Data import > 2 Go



Data import > 2 Go



Data import > 2 Go

Data import > 2 Go

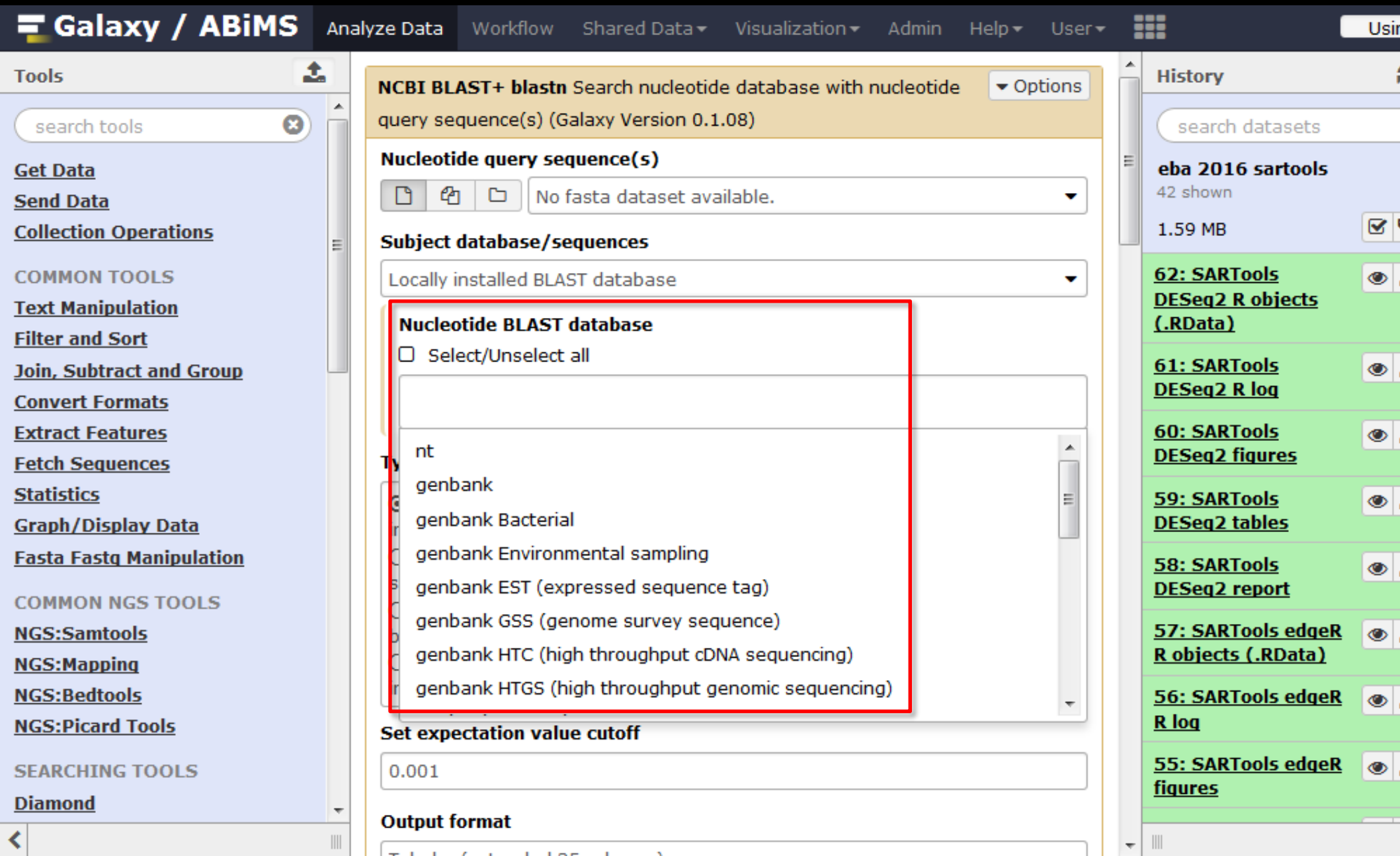
Data import > 2 Go

~~DATA IMPORT~~

~~Data import~~

For HUGE public resources: genome, databank …

--> Make a request to the support team



Hands-on

DATA IMPORT

Data import – Hands-on

1. Fetch the file with your internet browser (see given URL)

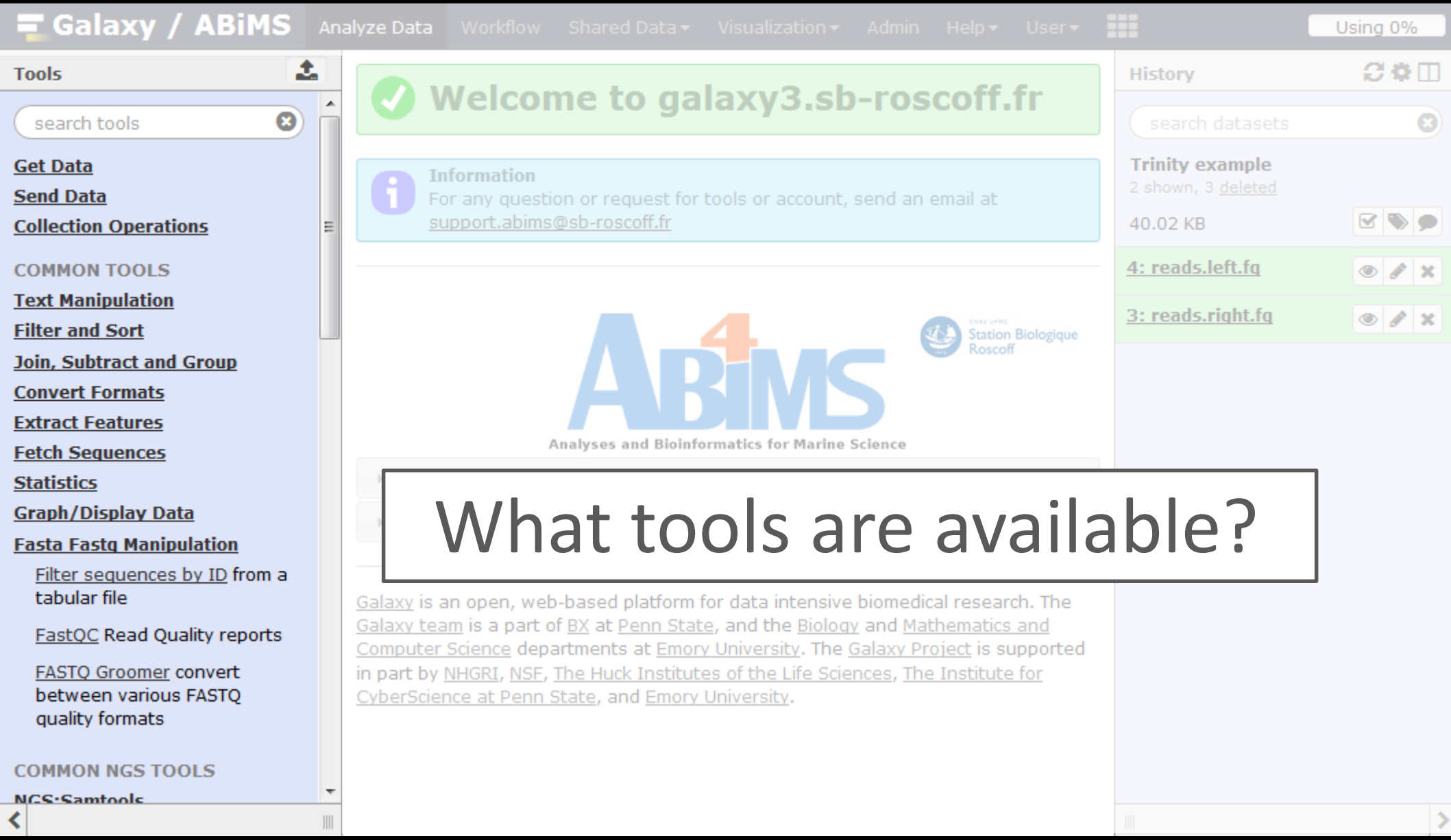
2. Upload this file into Galaxy
   a. First, as you want
   b. Consider that it is >2 Go

TOOLS

Tools - panel


Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 0%
Tools
search tools
Get Data
Send Data
Collection Operations
COMMON TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Statistics
Graph/Display Data
Fasta Fastq Manipulation
Filter sequences by ID from a tabular file
FastQC Read Quality reports
FASTQ Groomer convert between various FASTQ quality formats
COMMON NGS TOOLS
NGS:Samtools
Welcome to galaxy3.sb-roscoff.fr
Information
For any question or request for tools or account, send an email at support.abims@sb-roscoff.fr
Analyses and Bioinformatics for Marine Science
Station Biologique Roscoff
Changelog
Tutorials
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
Trinity example
2 shown, 3 deleted
40.02 KB
4: reads.left.fq
3: reads.right.fq

Tools - panel


Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 0%
Tools
search tools
Get Data
Send Data
Collection Operations
COMMON TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Statistics
Graph/Display Data
Fasta Fastq Manipulation
Filter sequences by ID from a tabular file
FastQC Read Quality reports
FASTQ Groomer convert between various FASTQ quality formats
COMMON NGS TOOLS
NGS:Samtools
Welcome to galaxy3.sb-roscoff.fr
Information
For any question or request for tools or account, send an email at support.abims@sb-roscoff.fr
Analyses and Bioinformatics for Marine Science
Station Biologique Roscoff
What tools are available?
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
Trinity example
2 shown, 3 deleted
40.02 KB
4: reads.left.fq
3: reads.right.fq

Tools - panel

>80 public Galaxy servers available:
https://galaxyproject.org/public-galaxy-servers

Tools - panel

>80 public Galaxy servers available:
https://galaxyproject.org/public-galaxy-servers

RNAseq: http://galaxy3.sb-roscoff.fr
SBR tools: http://webtools.sb-roscoff.fr
Metagenomics: http://galaxy4frogs.sb-roscoff.fr

Metabolomics:

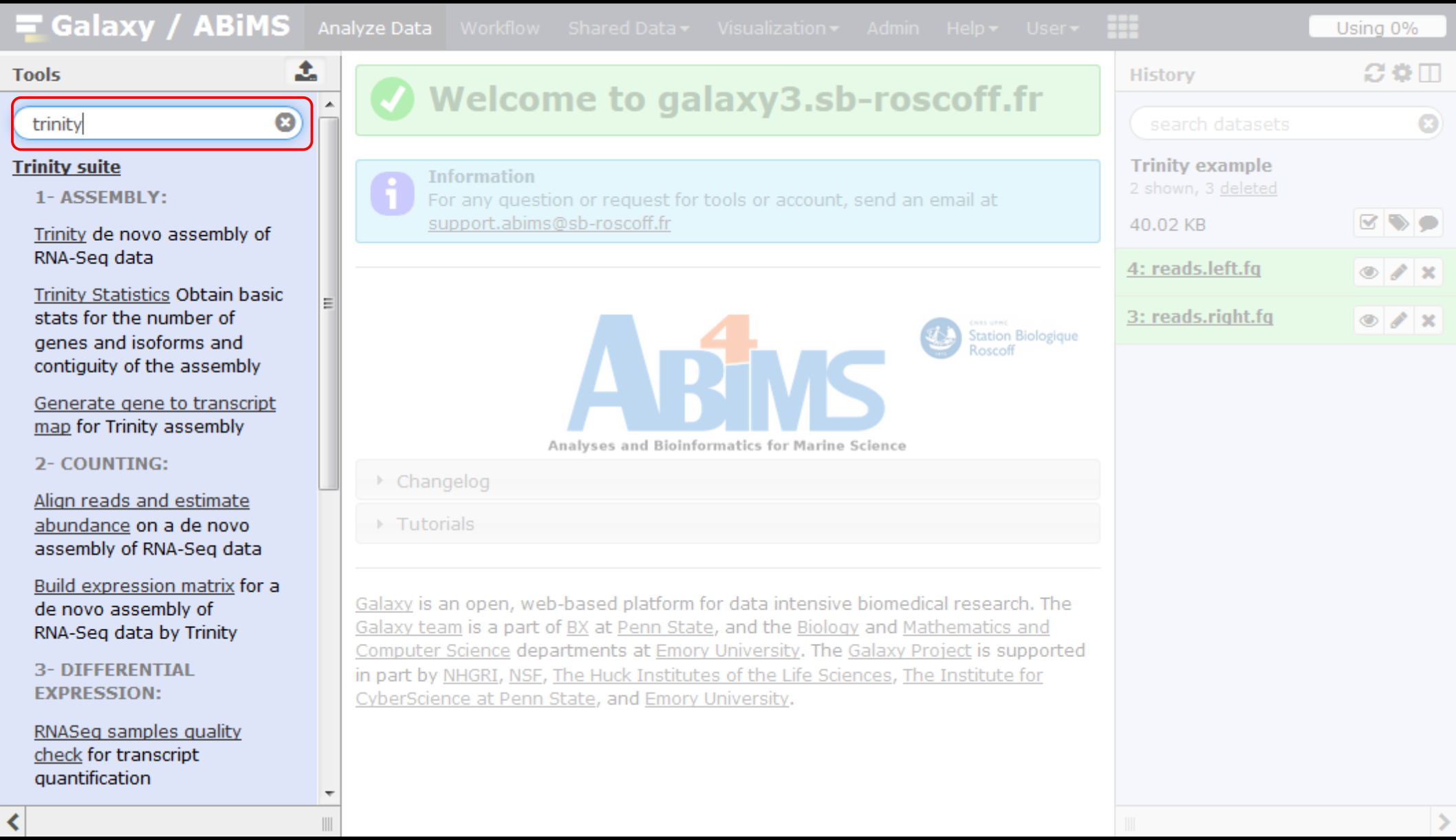
Workflow4metabolomics

ChIP-seq:

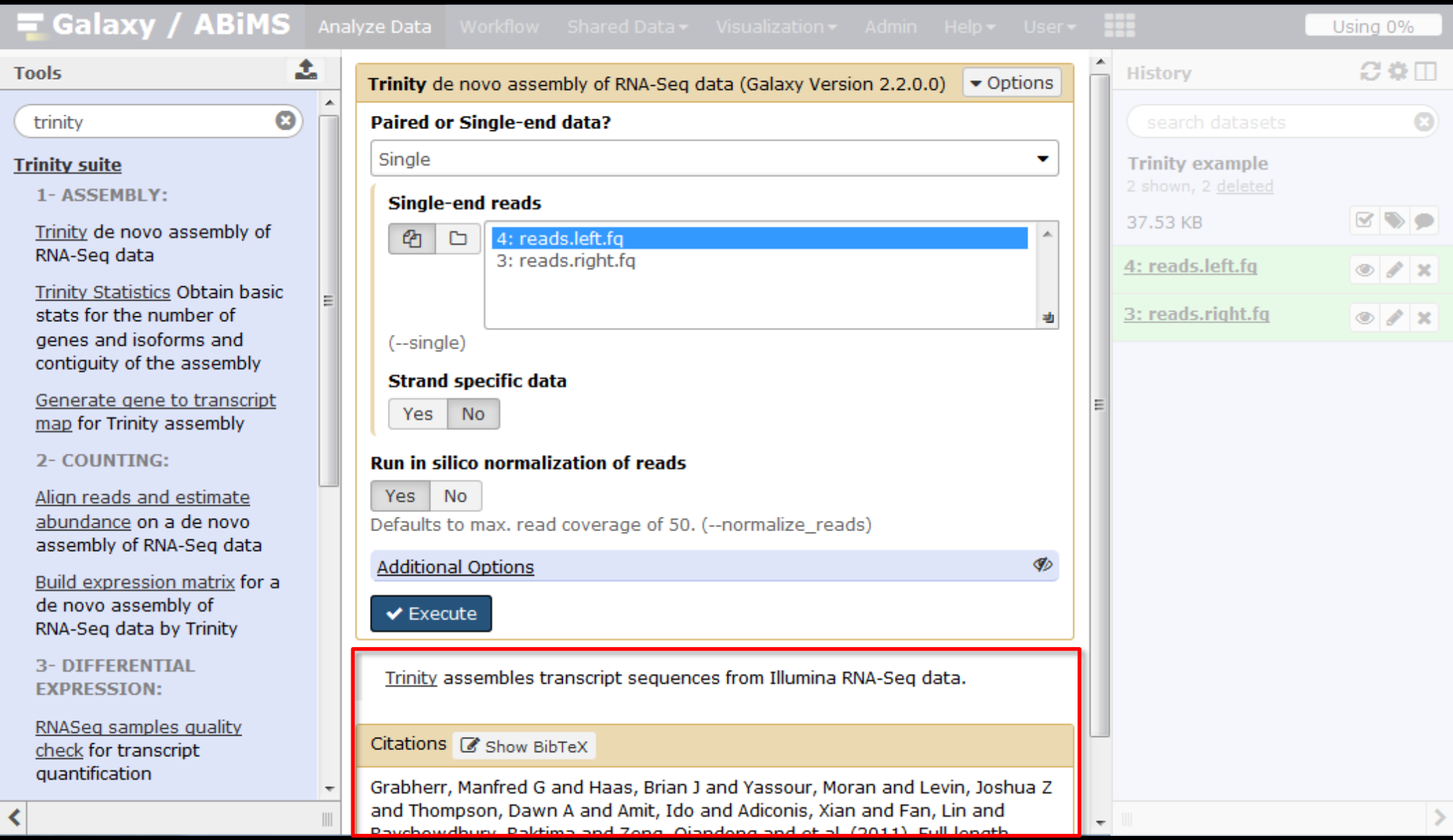
GalaxEast
an open and powerful Galaxy instance
for integrative Omics data analysis

nebula

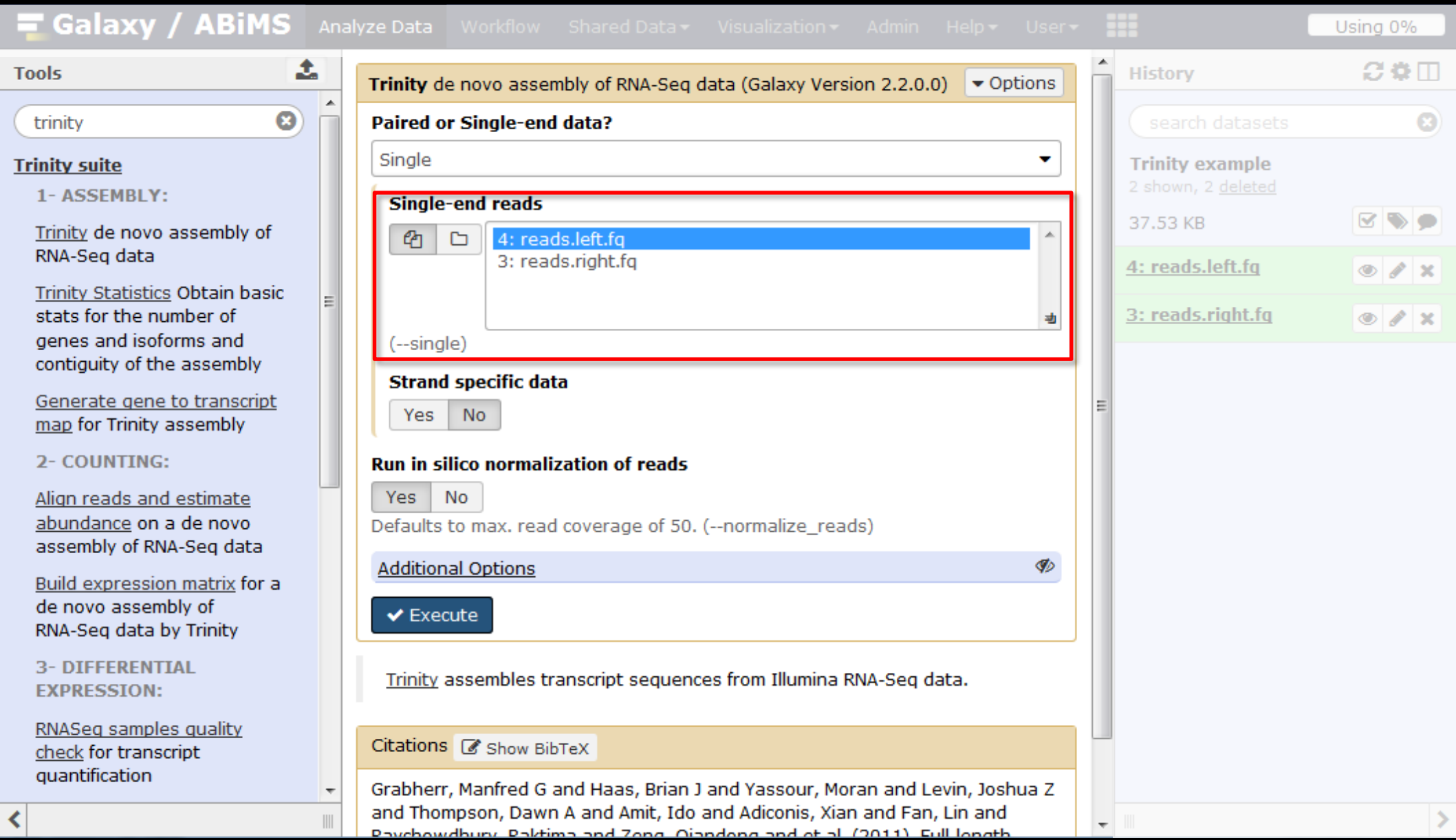
Tools - panel


Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 0%
Tools
trinity
Trinity suite
1- ASSEMBLY:
Trinity de novo assembly of RNA-Seq data
Trinity Statistics Obtain basic stats for the number of genes and isoforms and contiguity of the assembly
Generate gene to transcript map for Trinity assembly
2- COUNTING:
Align reads and estimate abundance on a de novo assembly of RNA-Seq data
Build expression matrix for a de novo assembly of RNA-Seq data by Trinity
3- DIFFERENTIAL EXPRESSION:
RNASeq samples quality check for transcript quantification
Welcome to galaxy3.sb-roscoff.fr
Information
For any question or request for tools or account, send an email at support.abims@sb-roscoff.fr
Analyses and Bioinformatics for Marine Science
Station Biologique Roscoff
Changelog
Tutorials
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
Trinity example
2 shown, 3 deleted
40.02 KB
4: reads.left.fq
3: reads.right.fq

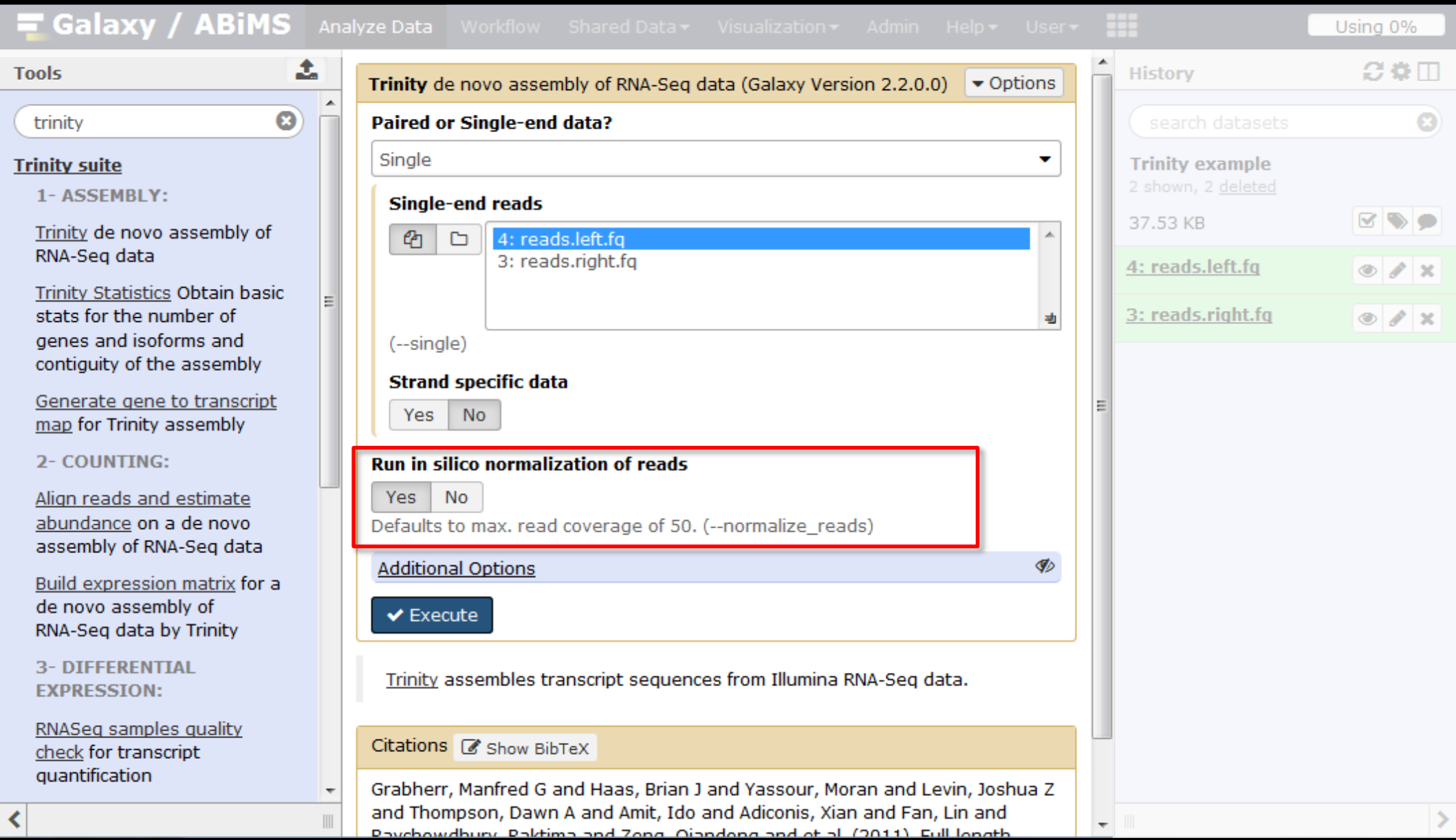
Tools - form


Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 0%
Tools
trinity
Trinity suite
1- ASSEMBLY:
Trinity de novo assembly of RNA-Seq data
Trinity Statistics Obtain basic stats for the number of genes and isoforms and contiguity of the assembly
Generate gene to transcript map for Trinity assembly
2- COUNTING:
Align reads and estimate abundance on a de novo assembly of RNA-Seq data
Build expression matrix for a de novo assembly of RNA-Seq data by Trinity
3- DIFFERENTIAL EXPRESSION:
RNASeq samples quality check for transcript quantification
Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.2.0.0)
Options
Paired or Single-end data?
Single
Single-end reads
4: reads.left.fq
3: reads.right.fq
(--single)
Strand specific data
Yes
No
Run in silico normalization of reads
Yes
No
Defaults to max. read coverage of 50. (--normalize_reads)
Additional Options
Execute
Trinity assembles transcript sequences from Illumina RNA-Seq data.
Citations
Show BibTeX
Grabherr, Manfred G and Haas, Brian J and Yassour, Moran and Levin, Joshua Z and Thompson, Dawn A and Amit, Ido and Adiconis, Xian and Fan, Lin and
History
search datasets
Trinity example
2 shown, 2 deleted
37.53 KB
4: reads.left.fq
3: reads.right.fq

Tools - form

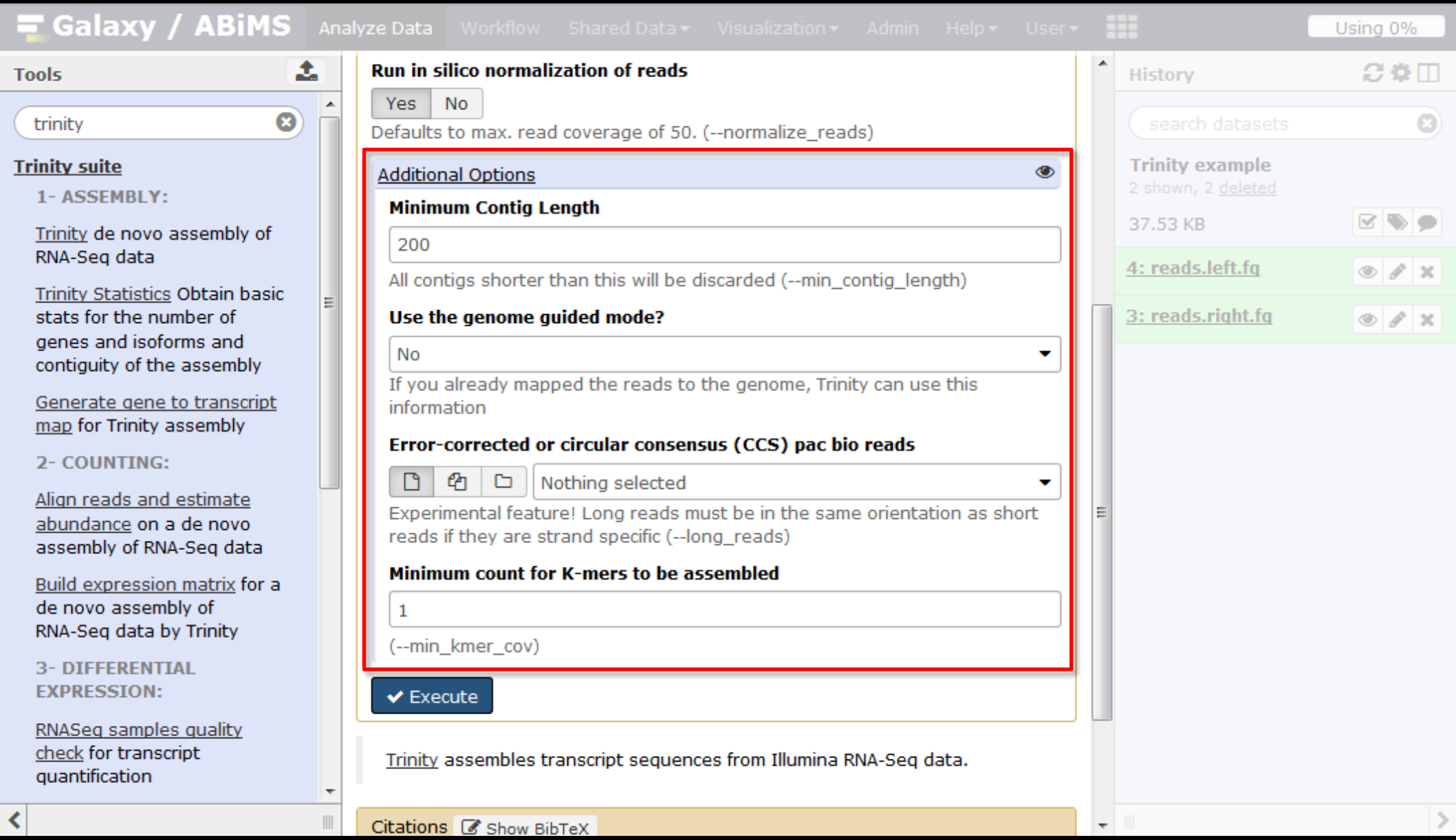

Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 0%
Tools
trinity
Trinity suite
1- ASSEMBLY:
Trinity de novo assembly of RNA-Seq data
Trinity Statistics Obtain basic stats for the number of genes and isoforms and contiguity of the assembly
Generate gene to transcript map for Trinity assembly
2- COUNTING:
Align reads and estimate abundance on a de novo assembly of RNA-Seq data
Build expression matrix for a de novo assembly of RNA-Seq data by Trinity
3- DIFFERENTIAL EXPRESSION:
RNASeq samples quality check for transcript quantification
Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.2.0.0)
Options
Paired or Single-end data?
Single
Single-end reads
4: reads.left.fq
3: reads.right.fq
(--single)
Strand specific data
Yes
No
Run in silico normalization of reads
Yes
No
Defaults to max. read coverage of 50. (--normalize_reads)
Additional Options
Execute
Trinity assembles transcript sequences from Illumina RNA-Seq data.
Citations
Show BibTeX
Grabherr, Manfred G and Haas, Brian J and Yassour, Moran and Levin, Joshua Z and Thompson, Dawn A and Amit, Ido and Adiconis, Xian and Fan, Lin and
History
search datasets
Trinity example
2 shown, 2 deleted
37.53 KB
4: reads.left.fq
3: reads.right.fq

Tools - form

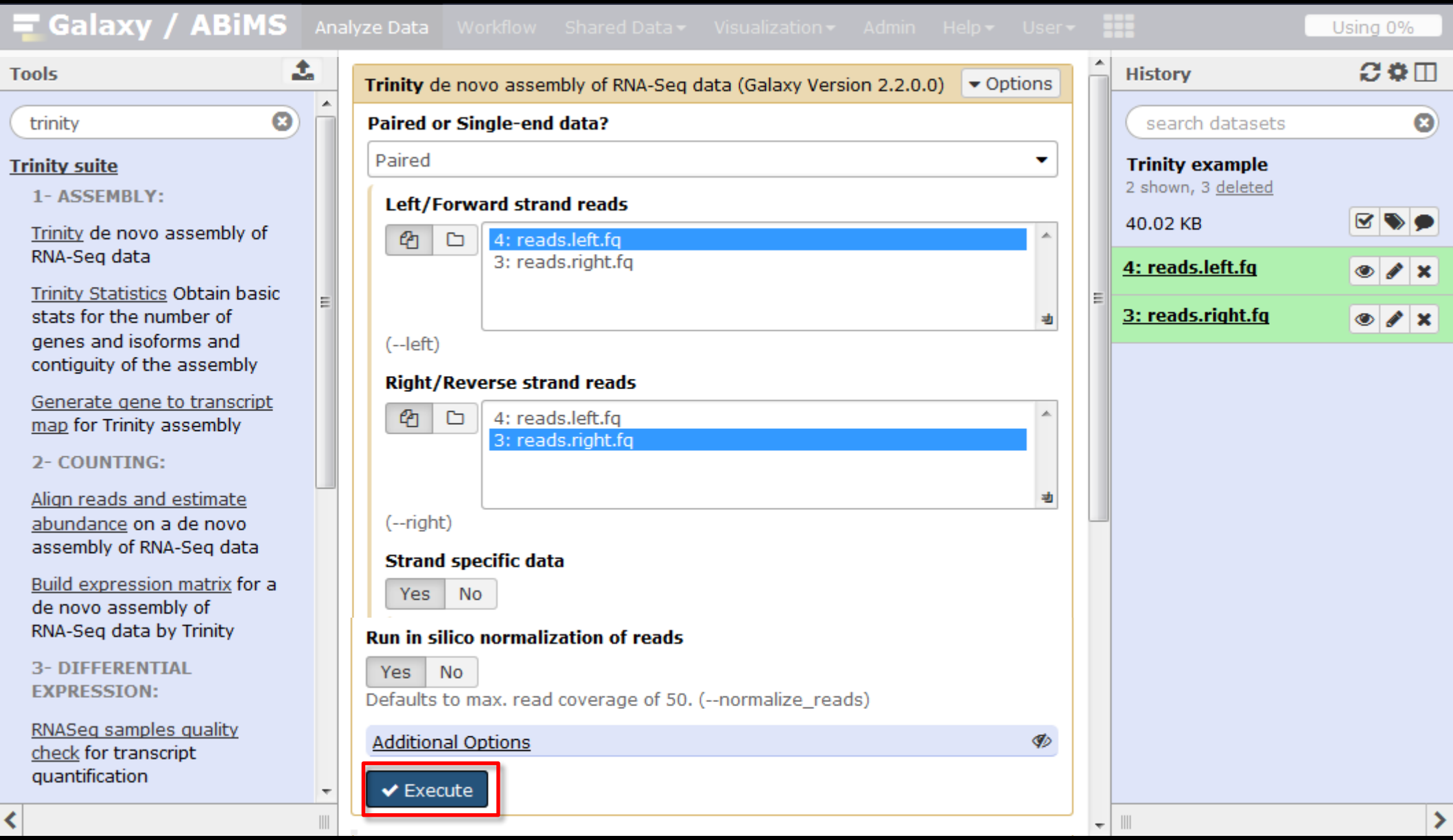

Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 0%
Tools
trinity
Trinity suite
1- ASSEMBLY:
Trinity de novo assembly of RNA-Seq data
Trinity Statistics Obtain basic stats for the number of genes and isoforms and contiguity of the assembly
Generate gene to transcript map for Trinity assembly
2- COUNTING:
Align reads and estimate abundance on a de novo assembly of RNA-Seq data
Build expression matrix for a de novo assembly of RNA-Seq data by Trinity
3- DIFFERENTIAL EXPRESSION:
RNASeq samples quality check for transcript quantification
Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.2.0.0)
Options
Paired or Single-end data?
Single
Single-end reads
4: reads.left.fq
3: reads.right.fq
(--single)
Strand specific data
Yes
No
Run in silico normalization of reads
Yes
No
Defaults to max. read coverage of 50. (--normalize_reads)
Additional Options
Execute
Trinity assembles transcript sequences from Illumina RNA-Seq data.
Citations
Show BibTeX
Grabherr, Manfred G and Haas, Brian J and Yassour, Moran and Levin, Joshua Z and Thompson, Dawn A and Amit, Ido and Adiconis, Xian and Fan, Lin and
History
search datasets
Trinity example
2 shown, 2 deleted
37.53 KB
4: reads.left.fq
3: reads.right.fq

Tools - form

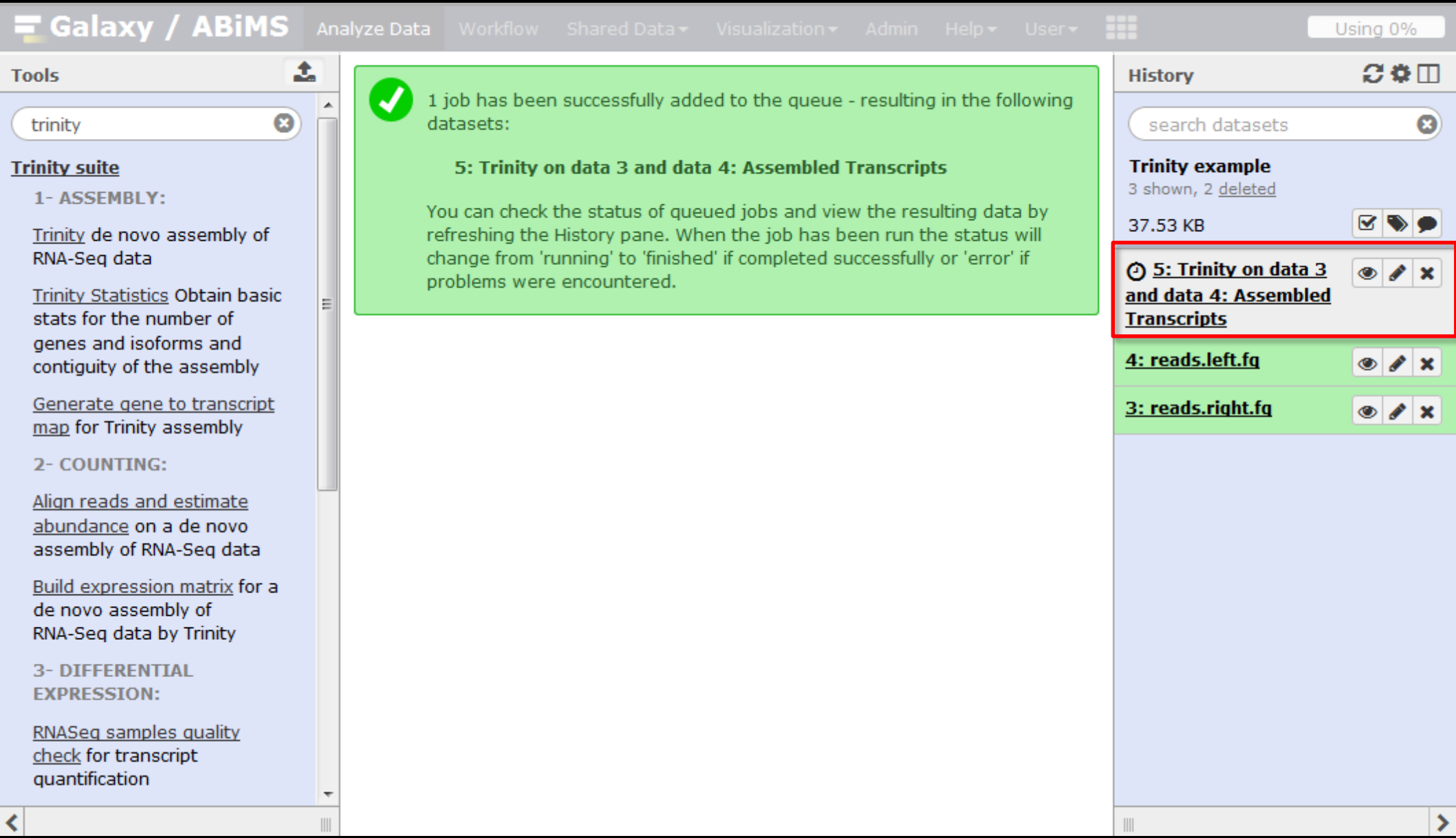
Tools can have some advanced options



Tools - form


Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 0%
Tools
trinity
Trinity suite
1- ASSEMBLY:
Trinity de novo assembly of RNA-Seq data
Trinity Statistics Obtain basic stats for the number of genes and isoforms and contiguity of the assembly
Generate gene to transcript map for Trinity assembly
2- COUNTING:
Align reads and estimate abundance on a de novo assembly of RNA-Seq data
Build expression matrix for a de novo assembly of RNA-Seq data by Trinity
3- DIFFERENTIAL EXPRESSION:
RNASeq samples quality check for transcript quantification
Trinity de novo assembly of RNA-Seq data (Galaxy Version 2.2.0.0)
Options
Paired or Single-end data?
Paired
Left/Forward strand reads
4: reads.left.fq
3: reads.right.fq
(--left)
Right/Reverse strand reads
4: reads.left.fq
3: reads.right.fq
(--right)
Strand specific data
Yes
No
Run in silico normalization of reads
Yes
No
Defaults to max. read coverage of 50. (--normalize_reads)
Additional Options
Execute
History
search datasets
Trinity example
2 shown, 3 deleted
40.02 KB
4: reads.left.fq
3: reads.right.fq

Tools - form


Galaxy / ABiMS
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 0%
Tools
trinity
Trinity suite
1- ASSEMBLY:
Trinity de novo assembly of RNA-Seq data
Trinity Statistics Obtain basic stats for the number of genes and isoforms and contiguity of the assembly
Generate gene to transcript map for Trinity assembly
2- COUNTING:
Align reads and estimate abundance on a de novo assembly of RNA-Seq data
Build expression matrix for a de novo assembly of RNA-Seq data by Trinity
3- DIFFERENTIAL EXPRESSION:
RNASeq samples quality check for transcript quantification
1 job has been successfully added to the queue - resulting in the following datasets:
5: Trinity on data 3 and data 4: Assembled Transcripts
You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.
History
search datasets
Trinity example
3 shown, 2 deleted
37.53 KB
5: Trinity on data 3 and data 4: Assembled Transcripts
4: reads.left.fq
3: reads.right.fq

Tools - Job status

Job is waiting to run

= the job is in the scheduler « queue »

Duration time of this status depends on the amount of actual queued jobs and on the requested number of processors

Tools - Job status

Job is currently running

= the job is being executed on the computing cluster

Duration time of this status depends on the job's attributes and the computing ressources allocated.

Some programs are executed with several processors (using 4, 8 or 16 Gb of RAM).

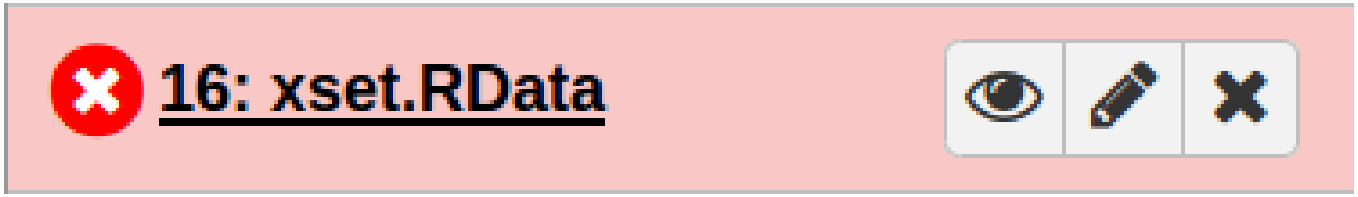
And others are mono-threaded ☹

Tools - Job status

Job is finished and status is OK

But warnings or errors can be hidden behind!

Tools - Job status

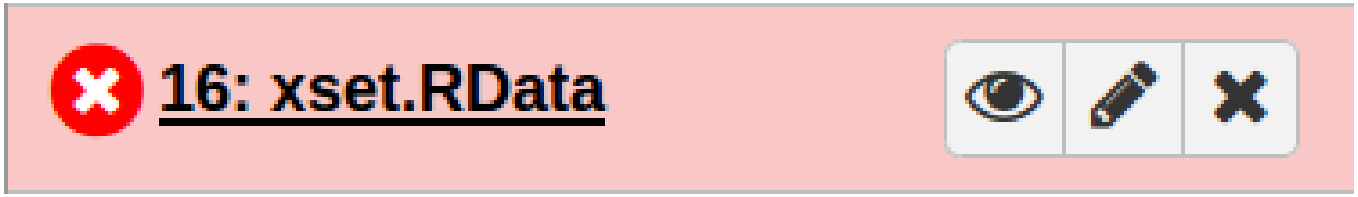


Job is finished but with an error status

= the program sends an error

The error is often explained by the program but sometimes … not.

Tools - Job status



Job is finished but with an error status

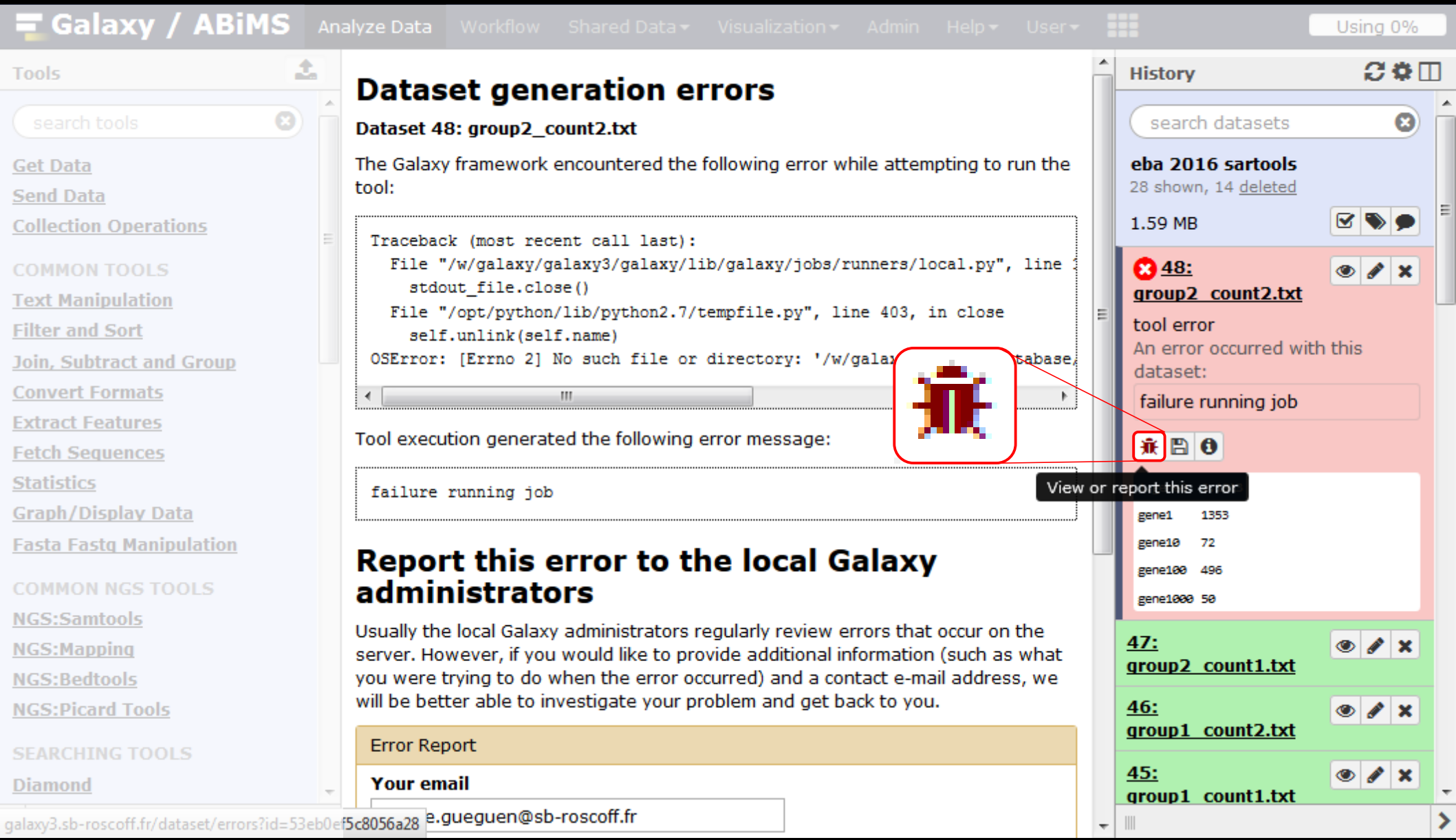
= the program sends an error

Possible causes of error :

- ~~The user~~ :P
- Bad usage : input file, format or option
- Bad integration of the program into Galaxy … sorry :/
- Non anticipated crash of the program

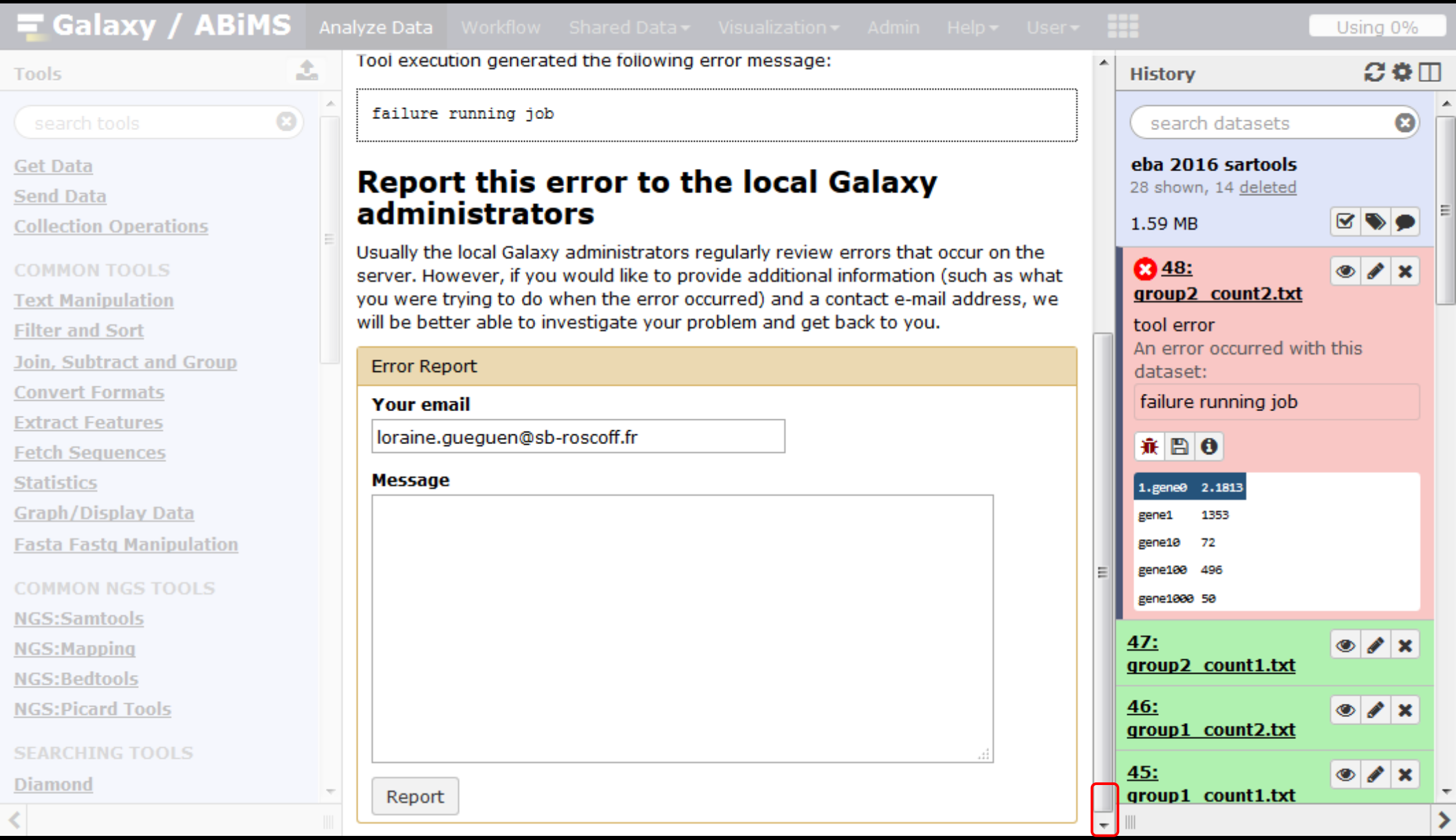
Tools - Handle errors

Tools - Handle errors



Tools - Handle errors

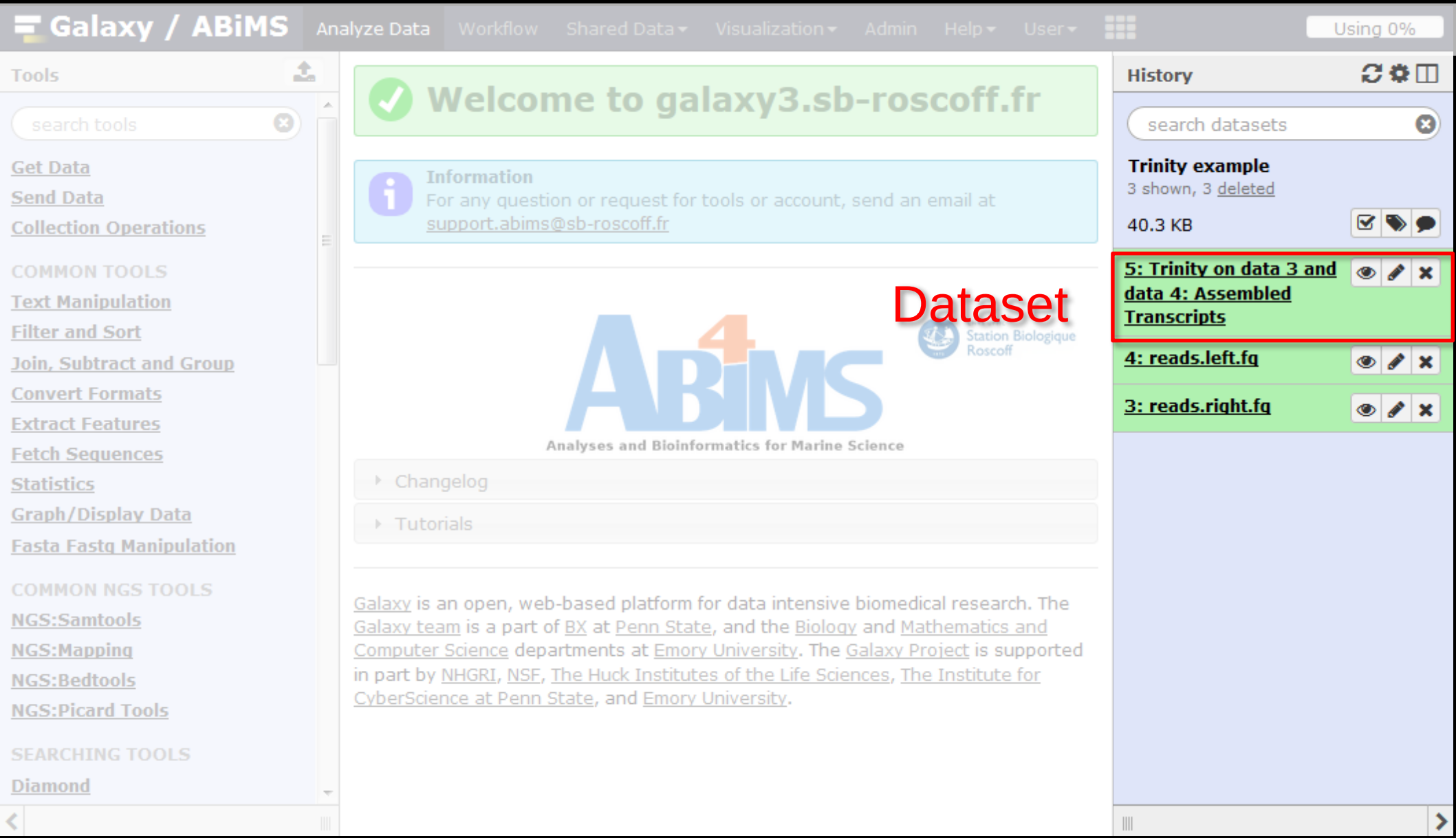
Sent to the support team



DATASET

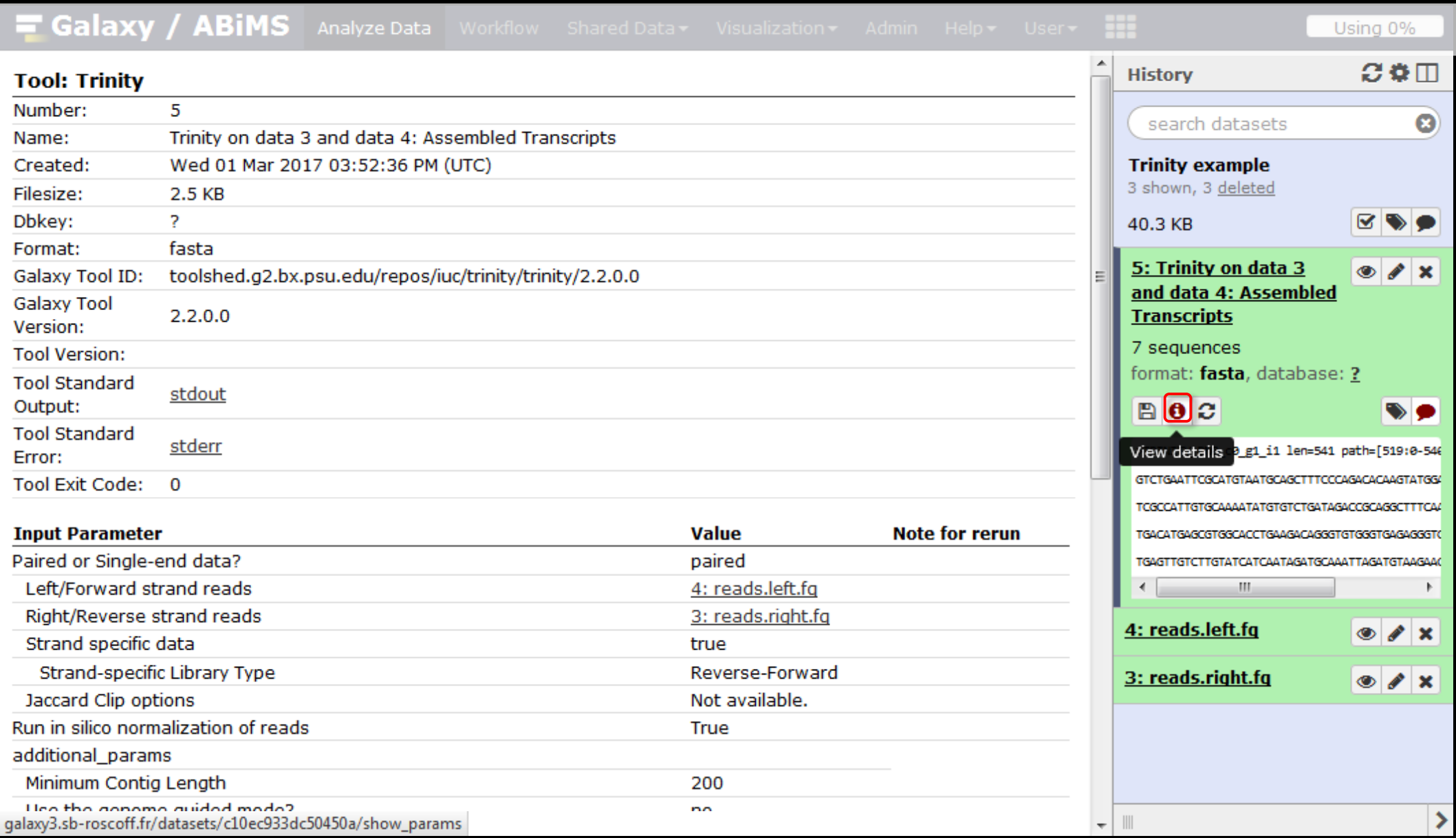
Dataset

Both inputs and outputs



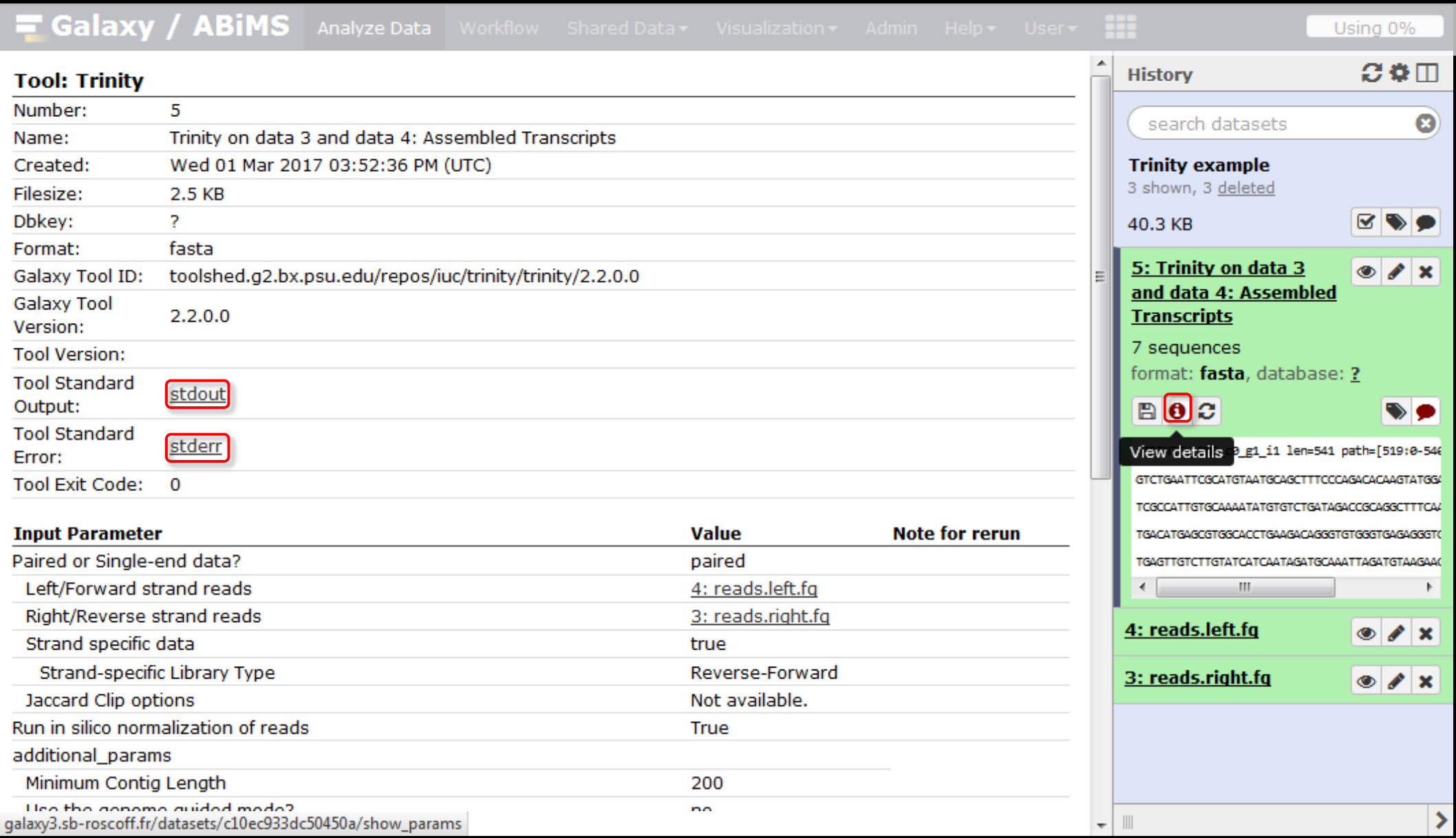
Dataset

Informations



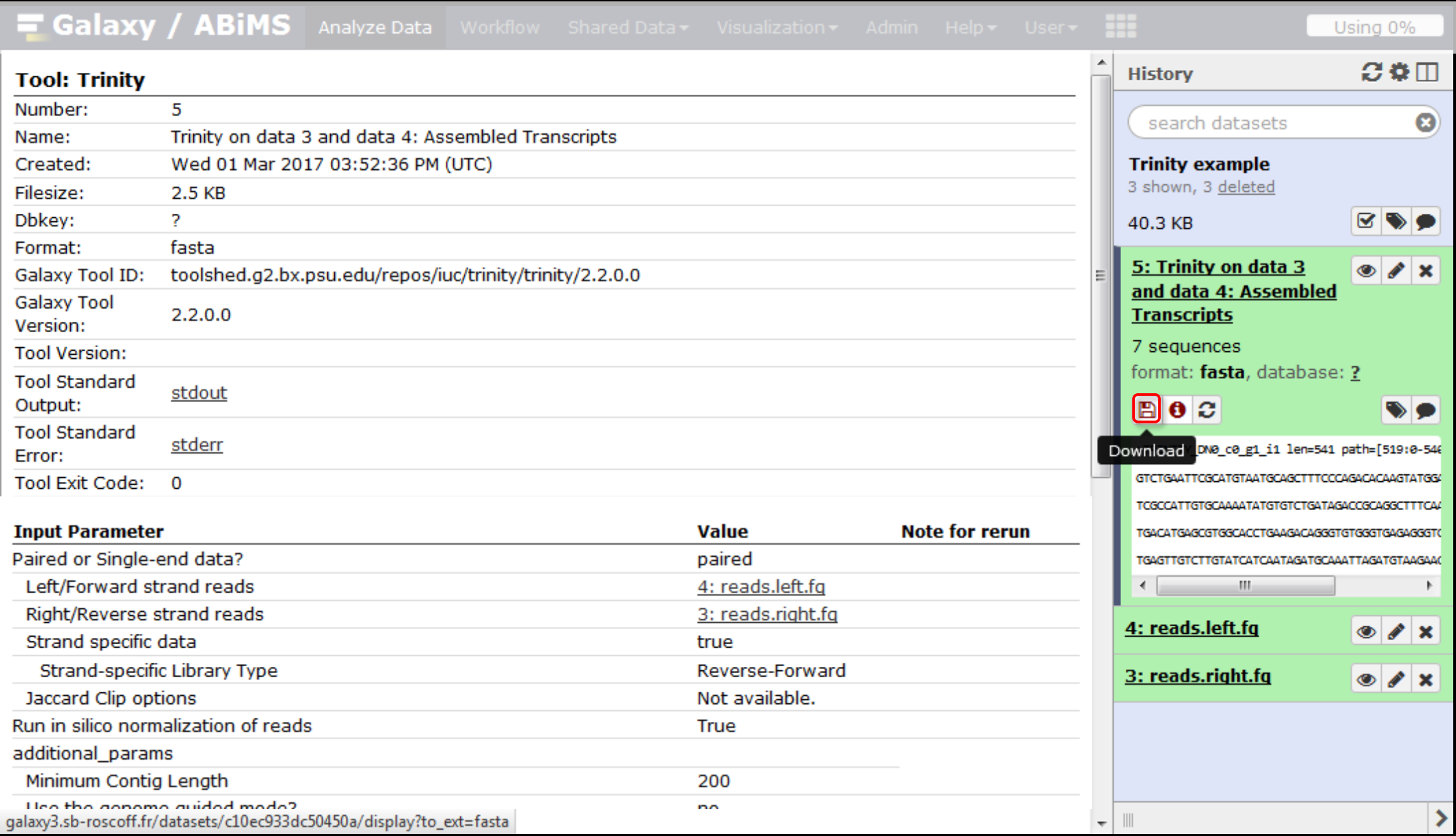
Dataset

Informations



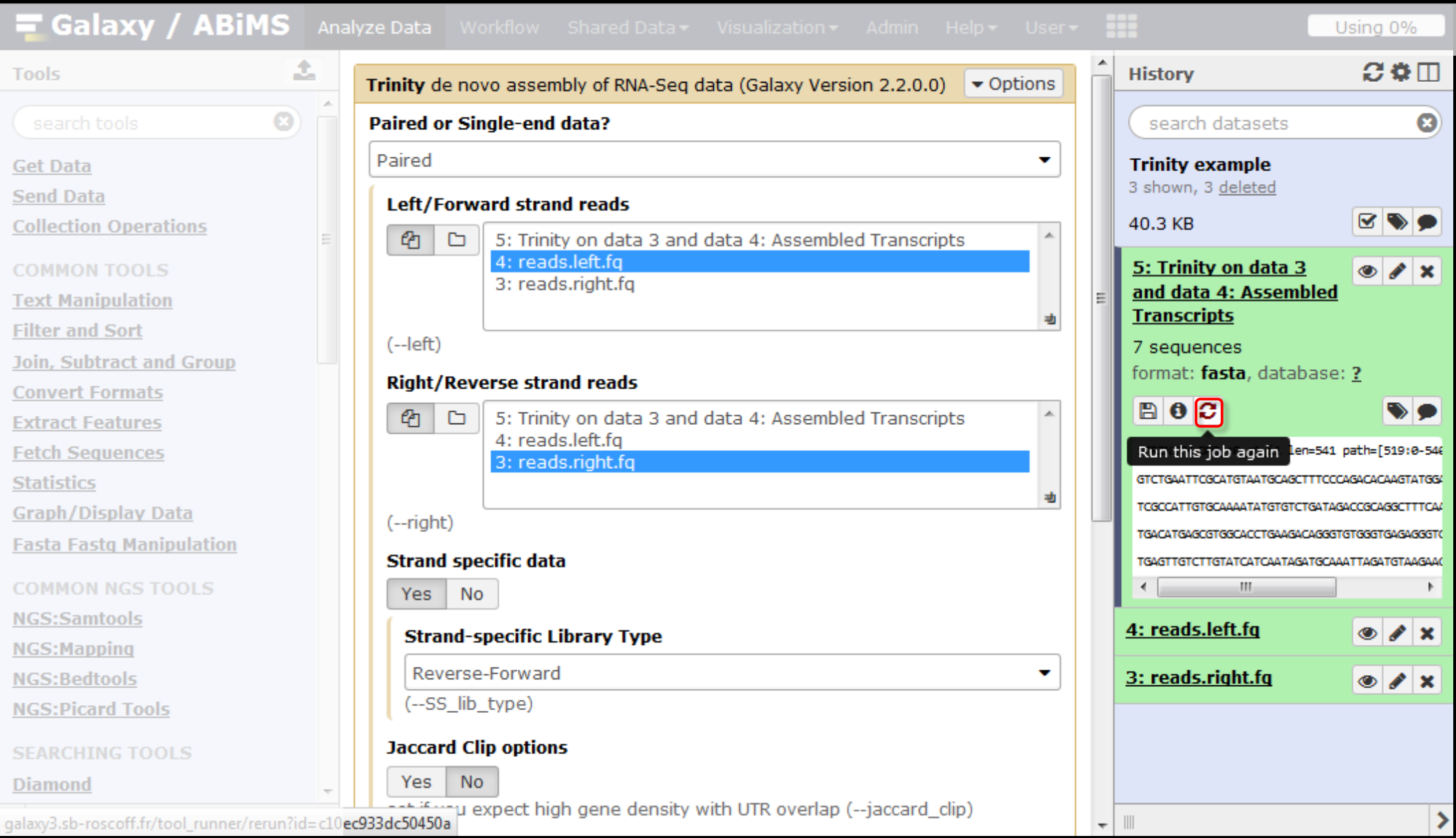
Dataset

Download



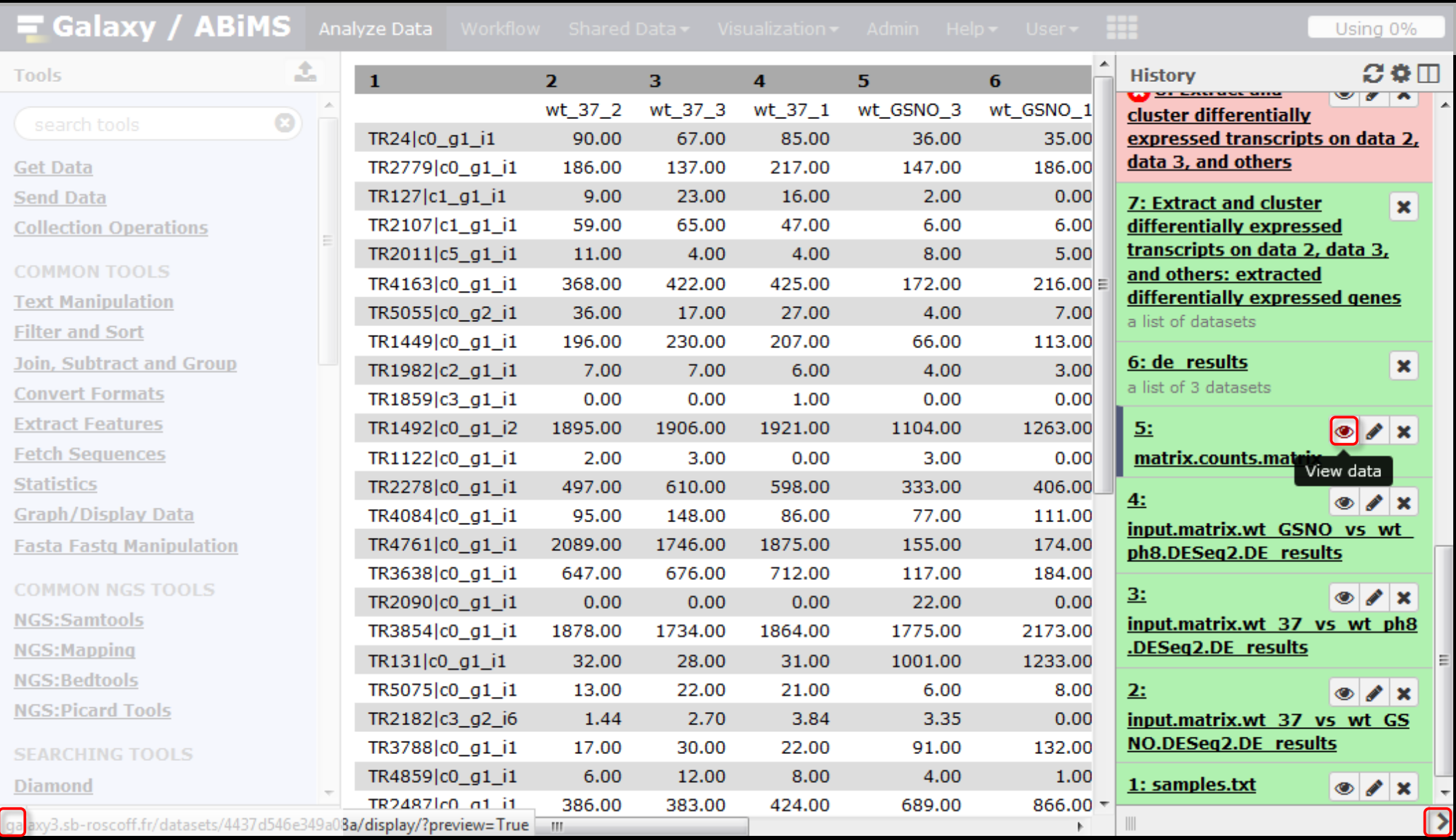
Dataset

Re-run a job



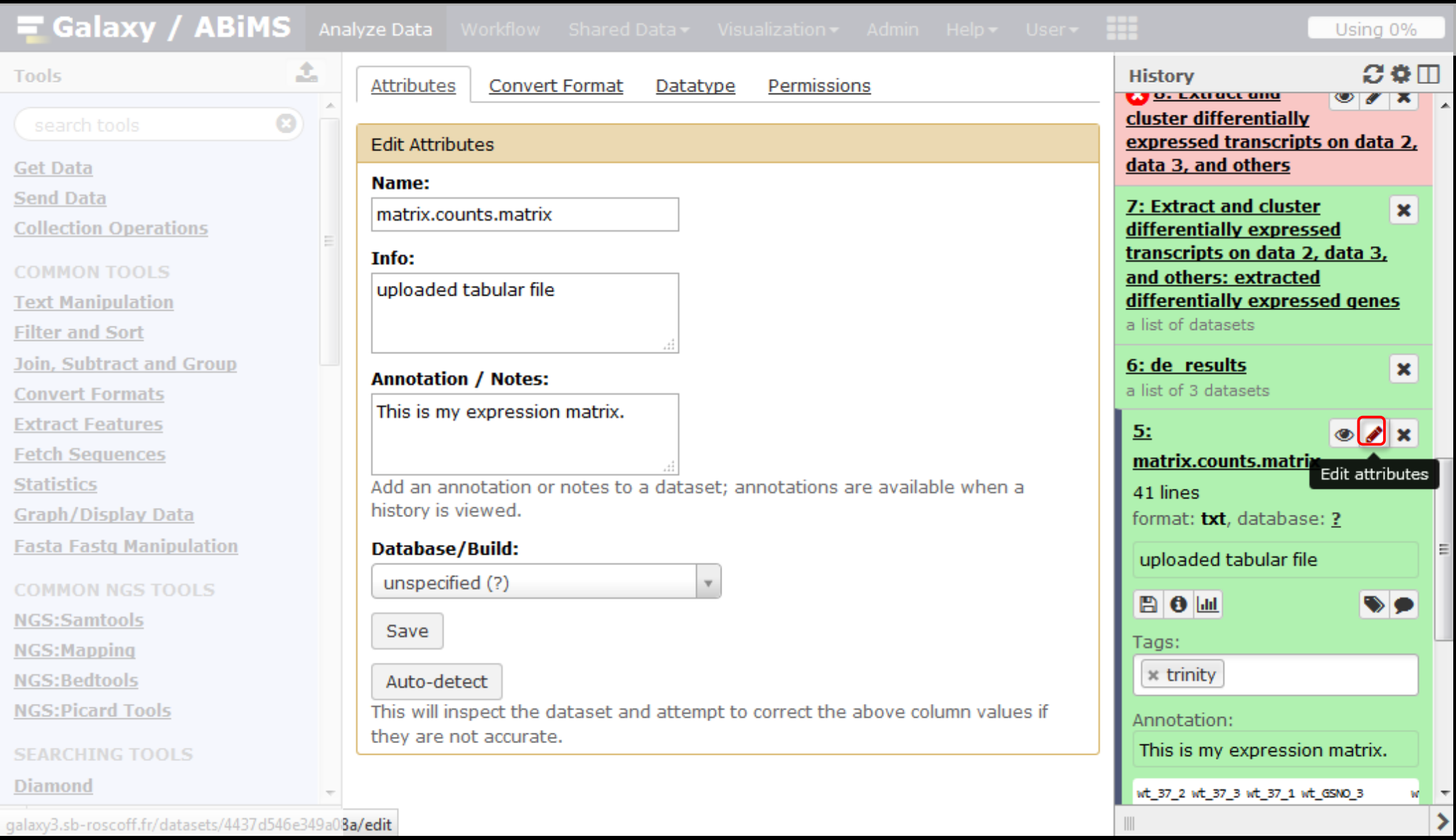
Dataset

Dataset display : text, tabular, pdf, picture, html …



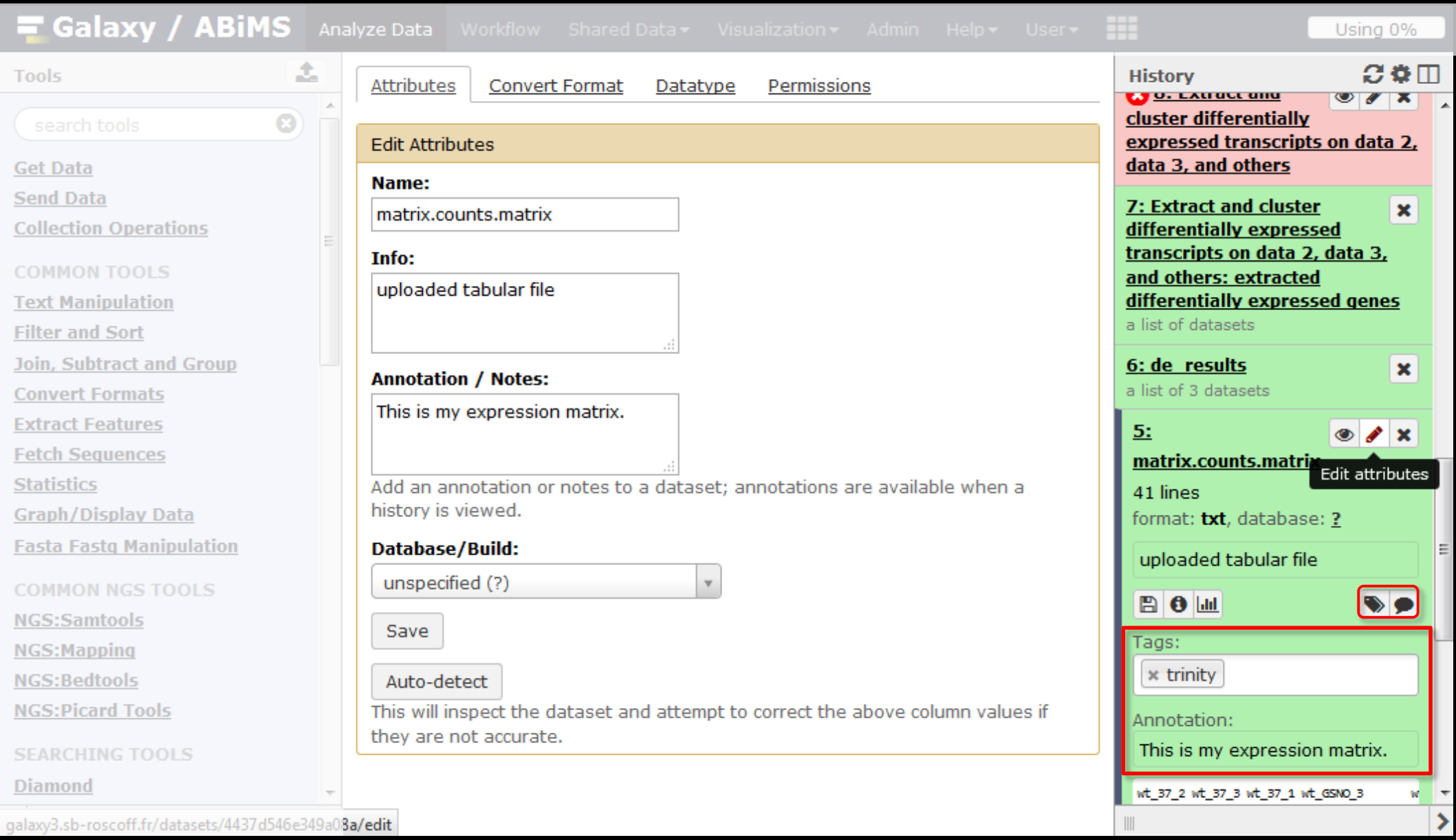
Dataset

Renaming and annotation



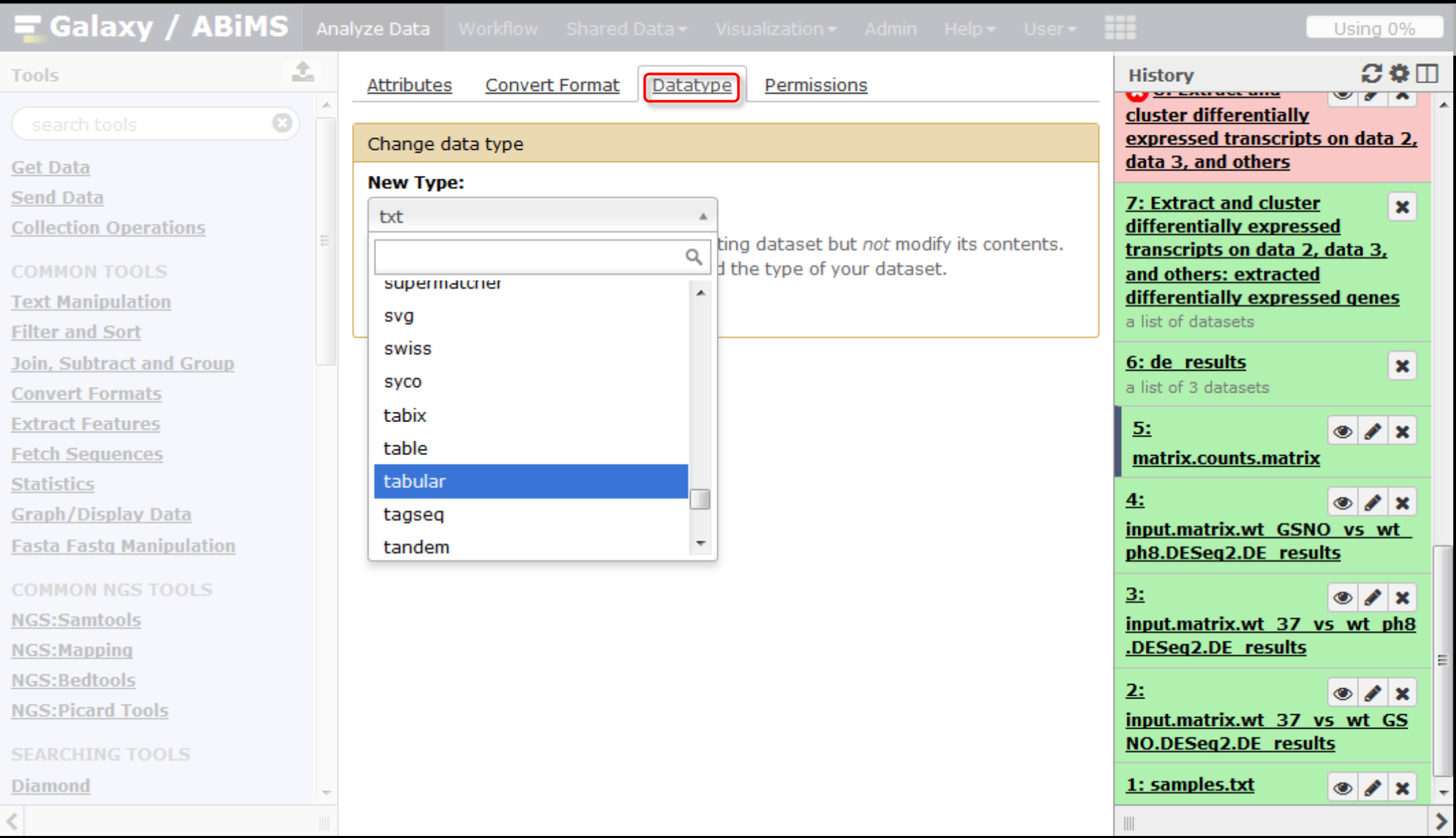
Dataset

Renaming and annotation



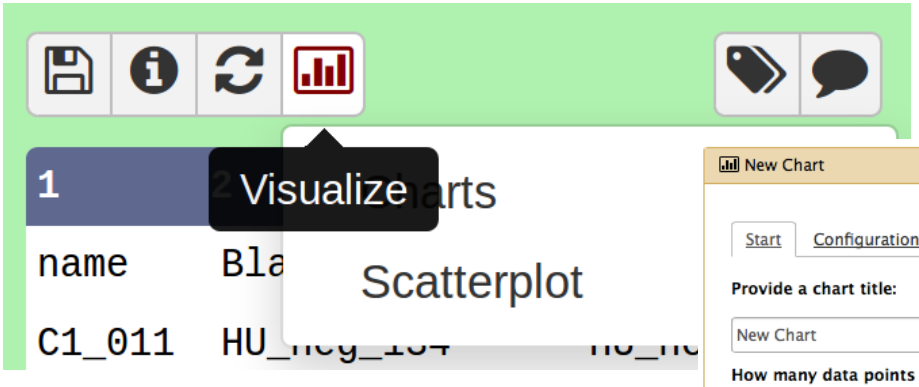
Dataset

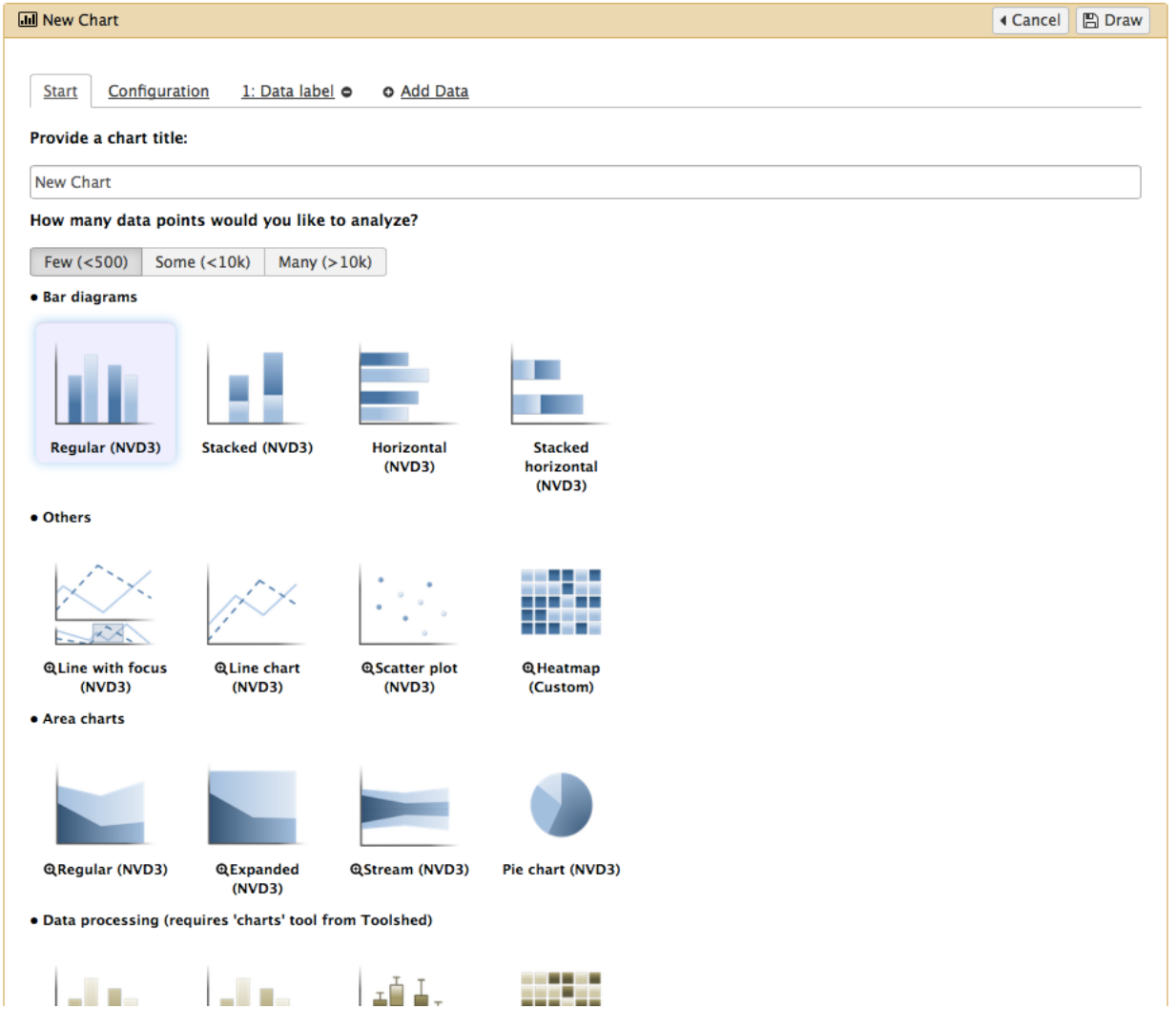
Change the Datatype of the Dataset



Dataset

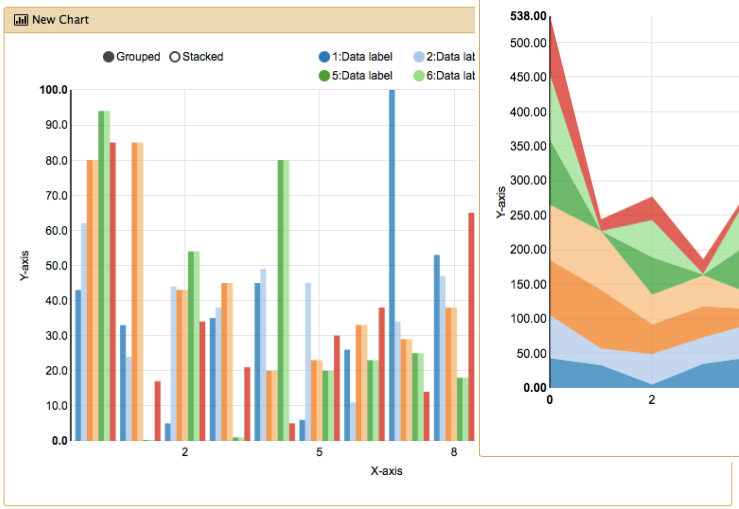
Graphics

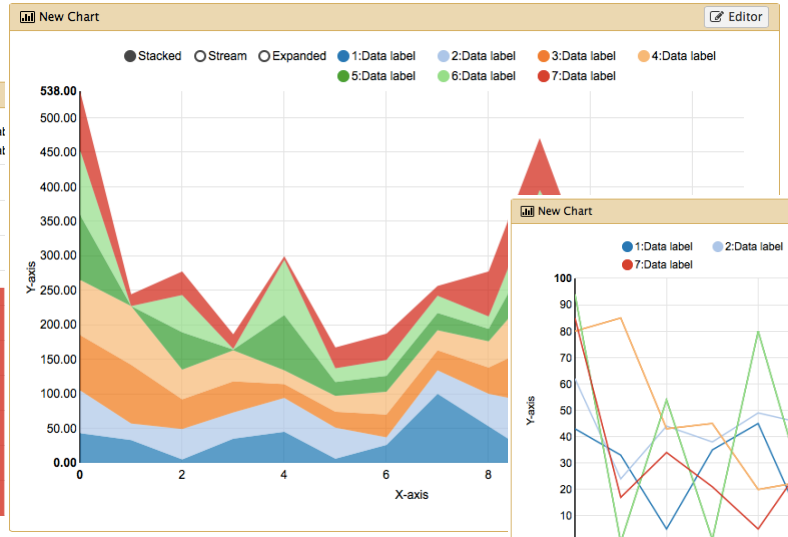


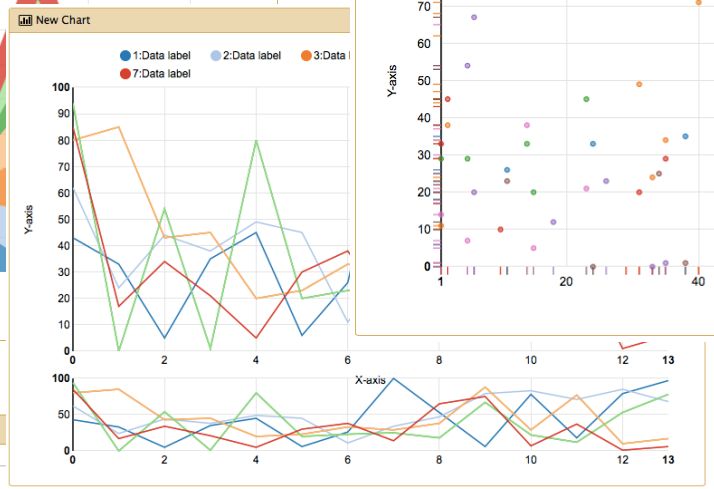


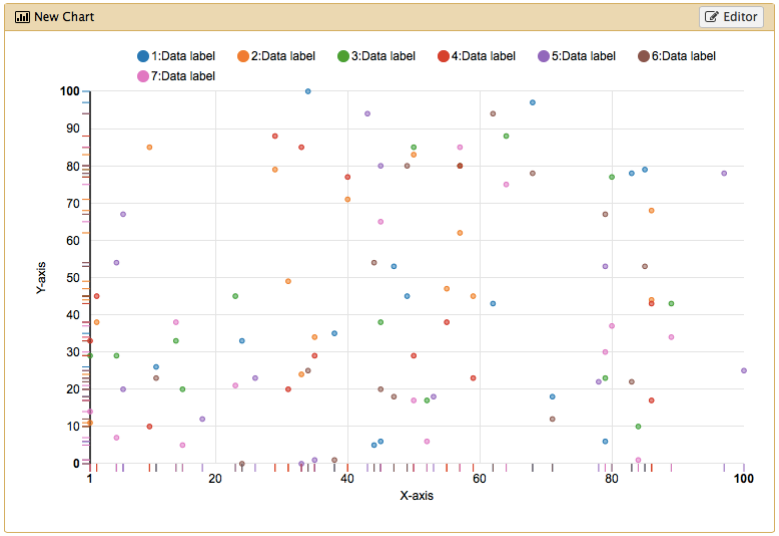
Dataset

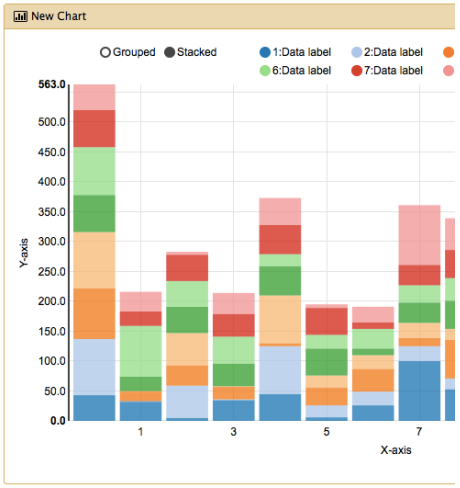
Graphics

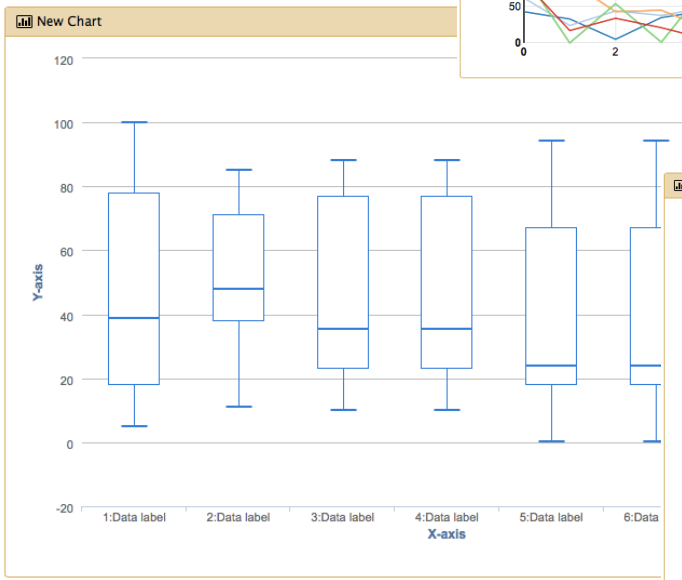


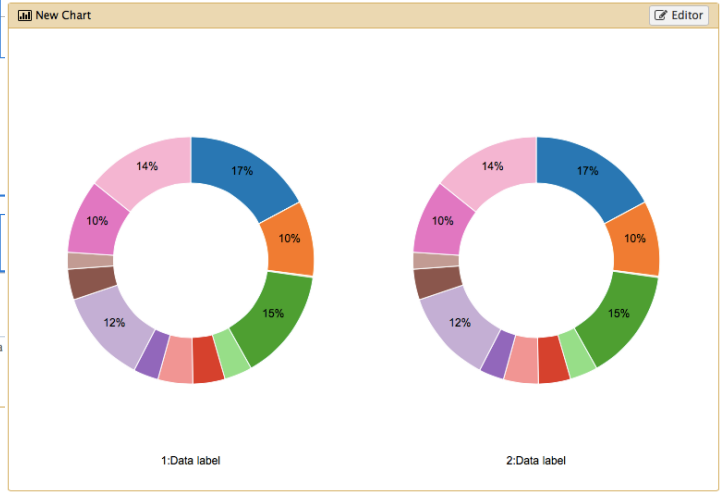












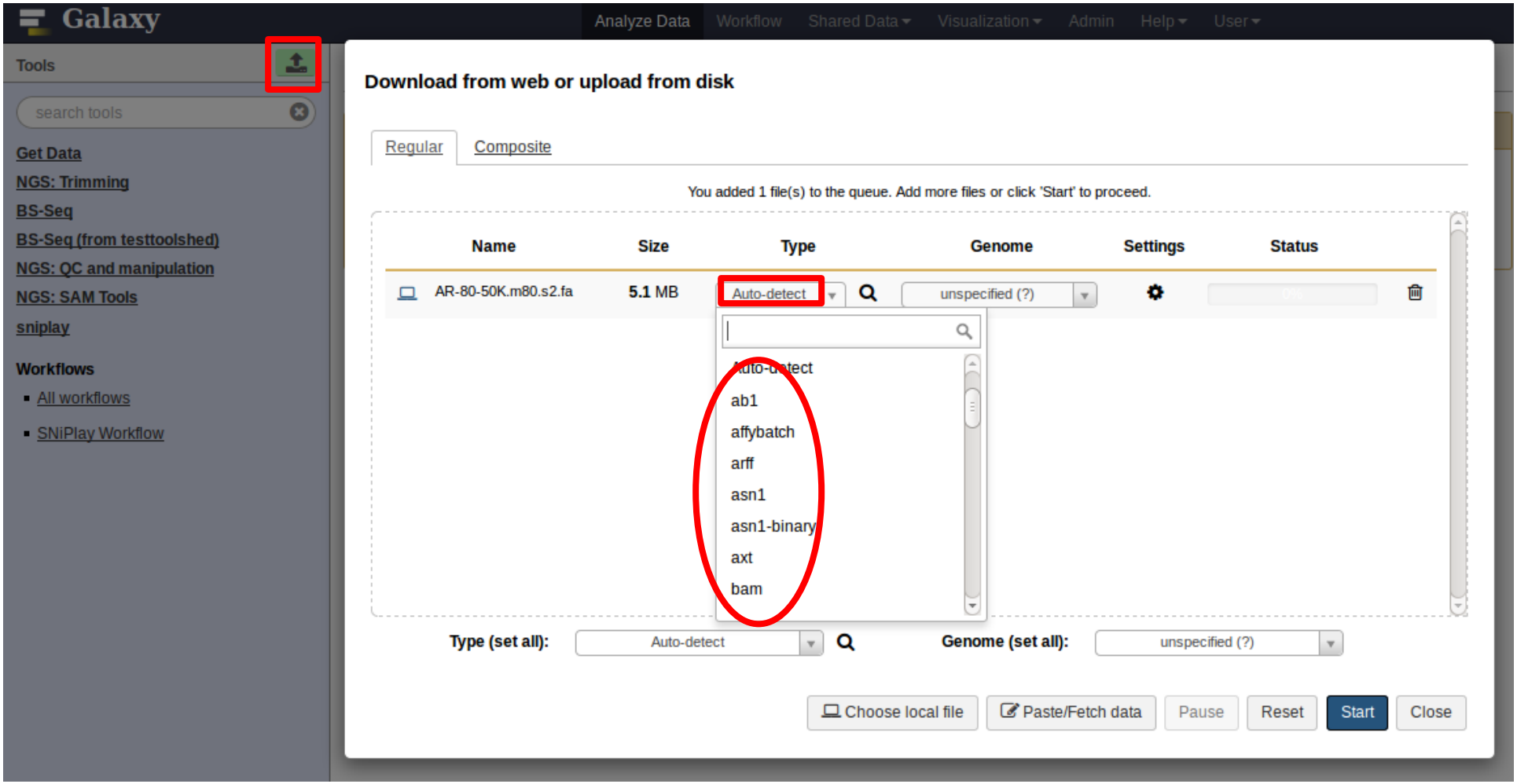
Datatypes

DATASET

Dataset

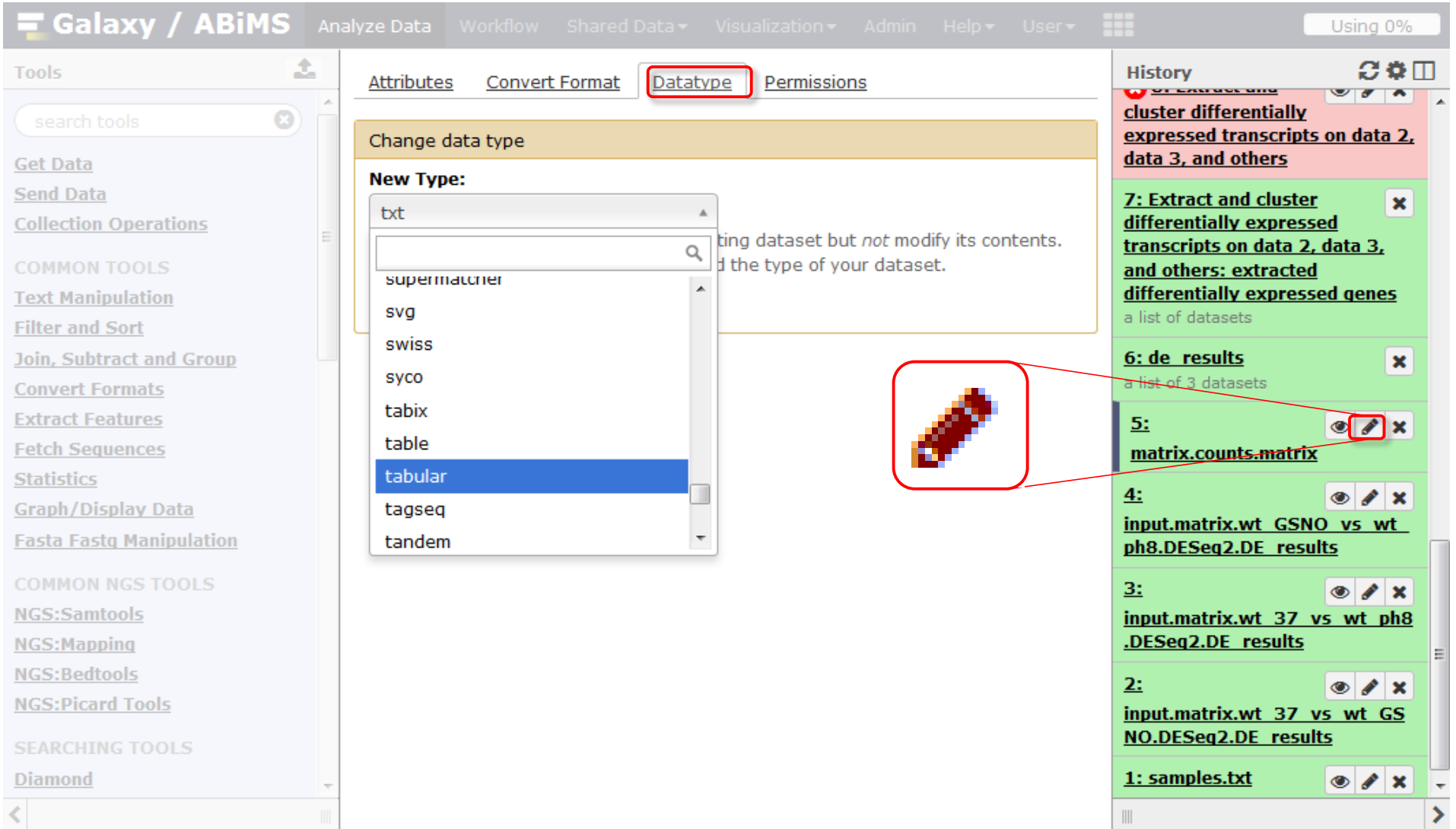
Datatypes

- Every Galaxy dataset is associated with a datatype.
- Datatype can be detected or user specified.



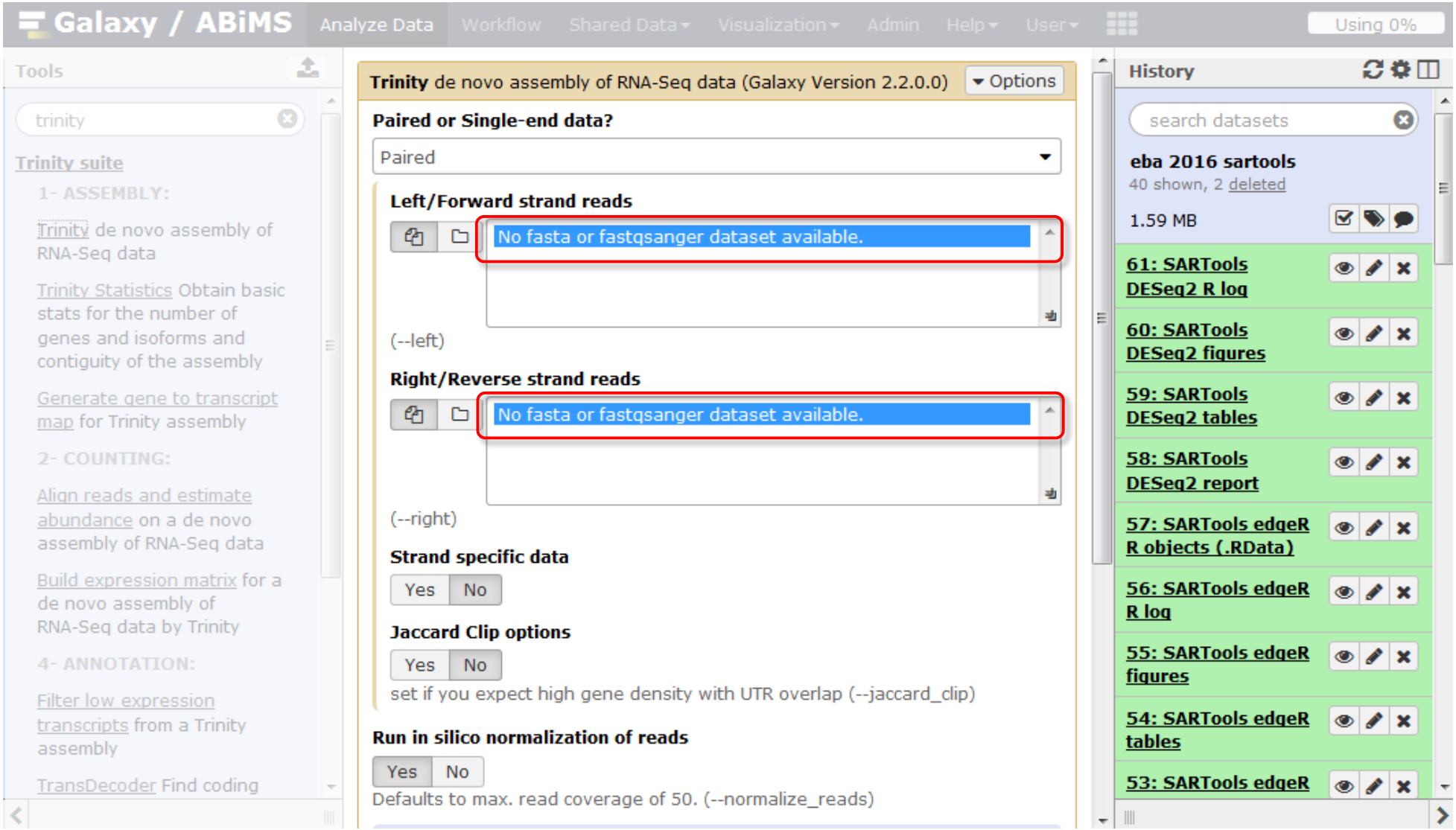
Dataset - Datatypes

- Every Galaxy dataset is associated with a datatype.
- Datatype can be detected or user specified.



Dataset - Datatypes

- Many tools will only accept input datasets with the appropriate datatype assigned.



Dataset - Datatypes

Common text formats:

- ***txt***: plain text ('.txt')
- ***tabular***: tab delimited ('.tab', '.txt', etc.)

```
                  wt_37_2  wt_37_3  wt_37_1
TR24|c0_g1_i1     90.00    67.00    85.00
TR2779|c0_g1_i1   186.00   137.00   217.00
TR127|c1_g1_i1    9.00     23.00    16.00
```

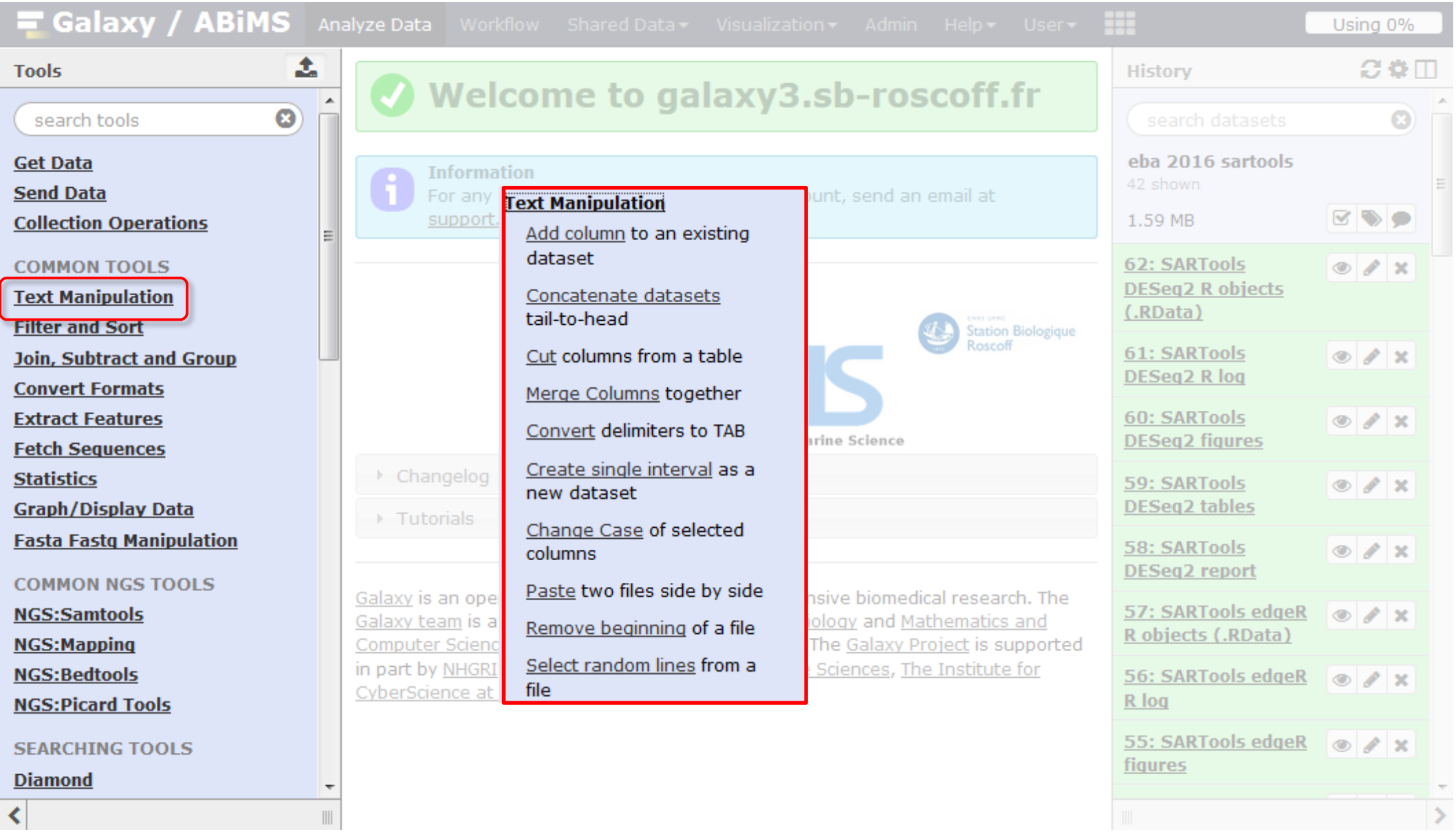
- ***csv***: comma-separated values ('.csv')

```
Year,Make,Model
1997,Ford,E350
2000,Mercury,Cougar
```

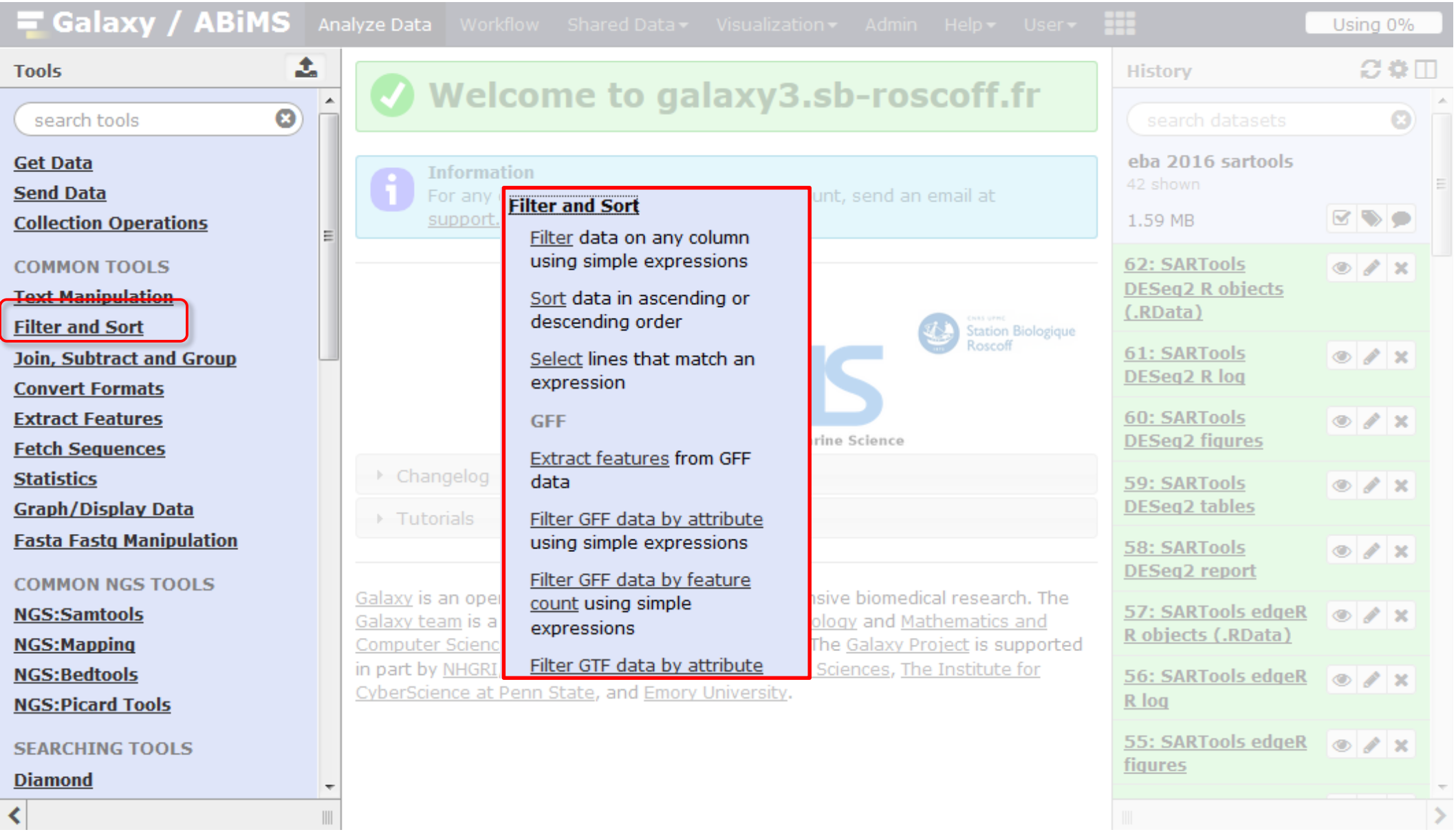
- ***html***: standard language for web pages

```
<!DOCTYPE html>
<html>
  <head>
    <title>This is a title</title>
  </head>
  <body>
    <p>Hello world!</p>
  </body>
</html>
```

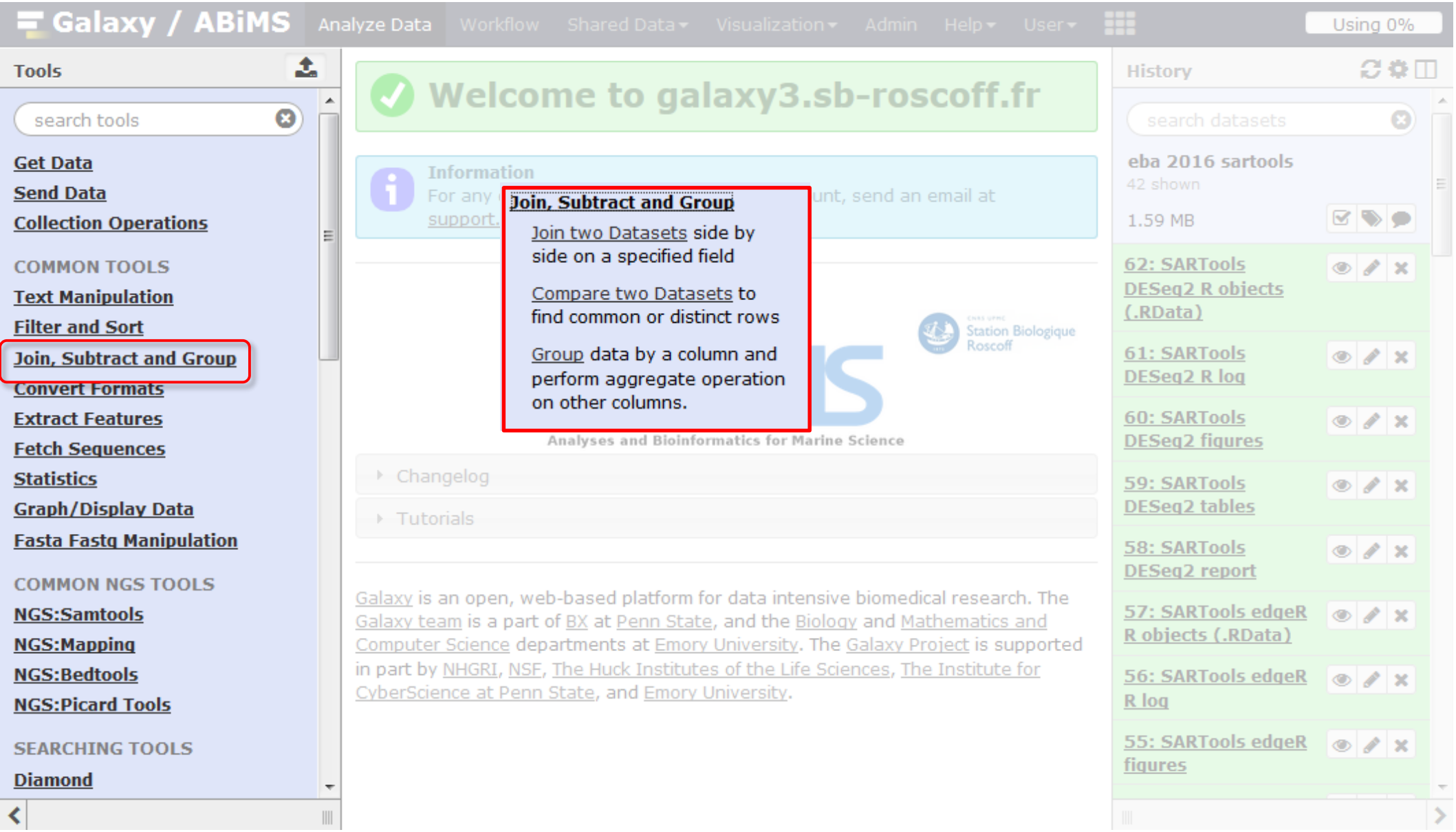
Dataset - Datatypes



Dataset - Datatypes



Dataset - Datatypes



Dataset - Datatypes

Common binary formats:

- ***data***: generic binary format
- ***zip, tar***: archives
- ***pdf, png, jpg, bmp, tiff, gif***: images
- ***rdata***: statistical computing program R
- ***bam*, wig, bigwig**: sequence alignment

Dataset - Datatypes

Sequence file formats:

- ***fasta***: a single-line description with '>', followed by lines of sequence data ('.fasta', '.fas')

```
>sequence1
atgcgtttgcgtgcatgcgtttgcgtgcatgcgtttgcgtgcatgcgtttgcgtgc
atgcgtttgcgtgc
>sequence2
tttcgtgcgtatagtttcgtgcgtatagtttcgtgcgtatagtttcgtgcgtatag
tggcgcggt
```

- ***fastq***: sequence + quality score ('.fastq', '.fq')

```
@SEQ_ID
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!''*((((***+))%%%++)(%%%%).1***-+*''))**55CCF>>>>>>CCCCCCC65
@SEQ_ID2
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!''*((((***+))%%%++)(%%%%).1***-+*''))**55CCF>>>>>>CCCCCCC65
```

Dataset - Datatypes

Dataset - Datatypes

Sequence file formats:

- ***gff3, bed, genbank***: sequence + annotations

bed

```
track name=pairedReads description="Clone Paired Reads" useScore=1
chr22 1000 5000 cloneA 960 + 1000 5000 0 2 567,488, 0,3512
chr22 2000 6000 cloneB 900 - 2000 6000 0 2 433,399, 0,3601
```

gff3

```
##gff-version 3
ctg123  .  exon  1300  1500  .  +  .  ID=exon00001
ctg123  .  exon  1050  1500  .  +  .  ID=exon00002
ctg123  .  exon  3000  3902  .  +  .  ID=exon00003
##FASTA
>ctg123
cttctgggcgtacccgattctcggagaacttgccgcaccattccgccttg
tgttcattgctgcctgcatgttcattgtctacctcggctacgtgtggcta
…
```

Cleanup

DATASET

Dataset

Delete a dataset

Dataset

The dataset isn't really deleted. It's in the Trash

Dataset

"Empty Trash" : to free up disk space

Galaxy / ABiMS | Analyze Data | Workflow | Shared Data | Visualization | Admin | Help | User | Using 0%

Tools

search tools

Get Data
Send Data
Collection Operations

COMMON TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Statistics
Graph/Display Data
Fasta Fastq Manipulation

COMMON NGS TOOLS
NGS:Samtools
NGS:Mapping
NGS:Bedtools
NGS:Picard Tools

SEARCHING TOOLS
Diamond

Welcome to galaxy3.sb-roscoff.fr

Information
For any question or request for tools or account, send an email at support.abims@sb-roscoff.fr

ABiMS
Analyses and Bioinformatics for Marine Science

Station Biologique Roscoff

Changelog

Tutorials

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

HISTORY LISTS
Saved Histories
Histories Shared with Me
HISTORY ACTIONS
Create New
Copy History
Share or Publish
Show Structure
Extract Workflow
Delete
Delete Permanently
DATASET ACTIONS
Copy Datasets
Dataset Security
Resume Paused Jobs
Collapse Expanded Datasets
Unhide Hidden Datasets
Delete Hidden Datasets
Purge Deleted Datasets
DOWNLOADS
Export Tool Citations
Export History to File
OTHER ACTIONS
Import from File

galaxy3.sb-roscoff.fr/history/purge_deleted_datasets

HISTORY

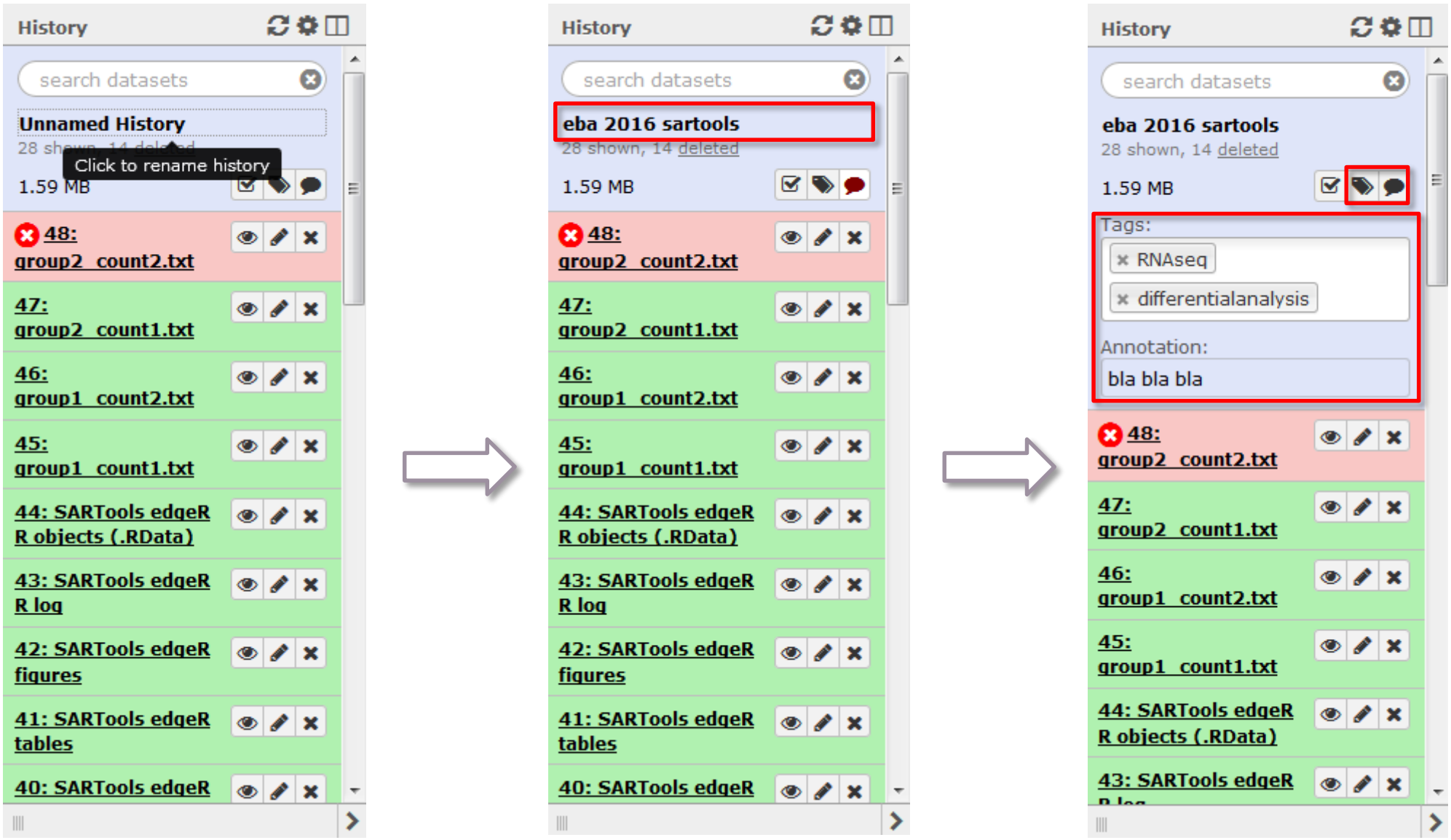
History panel

Both inputs and outputs

History panel

Both inputs and outputs

History panel renaming and annotation

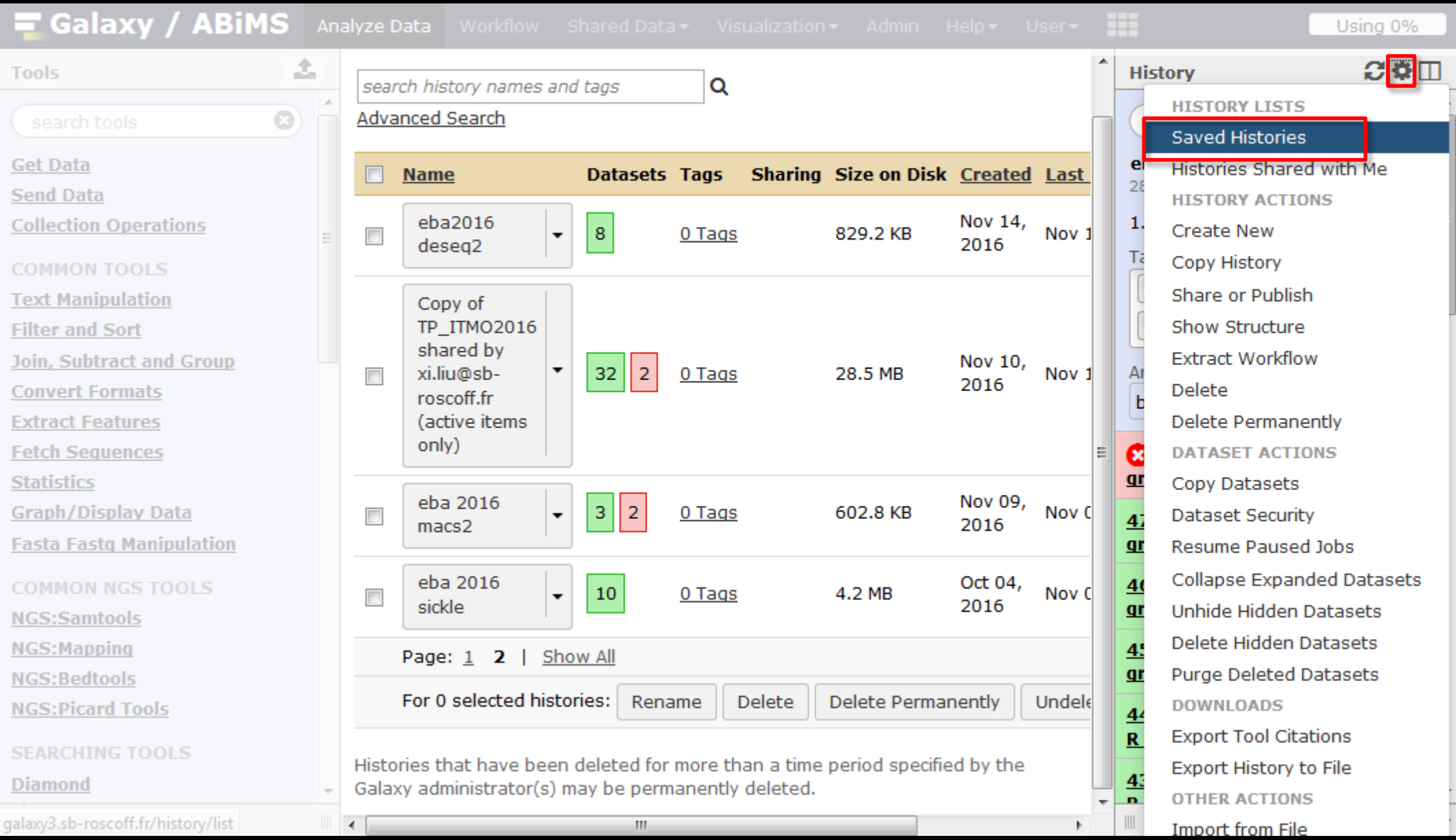


History panel

History menu: Create new, Rename, Delete, Delete Permanently

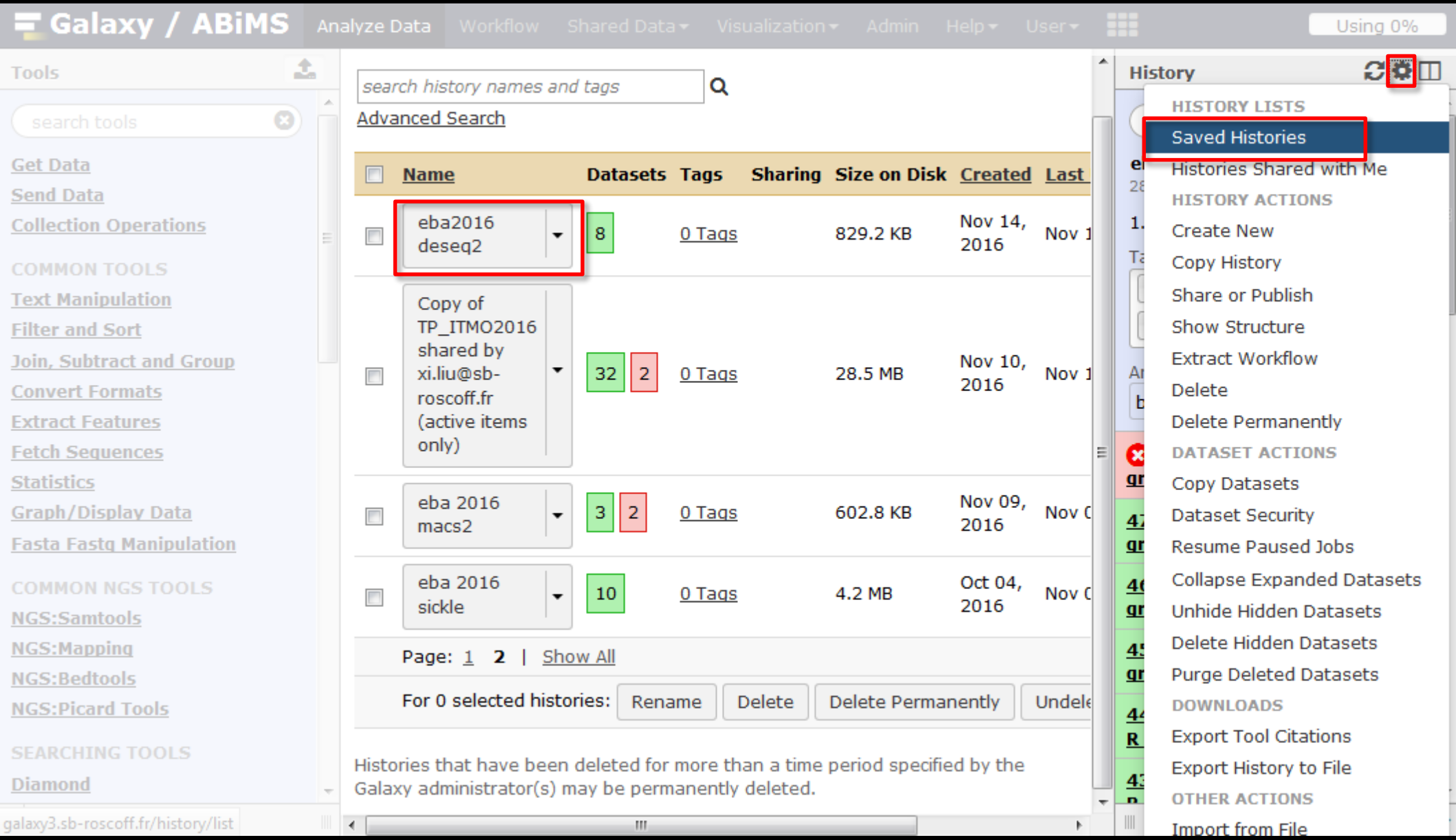
History panel

Saved histories



History panel

Saved histories: Switch histories



History panel

History menu: Create new, Rename, Delete, Delete Permanently

History panel

History panel

History panel

History panel

History panel

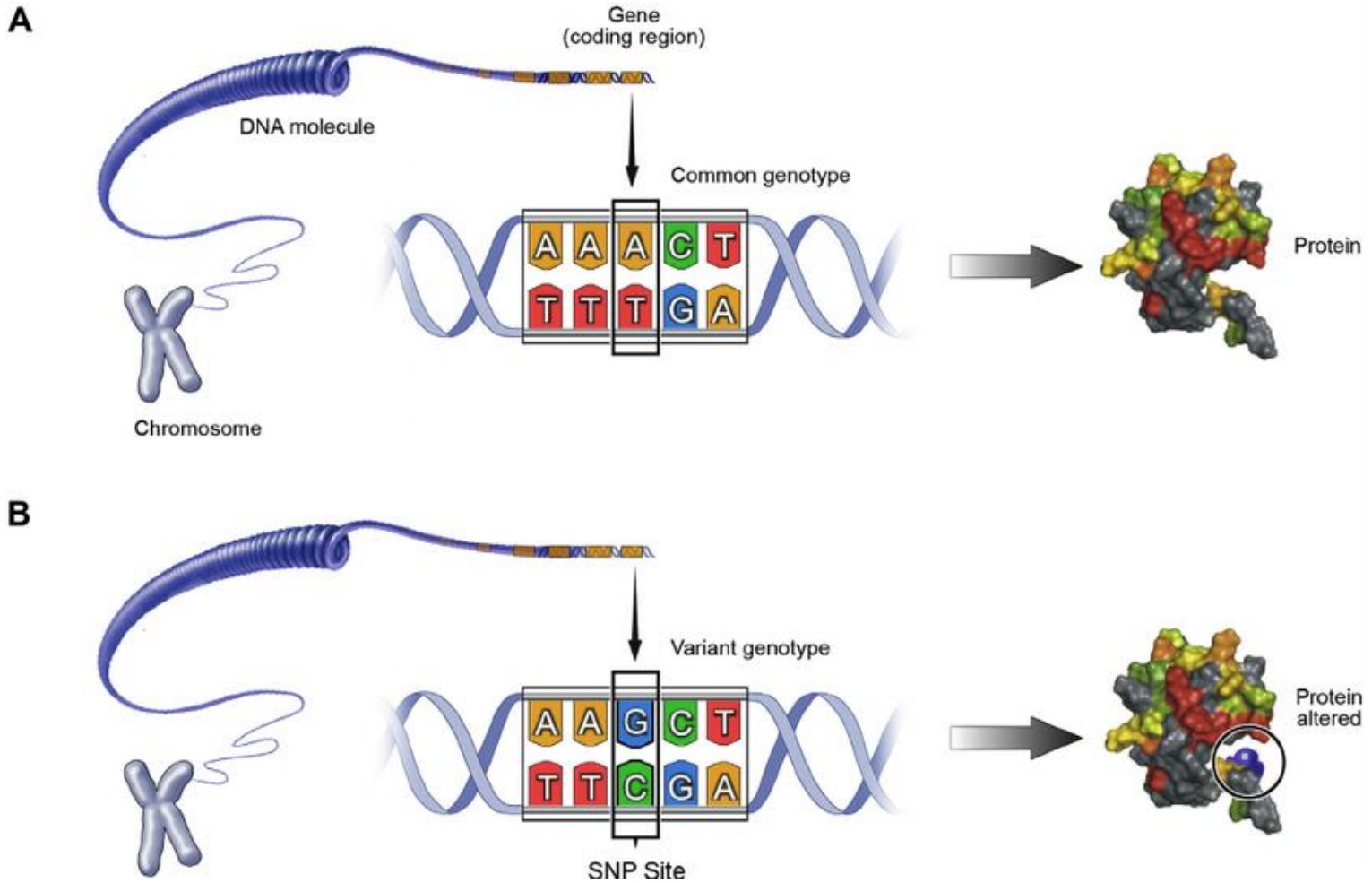
History panel

Hands-on

TOOLS

Tools – Hands-on

Which coding exon has the highest number of single nucleotide polymorphisms on chromosome 22?



DOI: 10.1016/j.jand.2013.12.001

Tools – Hands-on

Which coding exon has the highest number of single nucleotide polymorphisms on chromosome 22?

1. Get the data in a new history
2. Join exons with SNPs
3. Count the number of SNPs per exon
4. Sort exons by SNP count
5. Select top five
6. Build a bar diagram
7. Recover exon info and display data in genome browsers

Tools – Hands-on

Which coding exon has the highest number of single nucleotide polymorphisms on chromosome 22?

1. Get the data in a new history
2. Join exons with SNPs *Operate on Genomics Intervals -> Join*
3. Count the number of SNPs per exon *Join, Subtract, and Group -> Group*
4. Sort exons by SNP count *Filter and Sort -> Sort*
5. Select top five *Text Manipulation -> Select First*
6. Build a bar diagram *Visualize -> Charts*
7. Recover exon info and display data in genome browsers *Join, Subtract and Group -> Compare two Datasets*

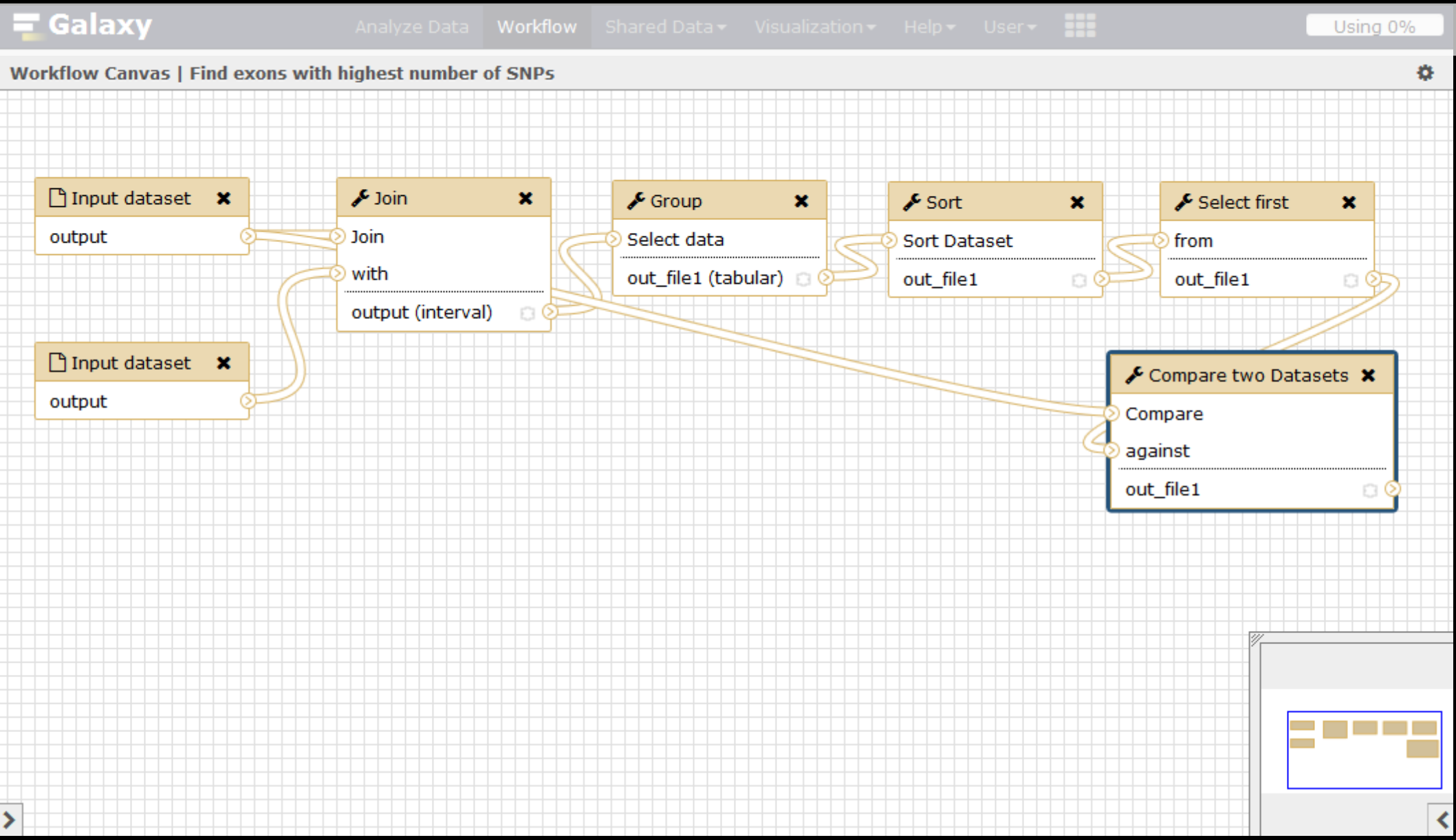
WORKFLOW

Workflow

- A workflow is a sequence of tool operations and parameters
- Can match the experiment protocol
- A workflow is built to be replayed (more or less strict)

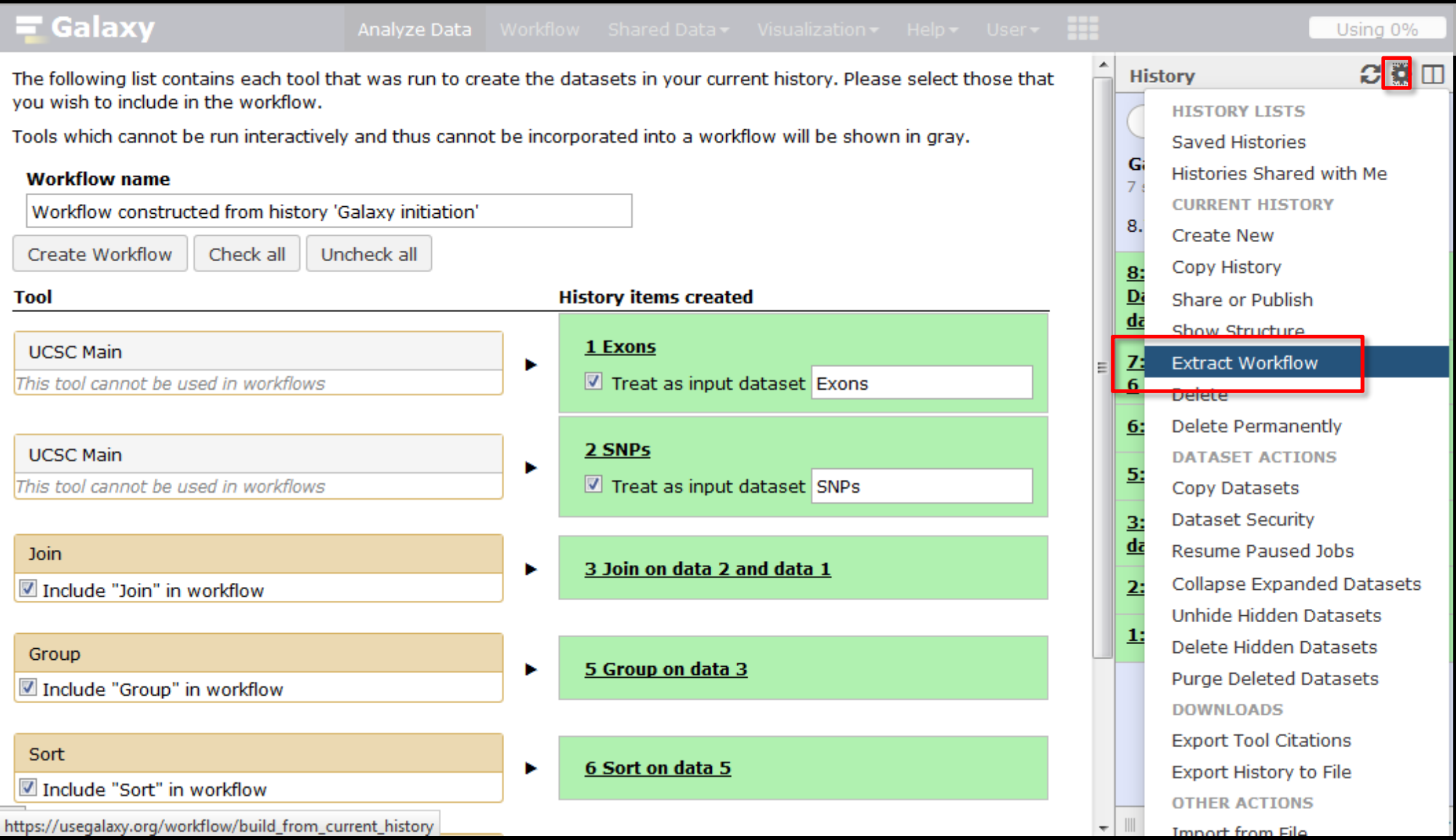
Workflow

Our workflow with Galaxy



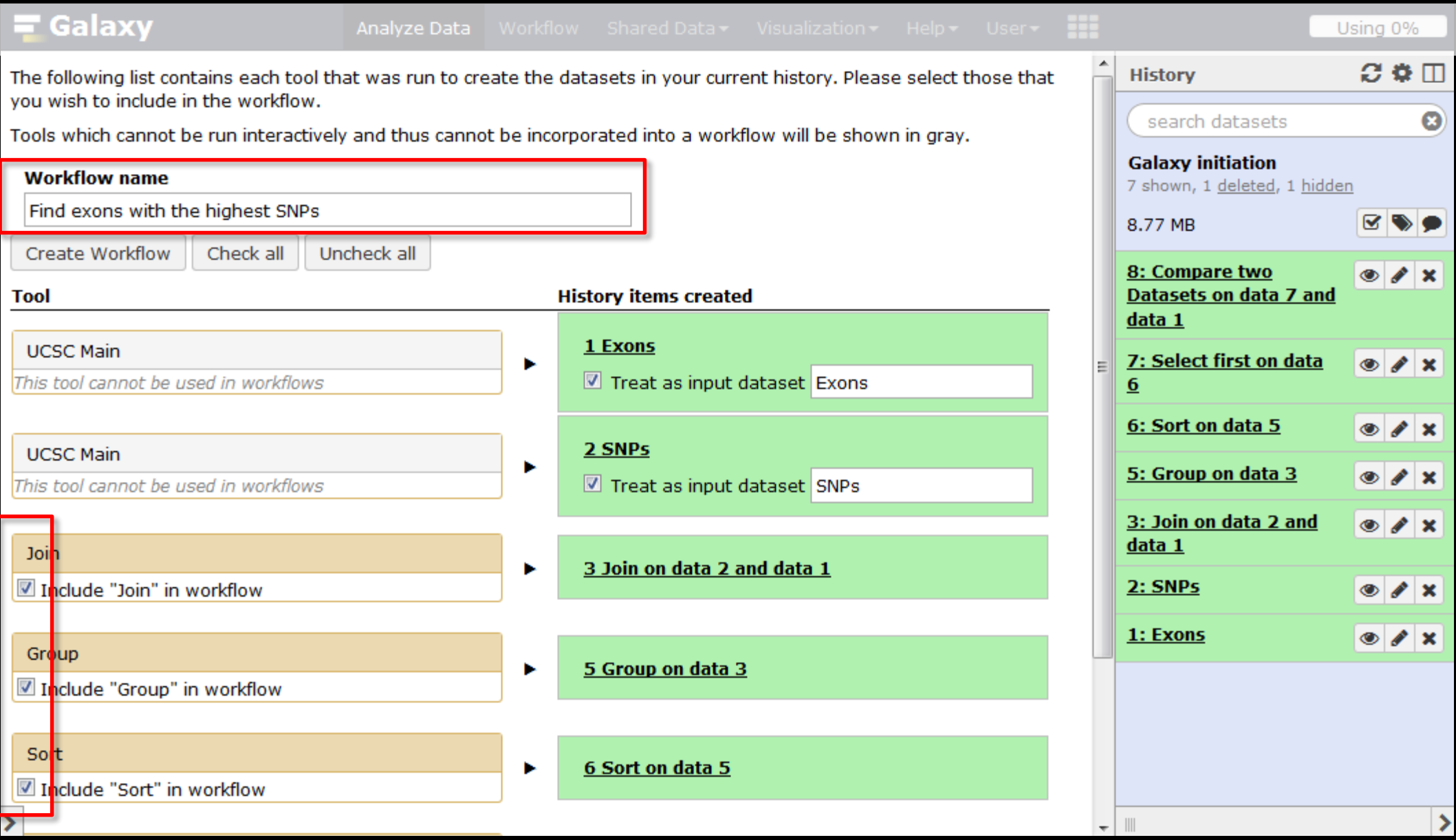
Workflow

From history



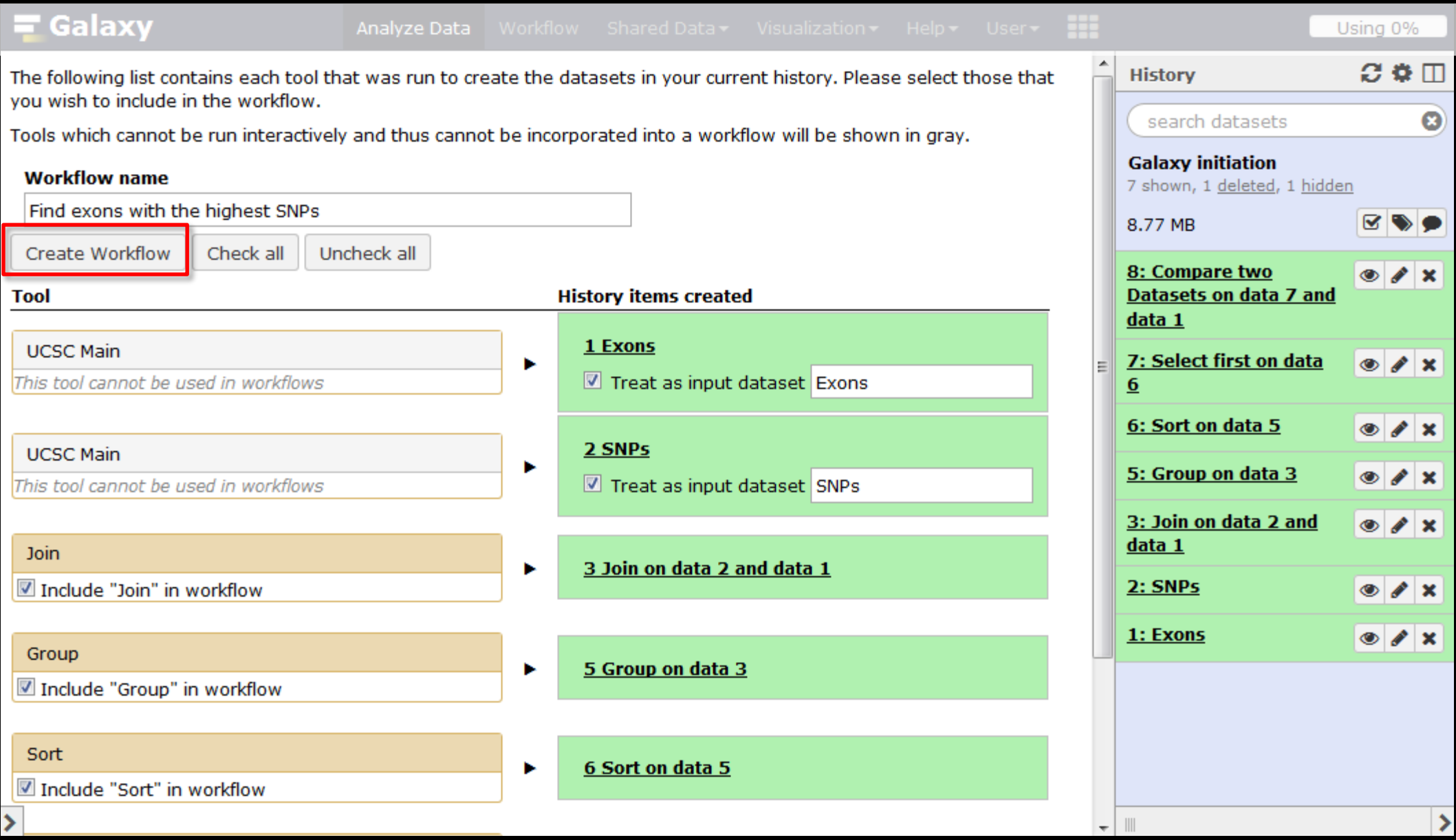
Workflow

From history



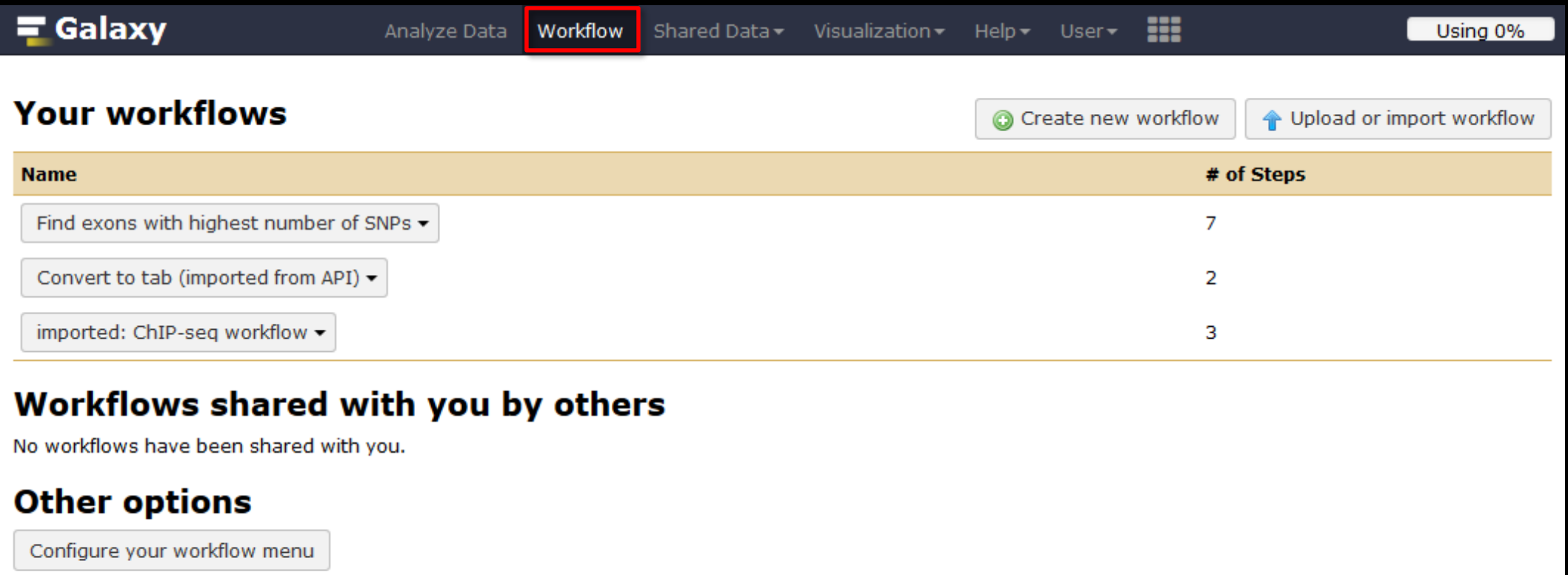
Workflow

From history



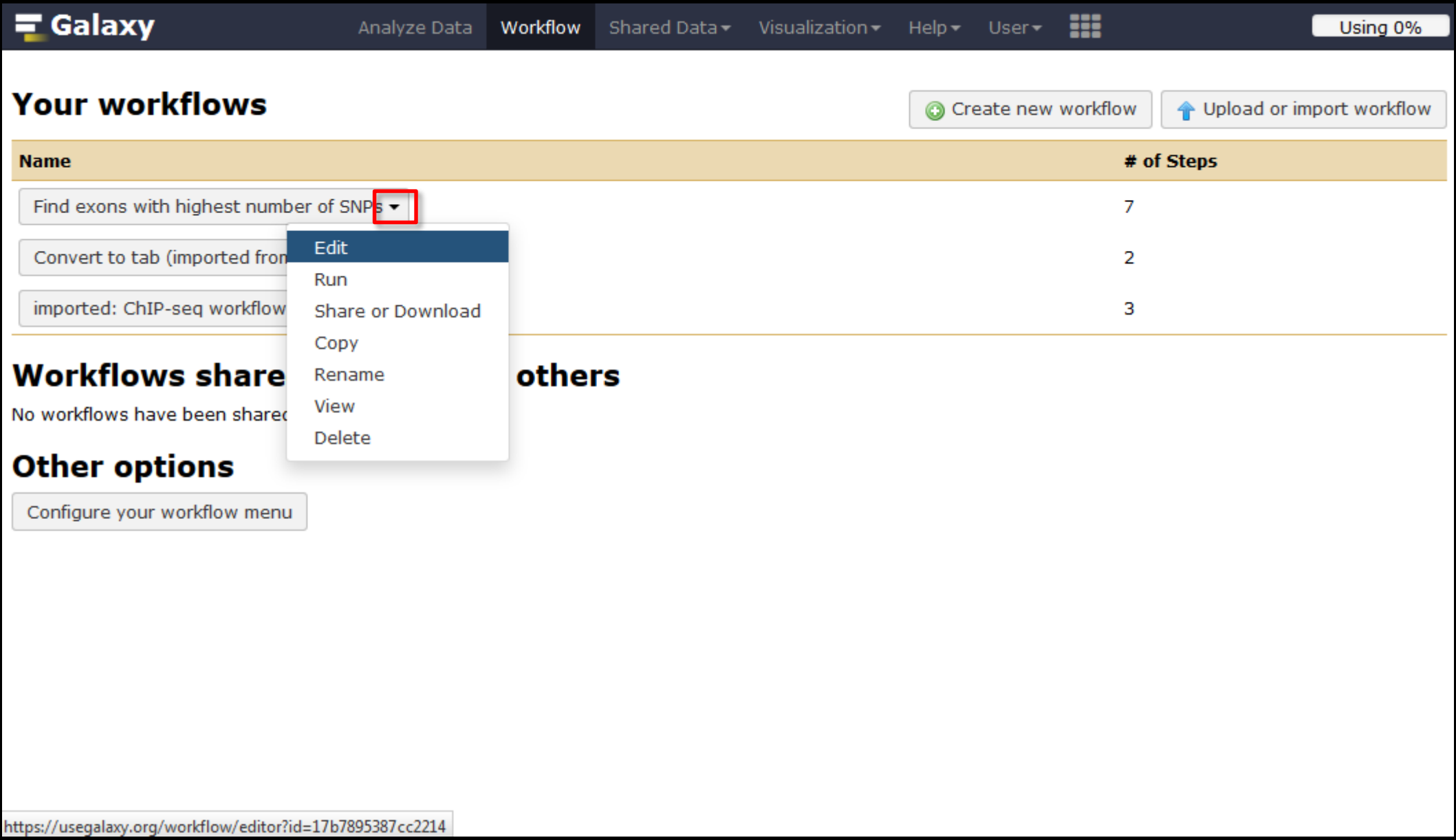
Workflow

Workflow manager



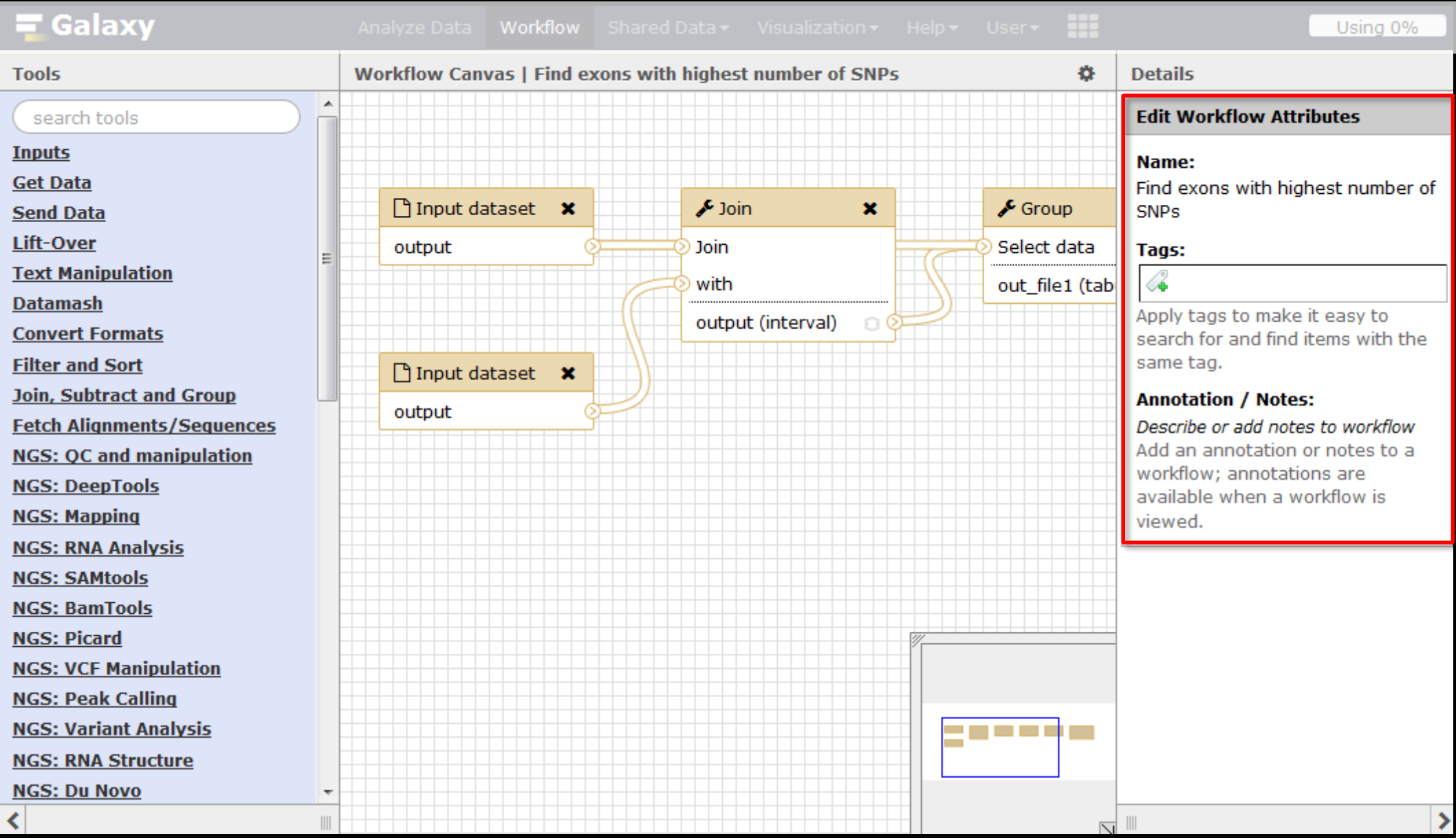
Workflow

Workflow manager



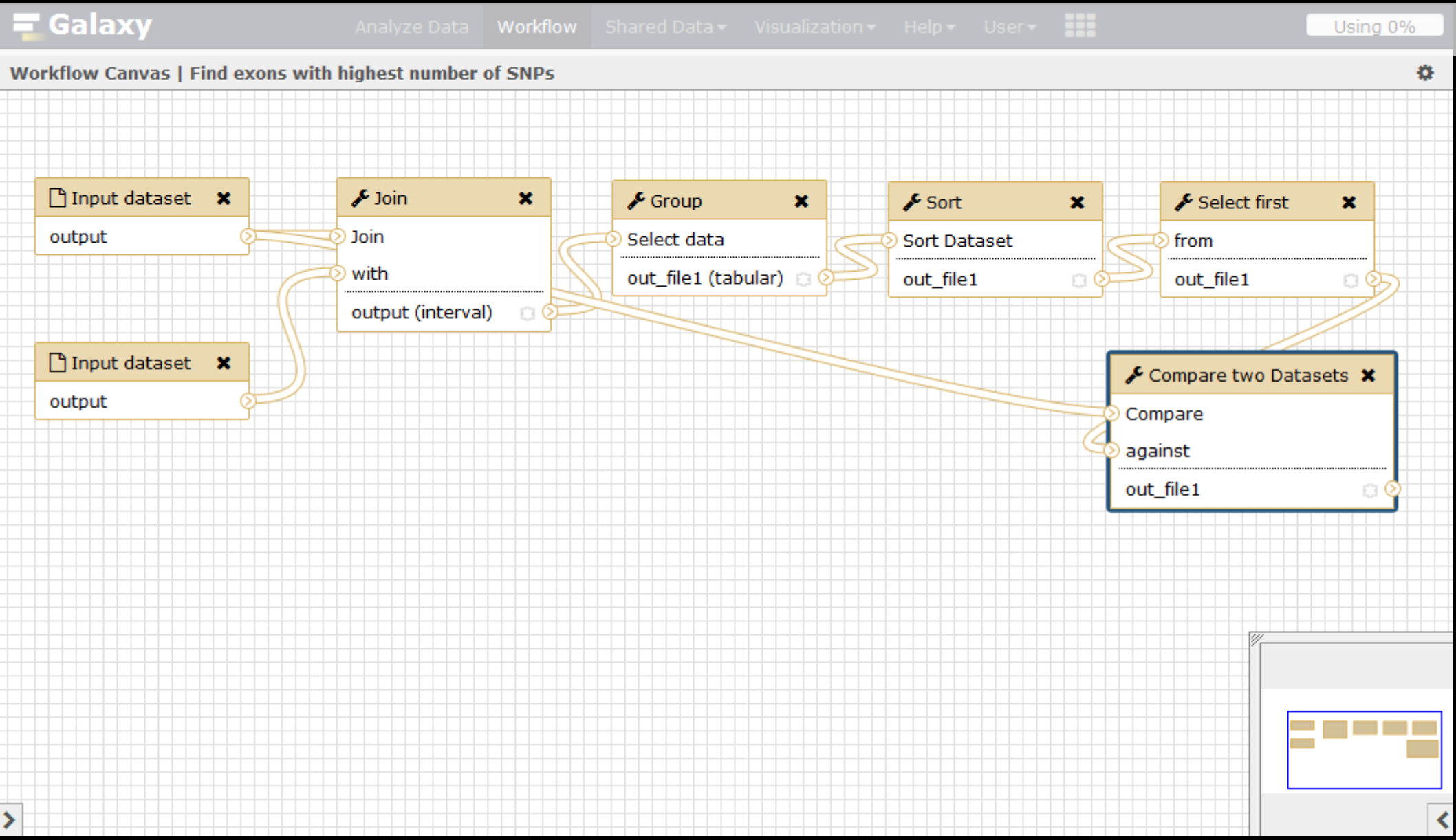
Workflow

Edit a workflow: add tags and annotation



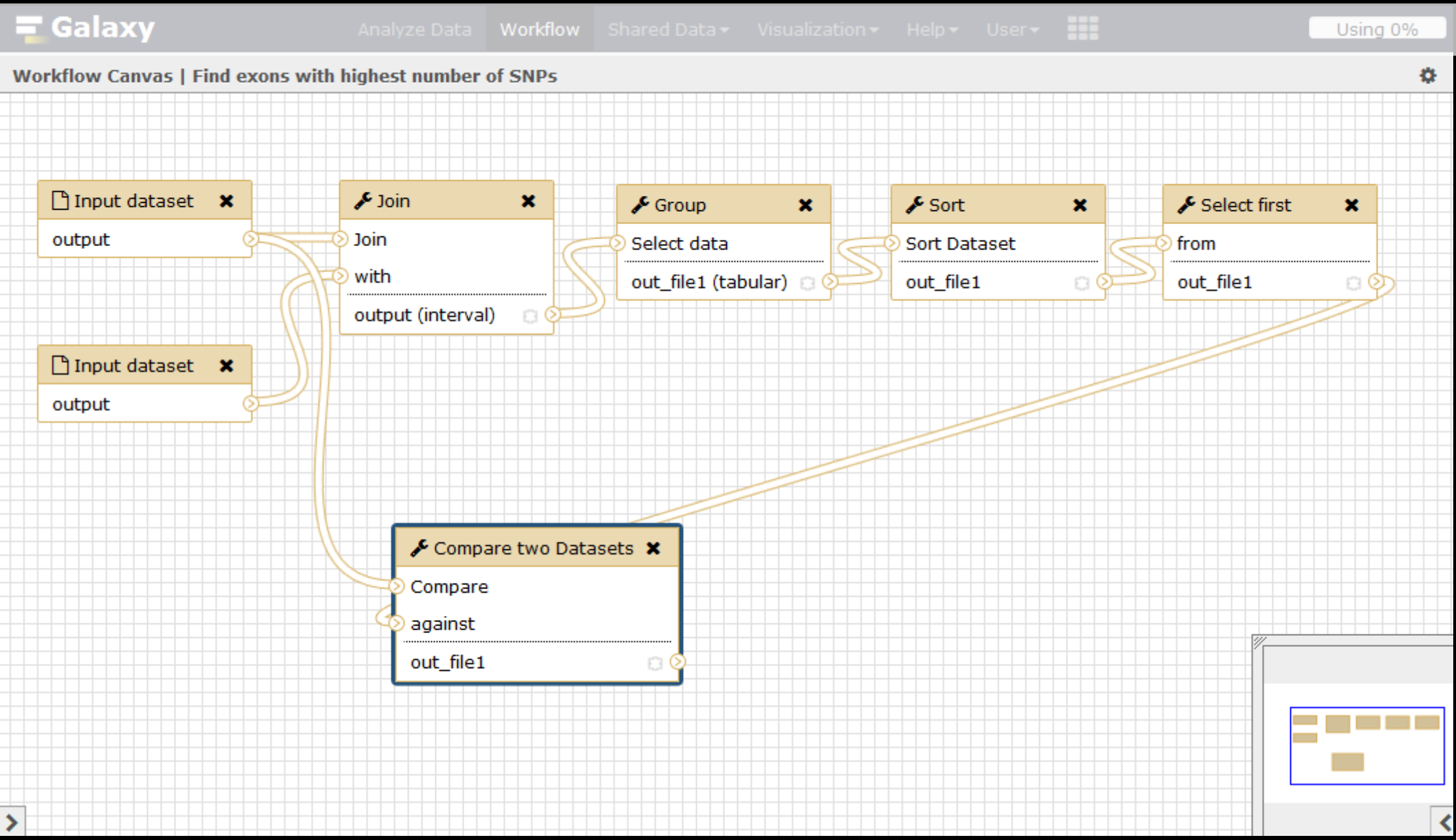
Workflow

Edit a workflow



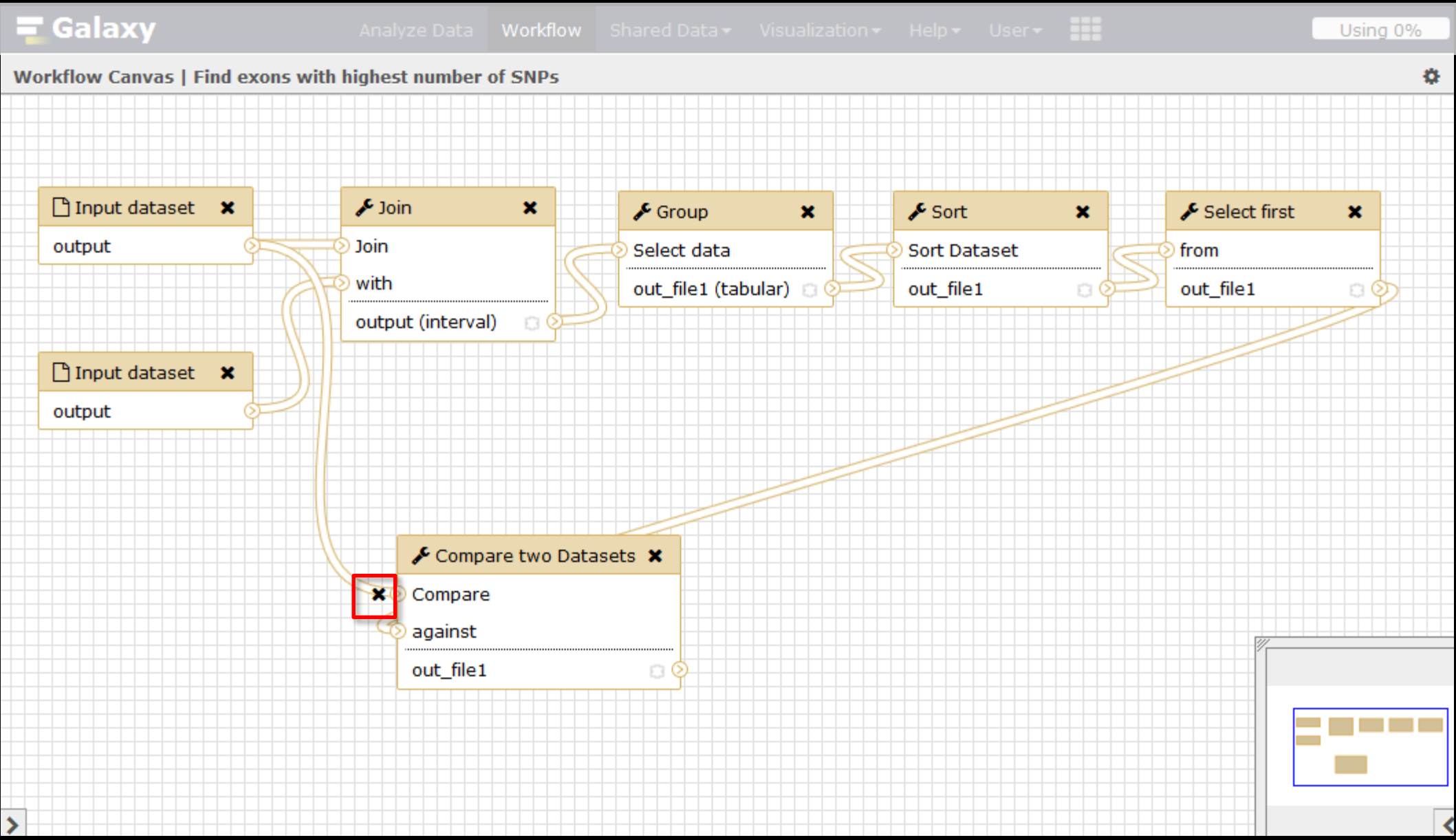
Workflow

Edit a workflow: drag and drop



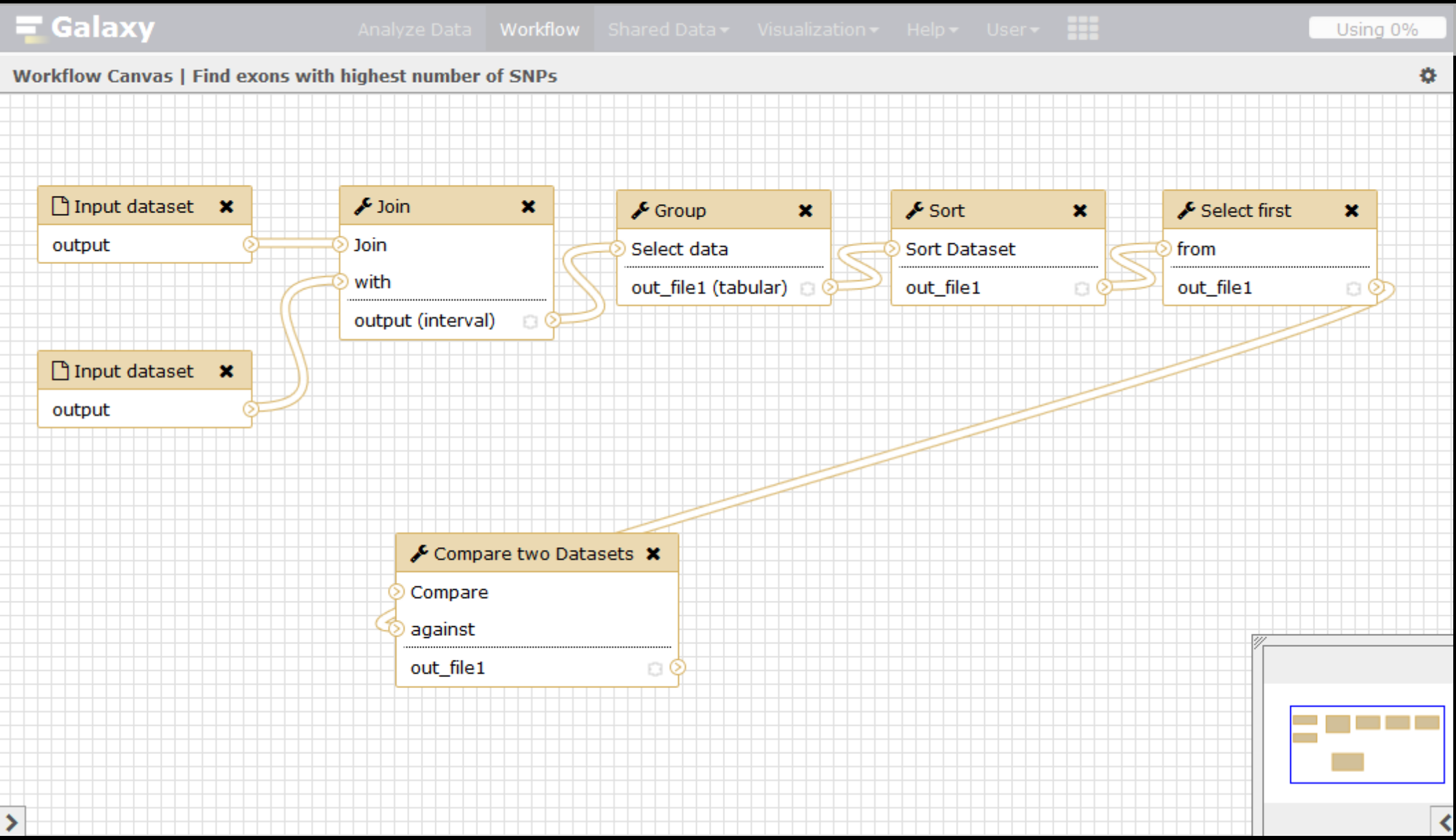
Workflow

Edit a workflow: delete a noodle



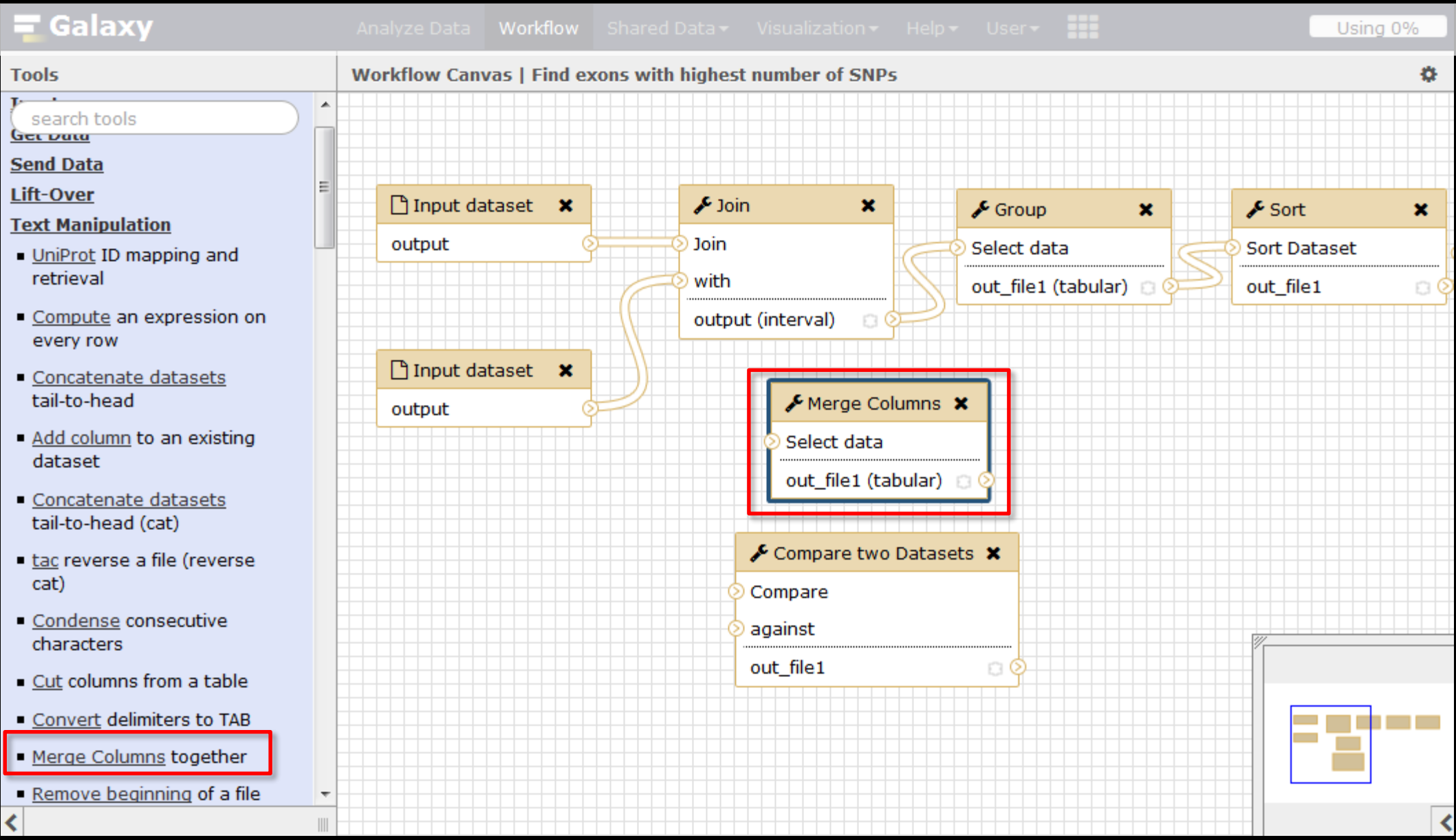
Workflow

Edit a workflow: delete a noodle



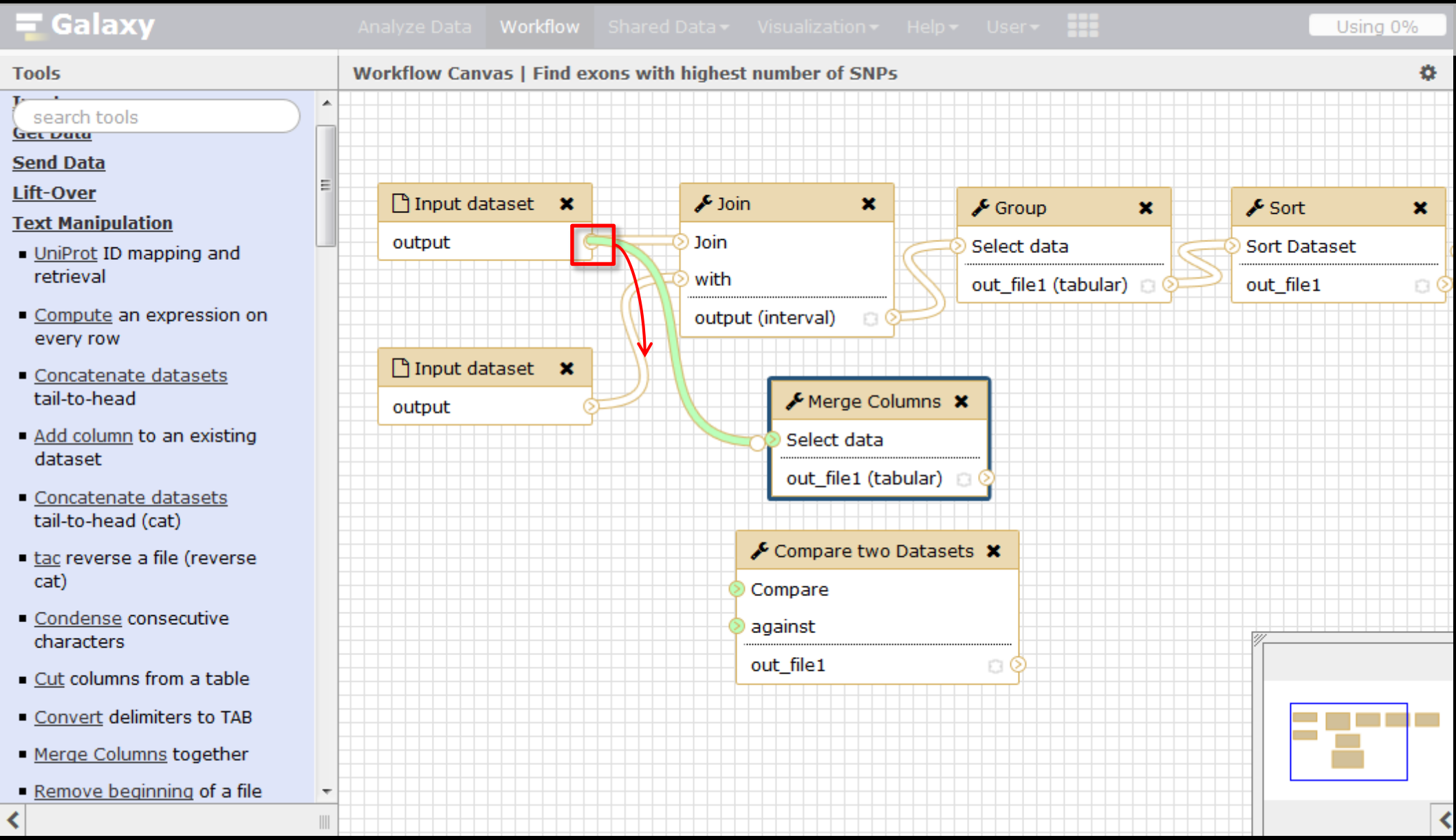
Workflow

Edit a workflow: add a tool



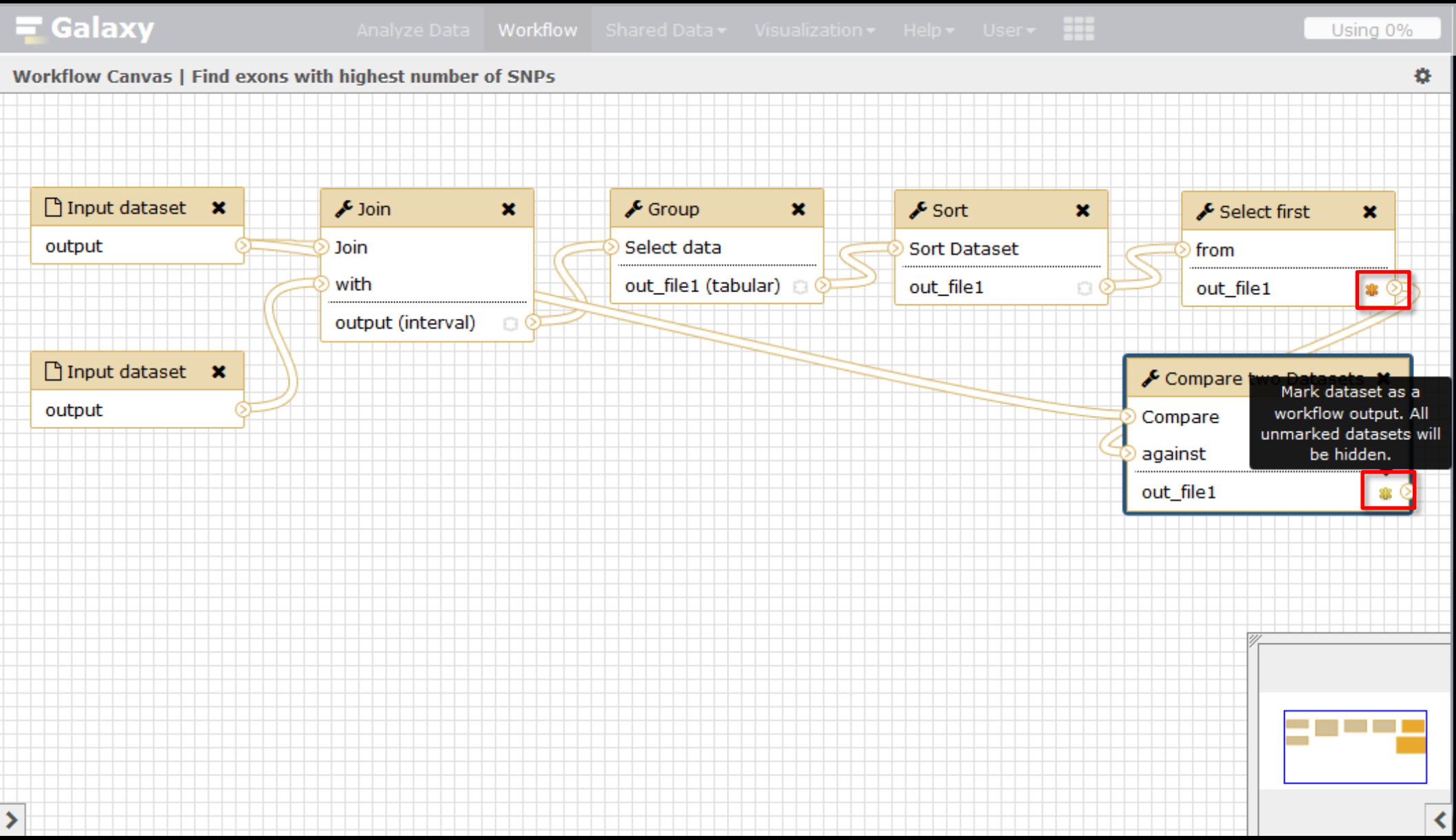
Workflow

Edit a workflow: add a noodle



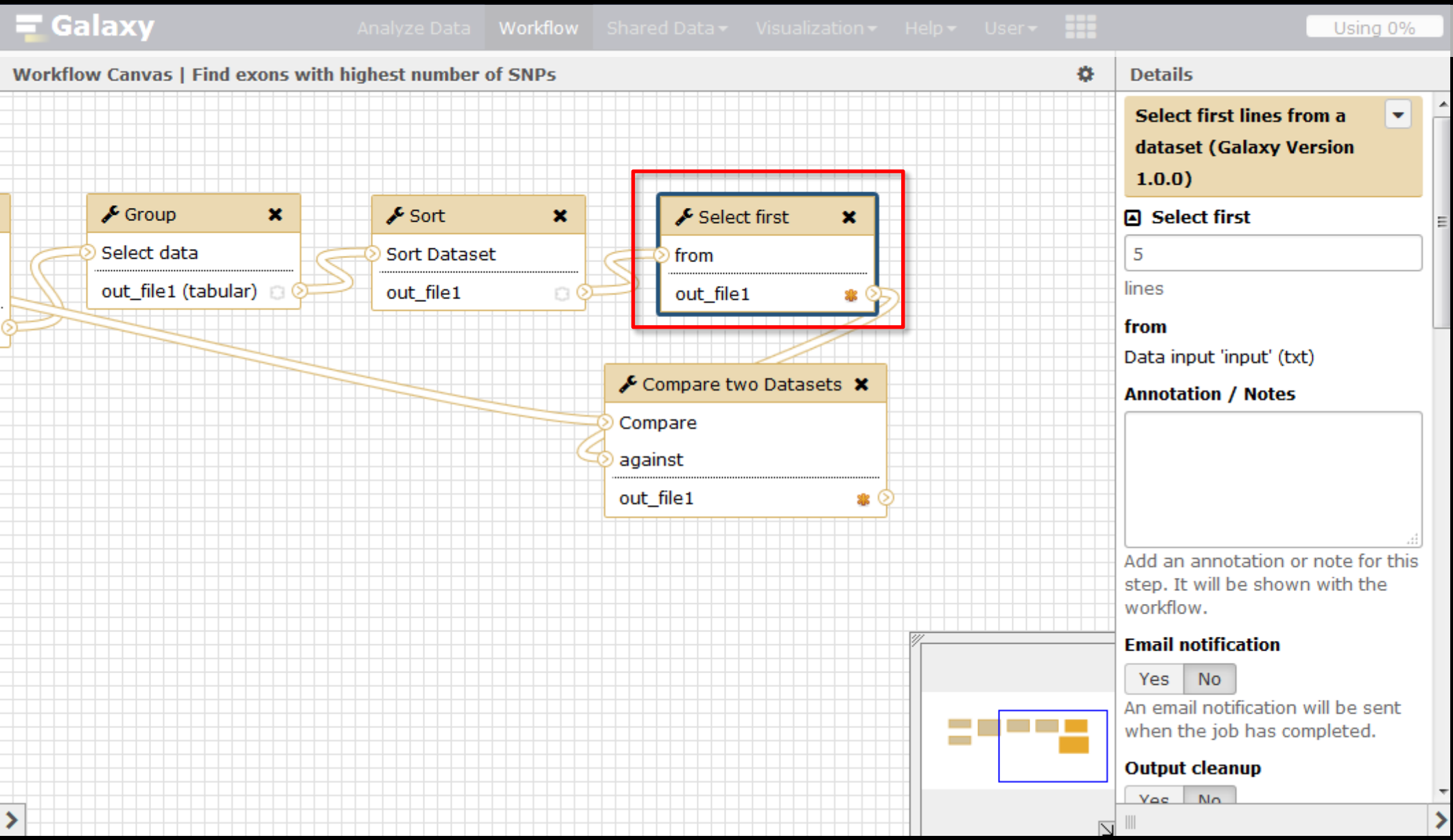
Workflow

Edit a workflow: hide intermediate steps



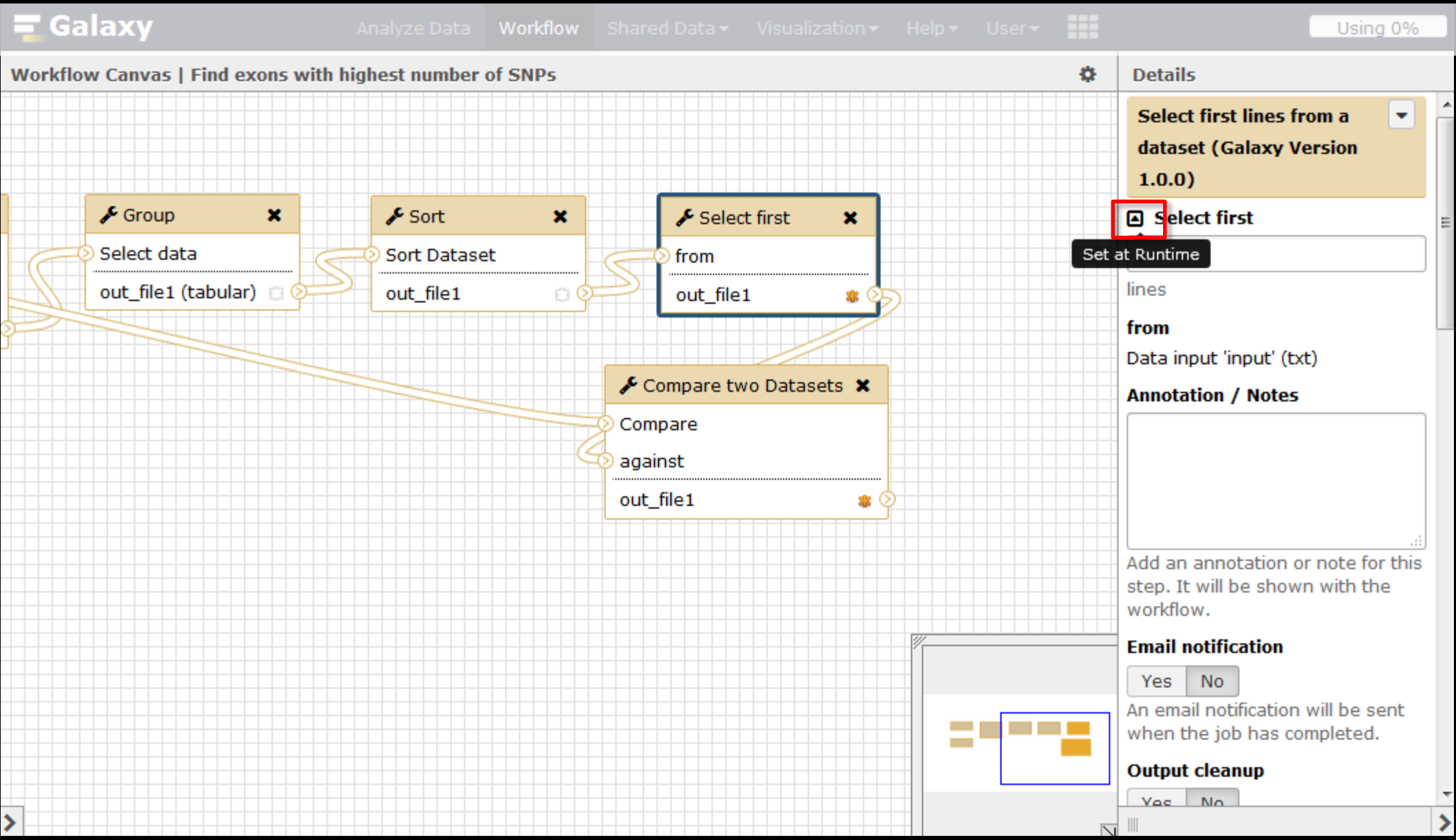
Workflow

Edit a workflow: set or release a parameter



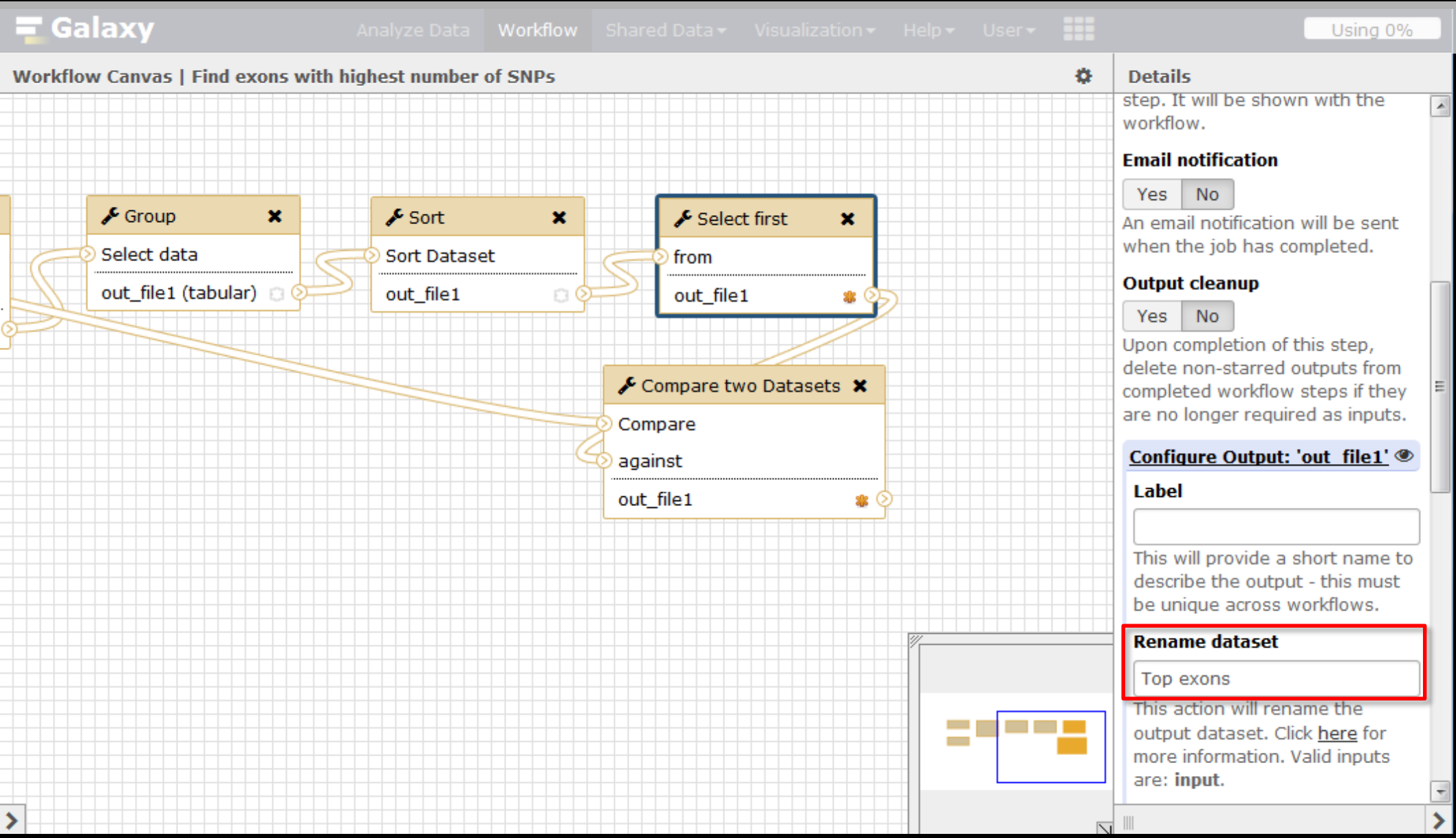
Workflow

Edit a workflow: set or release a parameter



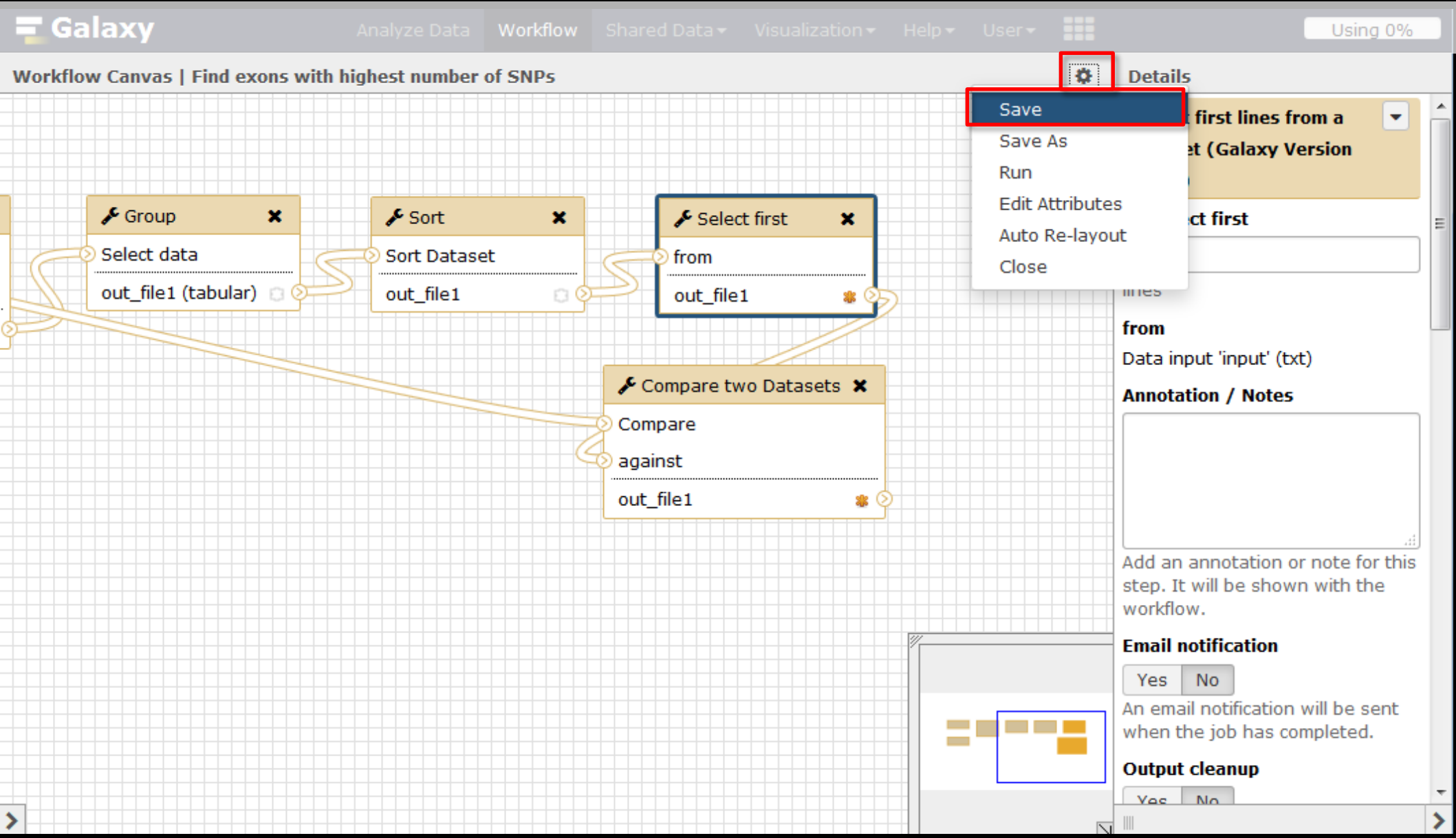
Workflow

Edit a workflow: rename the outputs



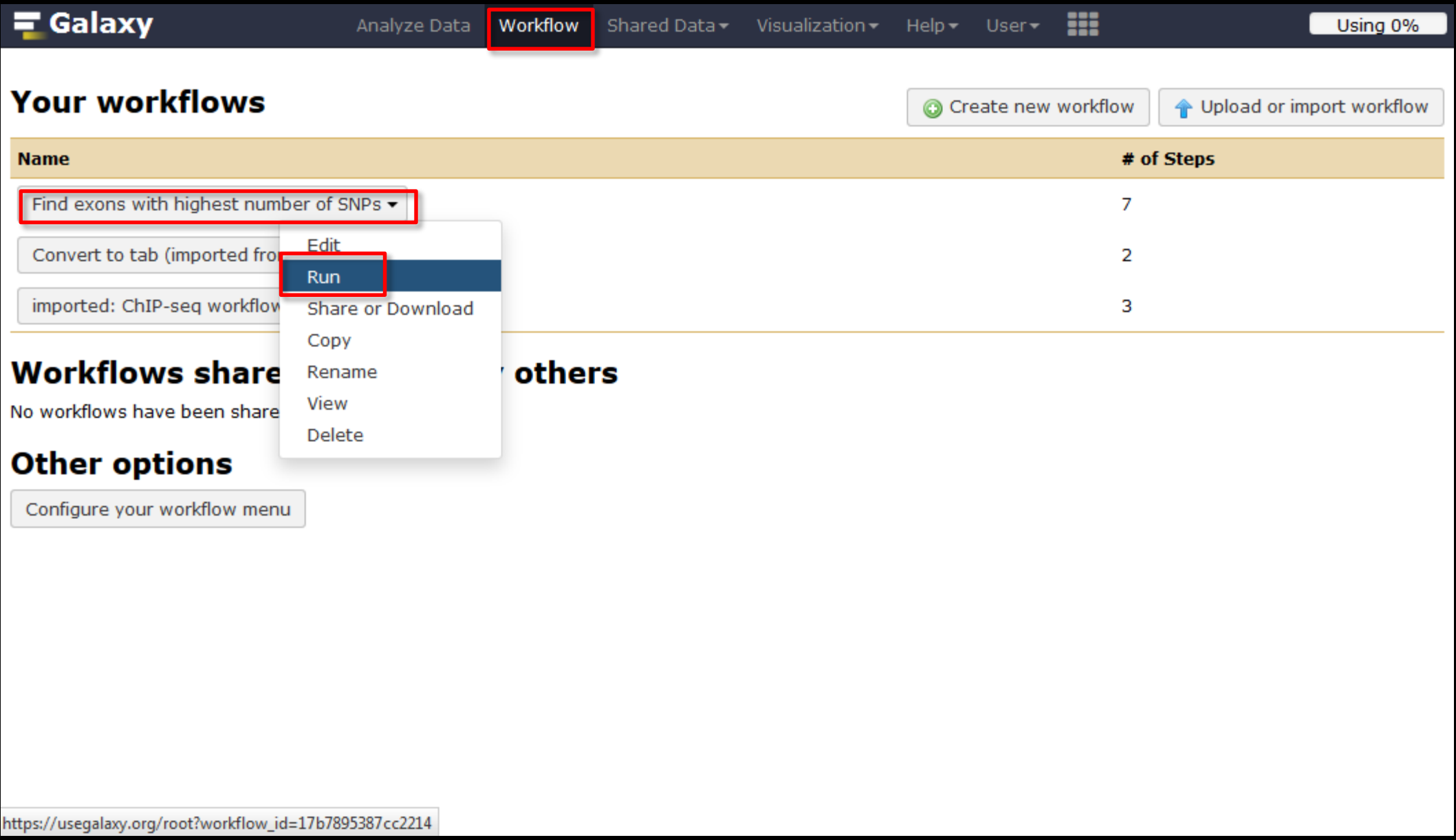
Workflow

Save



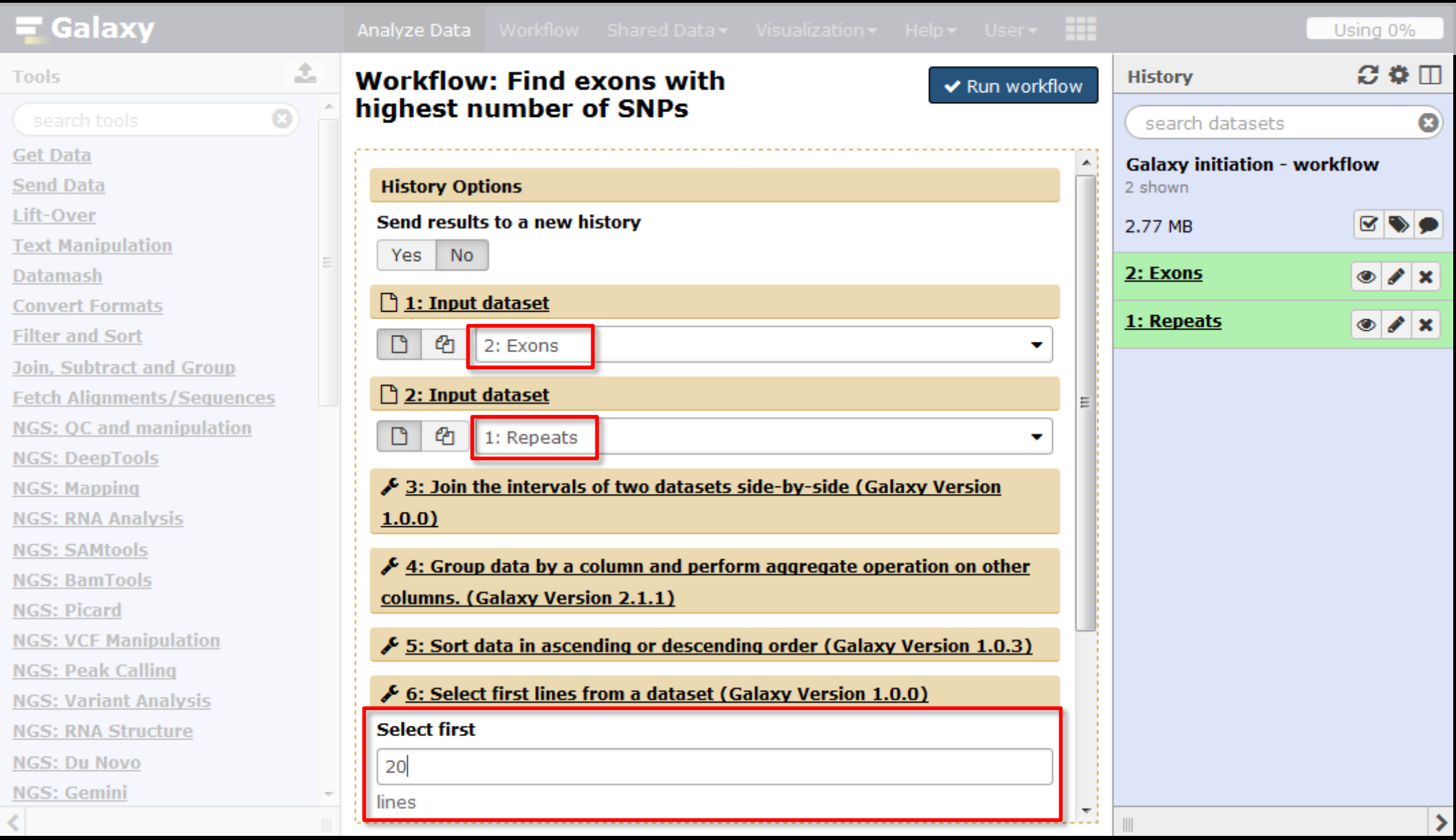
Workflow

Run a workflow



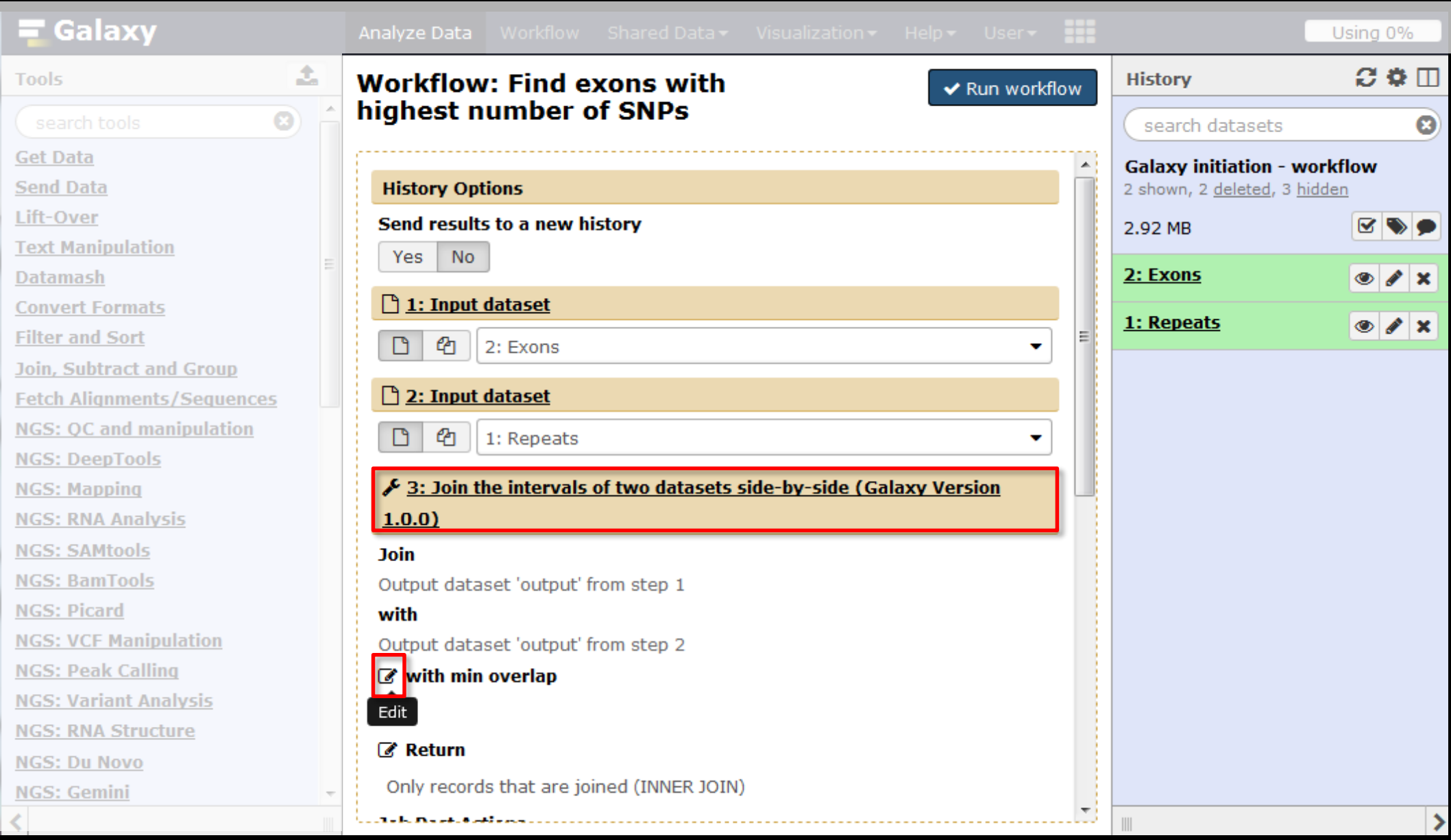
Workflow

Run a workflow



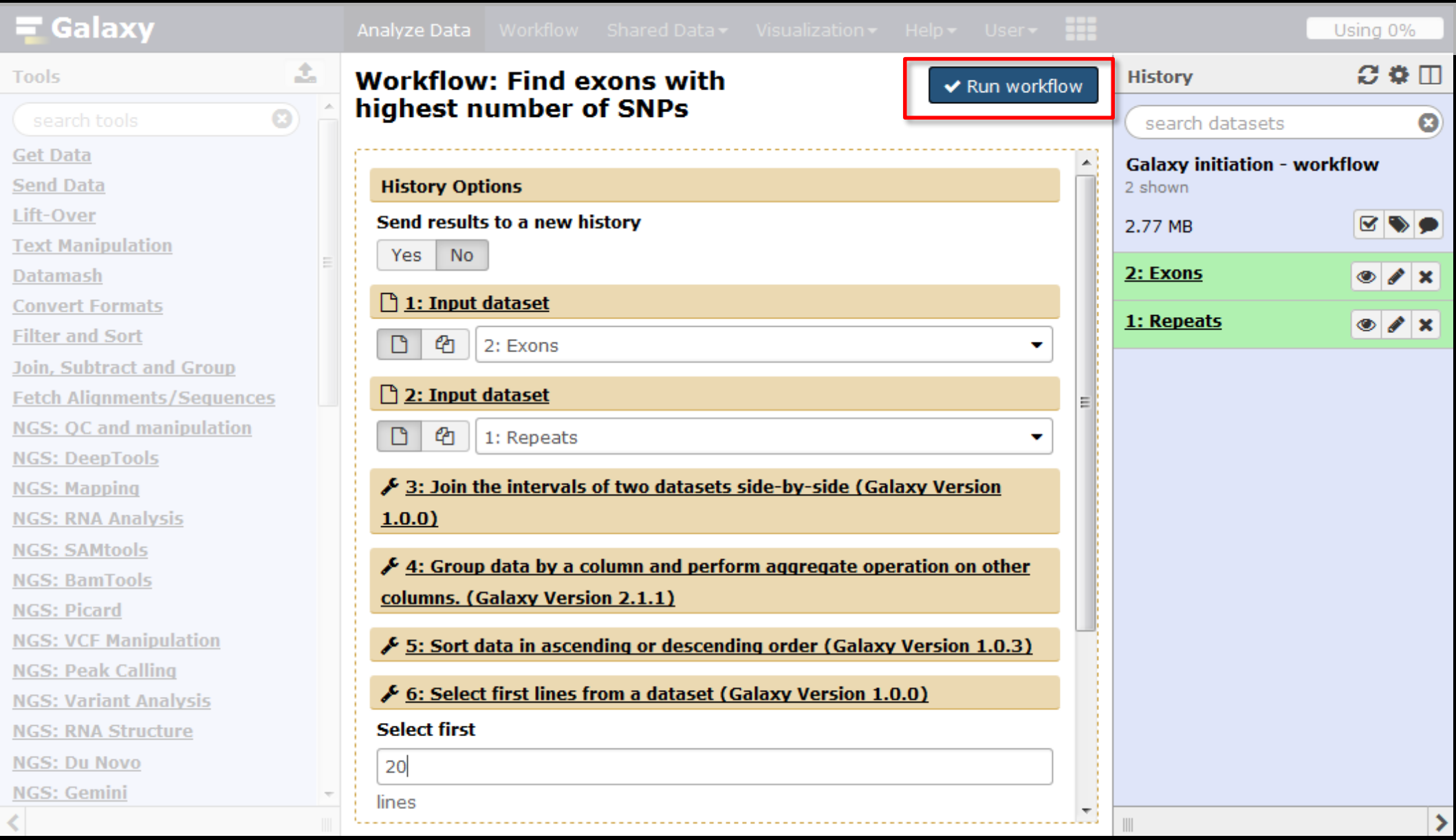
Workflow

Run a workflow



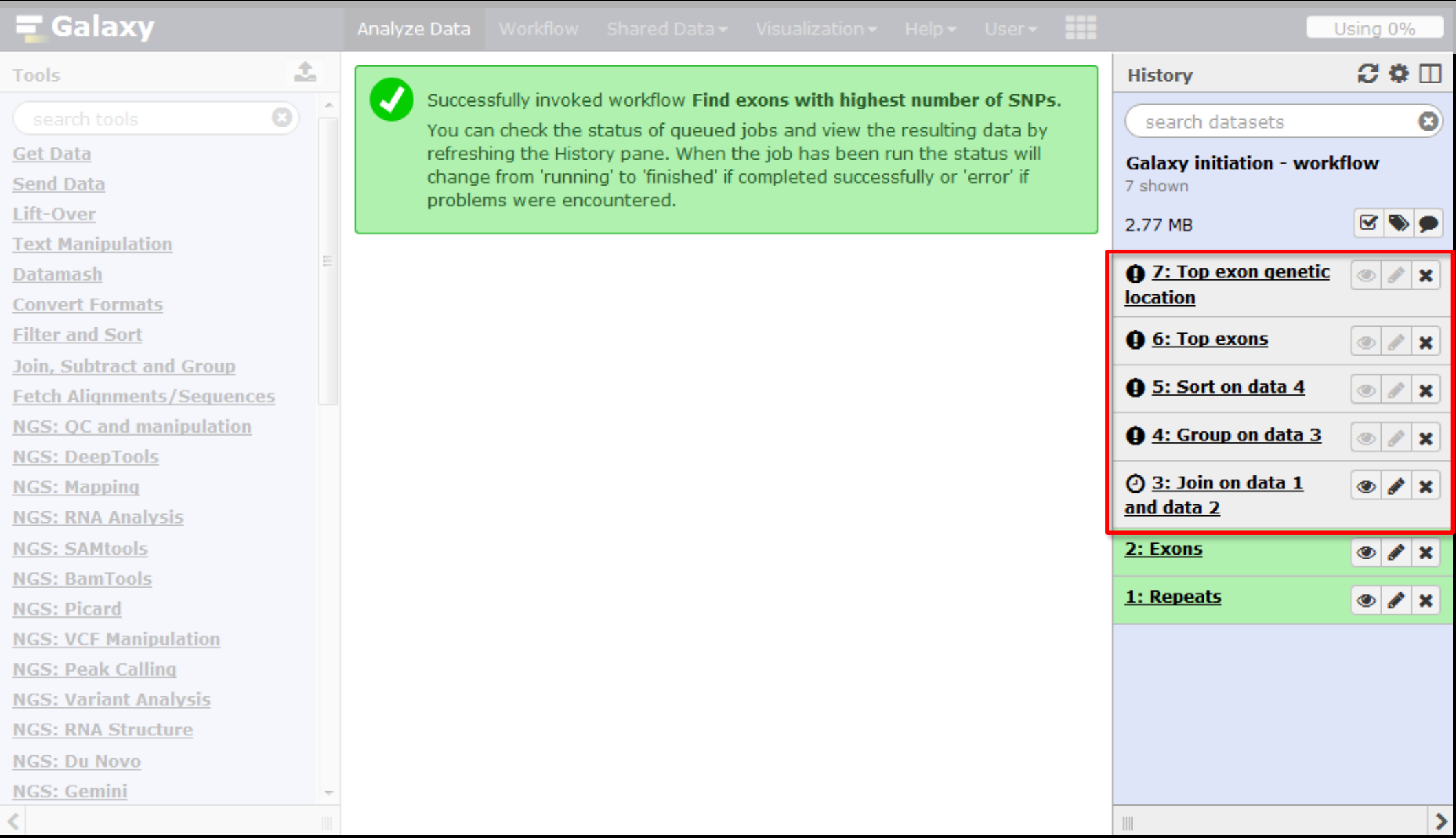
Workflow

Run a workflow



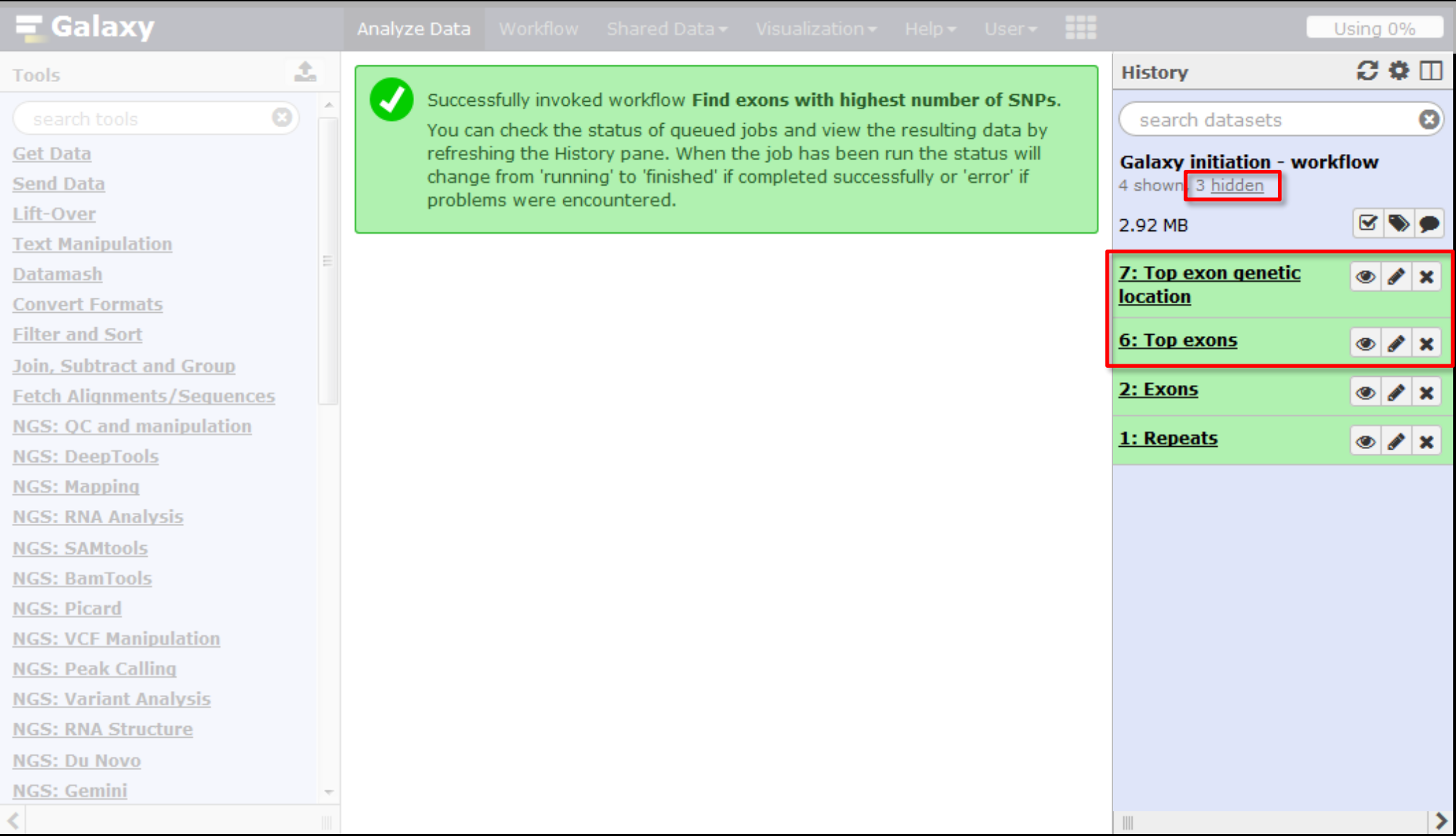
Workflow

Run a workflow



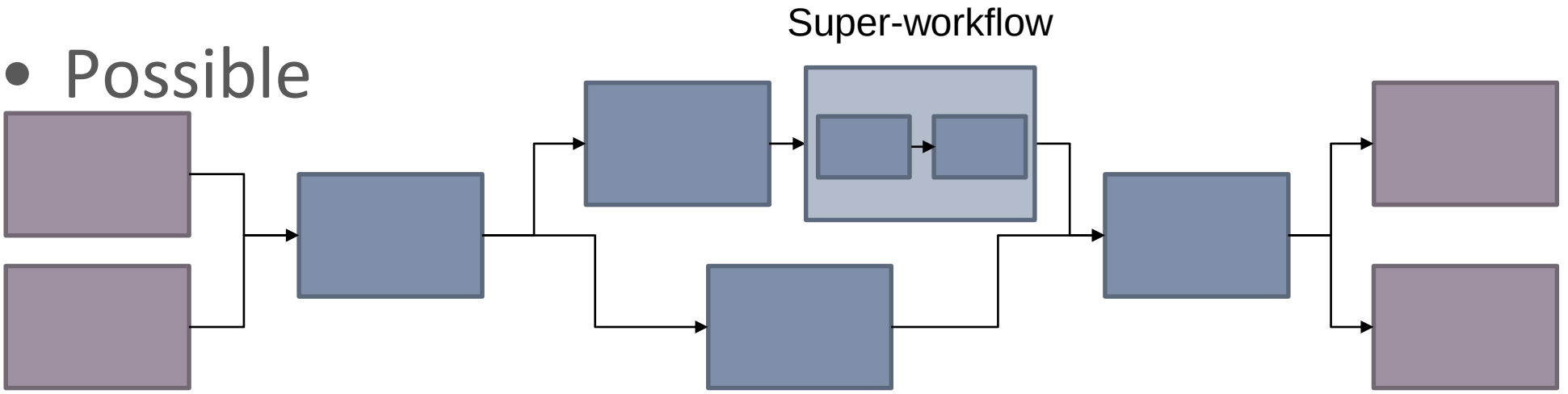
Workflow

Run a workflow

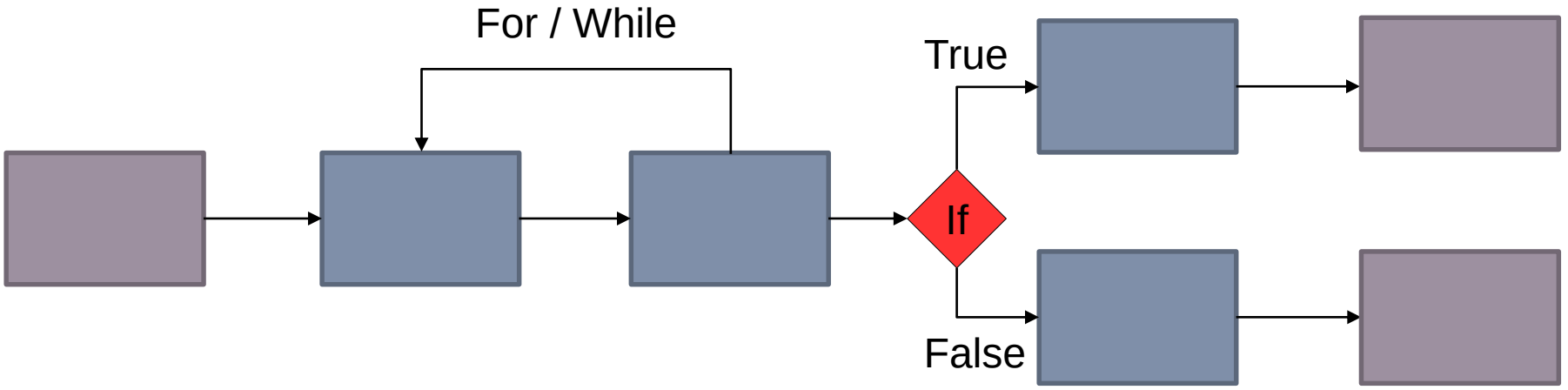


Workflow

- Possible



- Impossible (until now)



SHARE

biologist ↔ biologist

- Sharing histories or datasets
  - With or without linked workflow

bioanalyst ↔ biologist

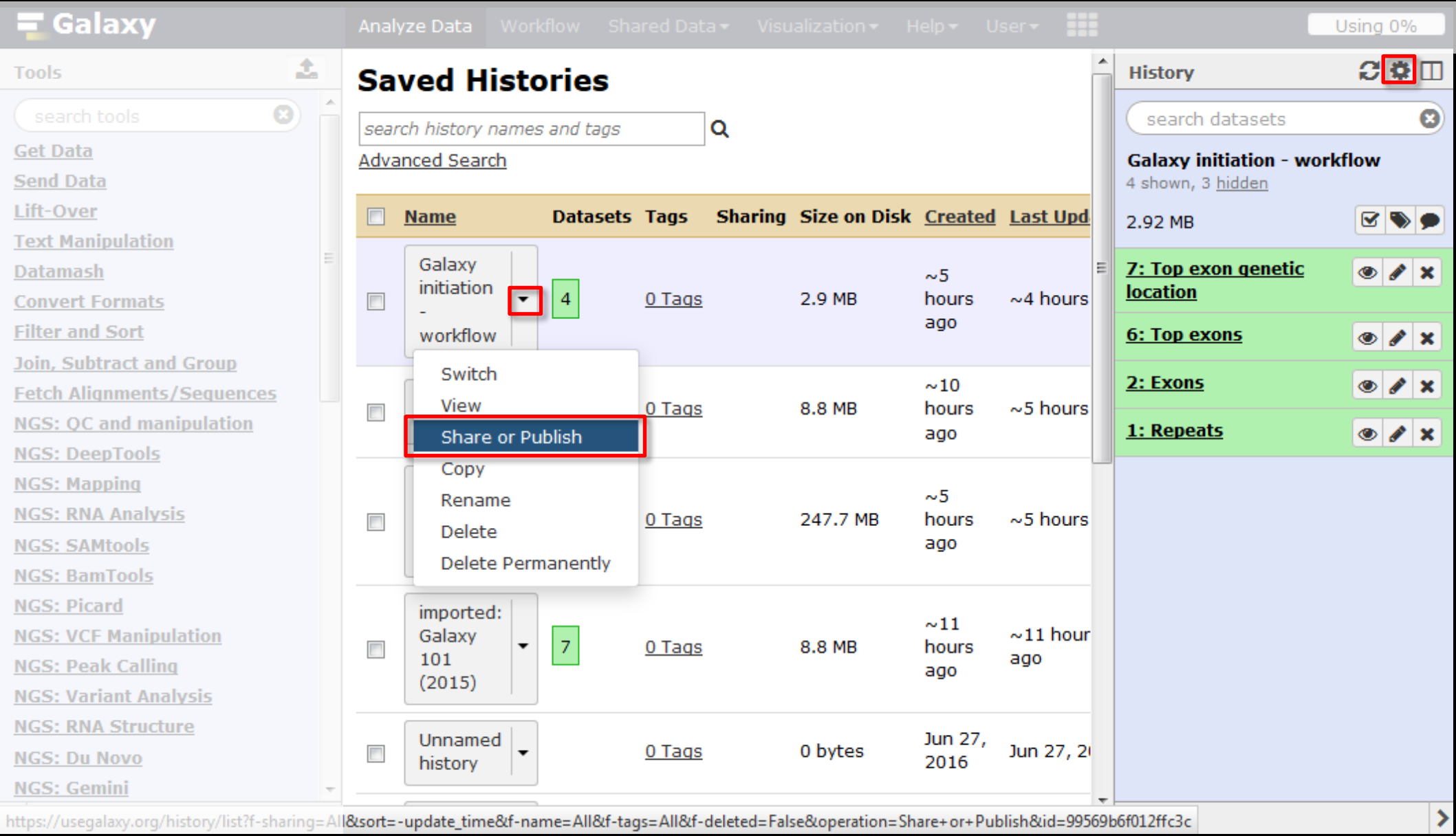
- Sharing workflows
  - Pre-configured parameters
  - With or without release parameters (set at runtime)
    - According to the user-end knowledge

bioinformatician ↔ bioinformatician

- Sharing tools ,scripts and wrappers
  - Toolshed

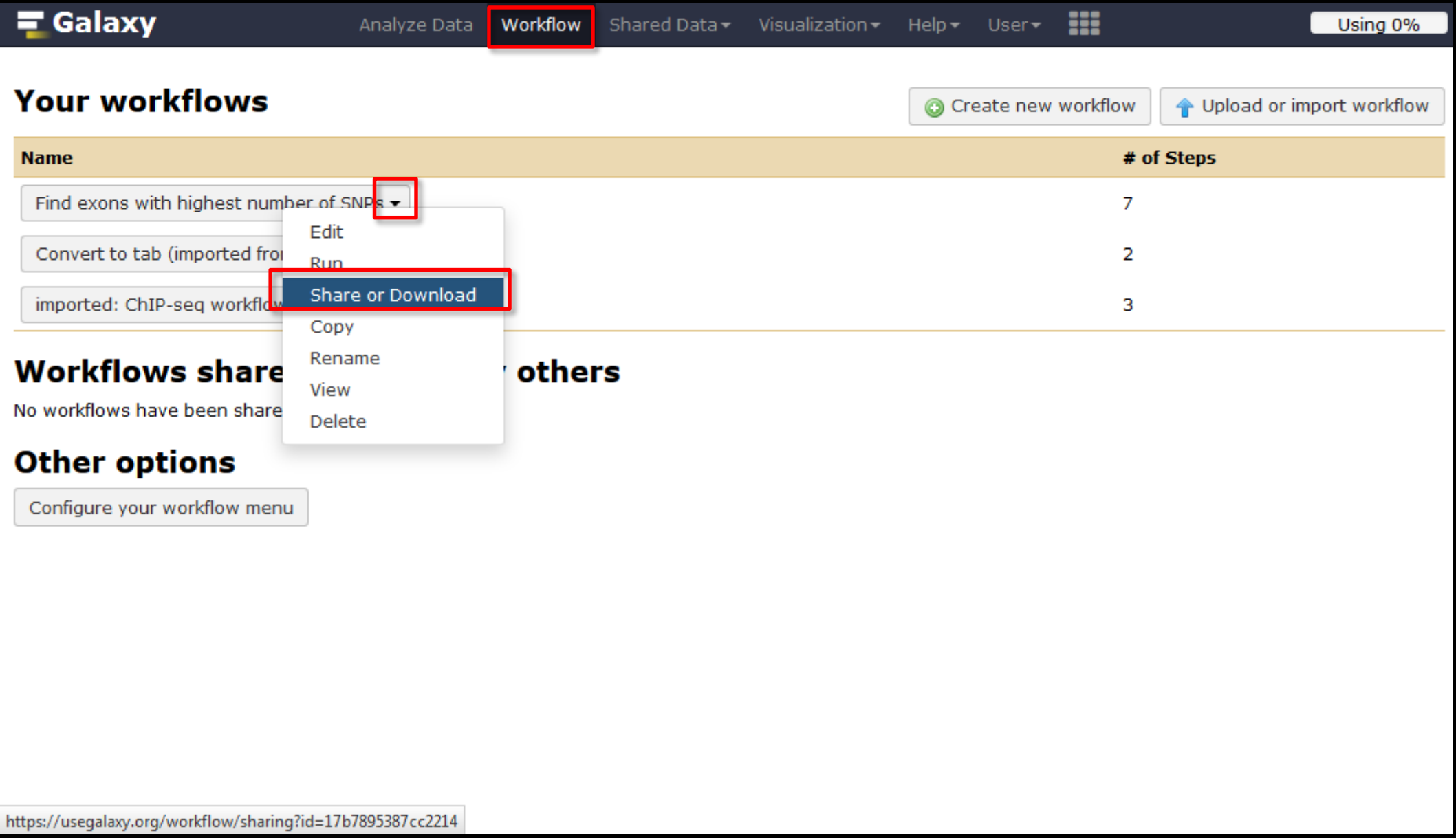
Share

History



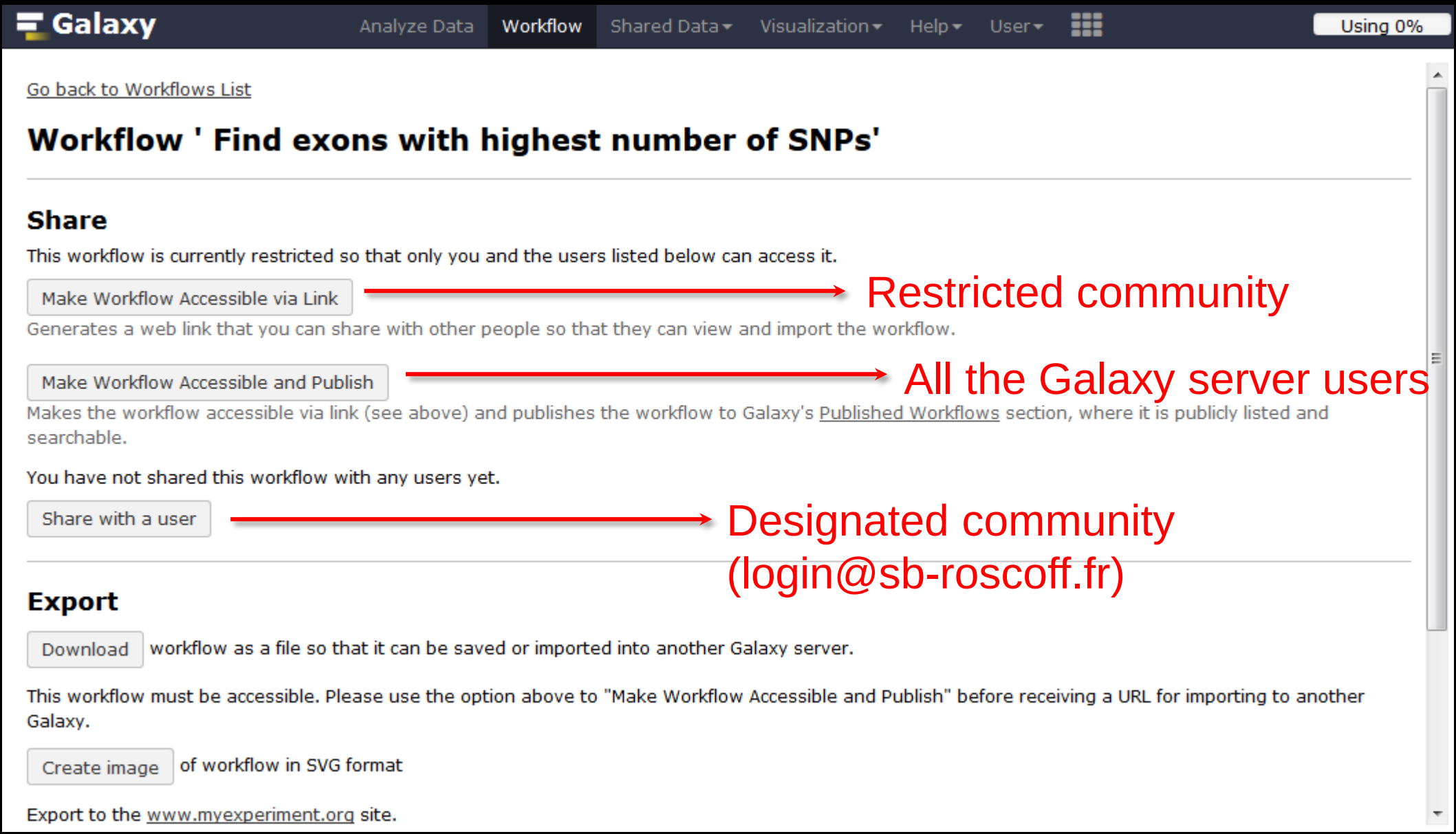
Share

Workflow



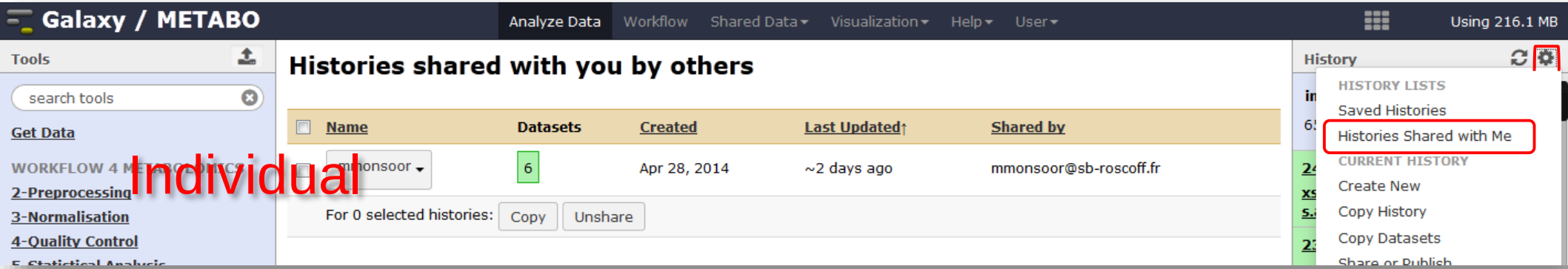
Share

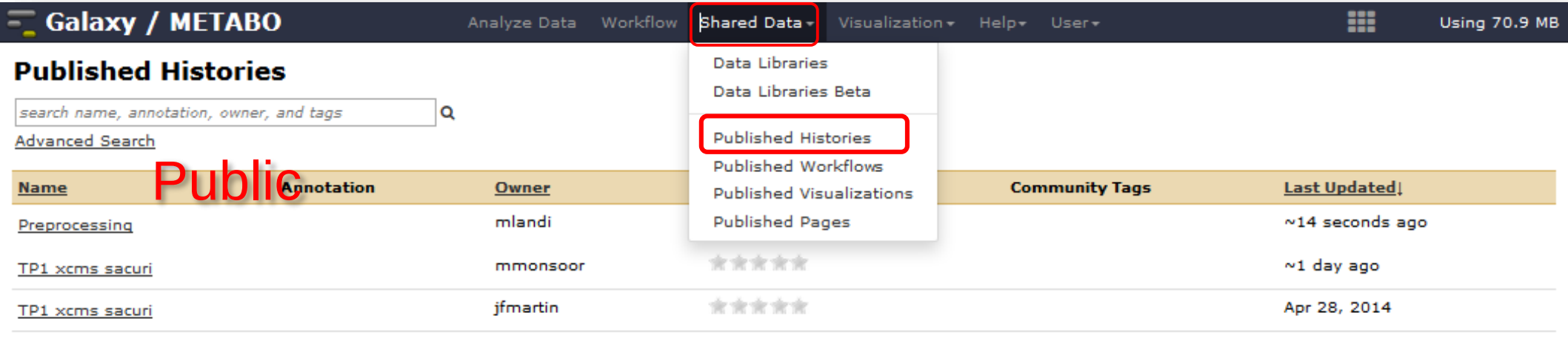
Mode


Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Help
User
Using 0%
Go back to Workflows List
Workflow ' Find exons with highest number of SNPs'
Share
This workflow is currently restricted so that only you and the users listed below can access it.
Make Workflow Accessible via Link
Restricted community
Generates a web link that you can share with other people so that they can view and import the workflow.
Make Workflow Accessible and Publish
All the Galaxy server users
Makes the workflow accessible via link (see above) and publishes the workflow to Galaxy's Published Workflows section, where it is publicly listed and searchable.
You have not shared this workflow with any users yet.
Share with a user
Designated community
(login@sb-roscoff.fr)
Export
Download
workflow as a file so that it can be saved or imported into another Galaxy server.
This workflow must be accessible. Please use the option above to "Make Workflow Accessible and Publish" before receiving a URL for importing to another Galaxy.
Create image
of workflow in SVG format
Export to the www.myexperiment.org site.

Share

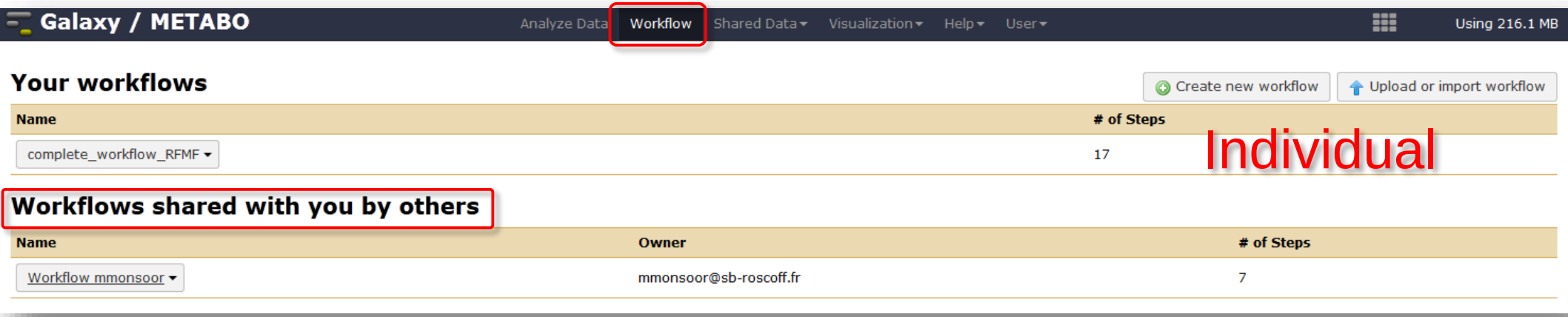
- Get shared histories

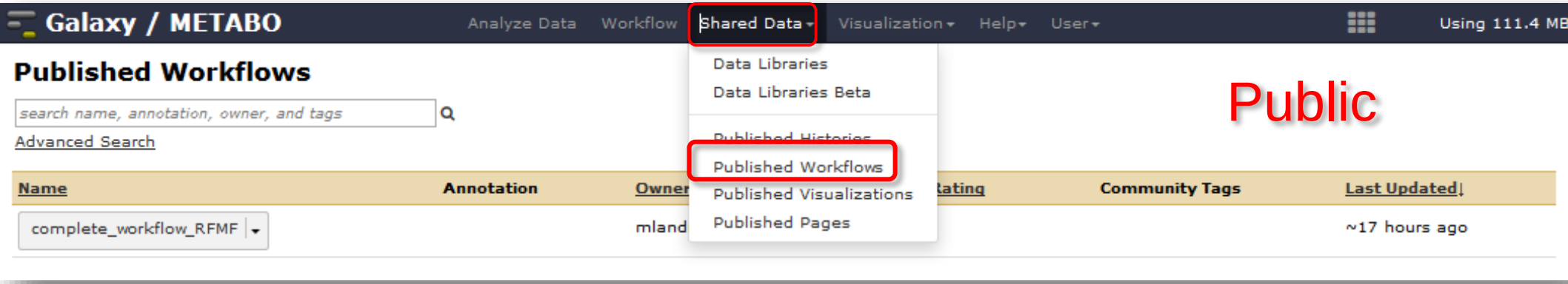




Share

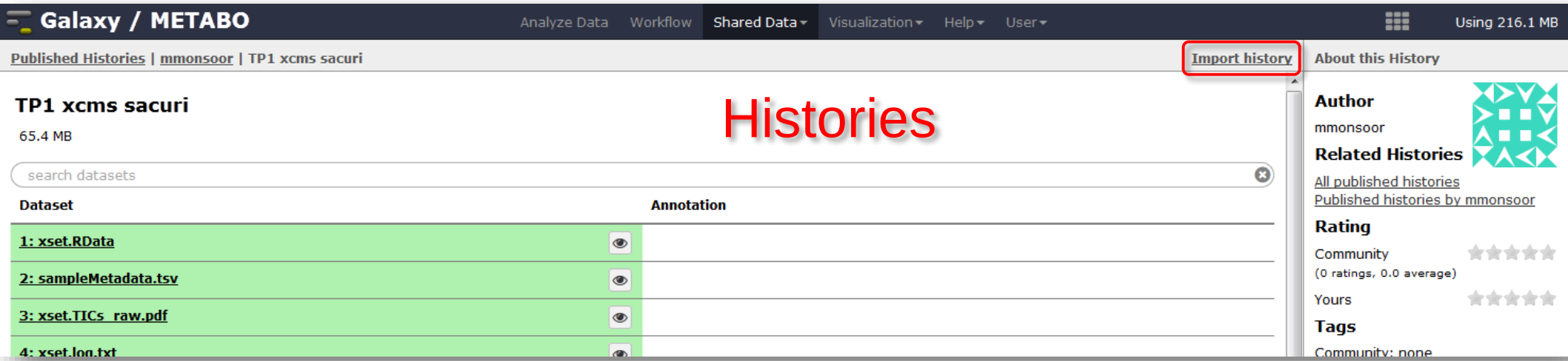
- Get shared workflows

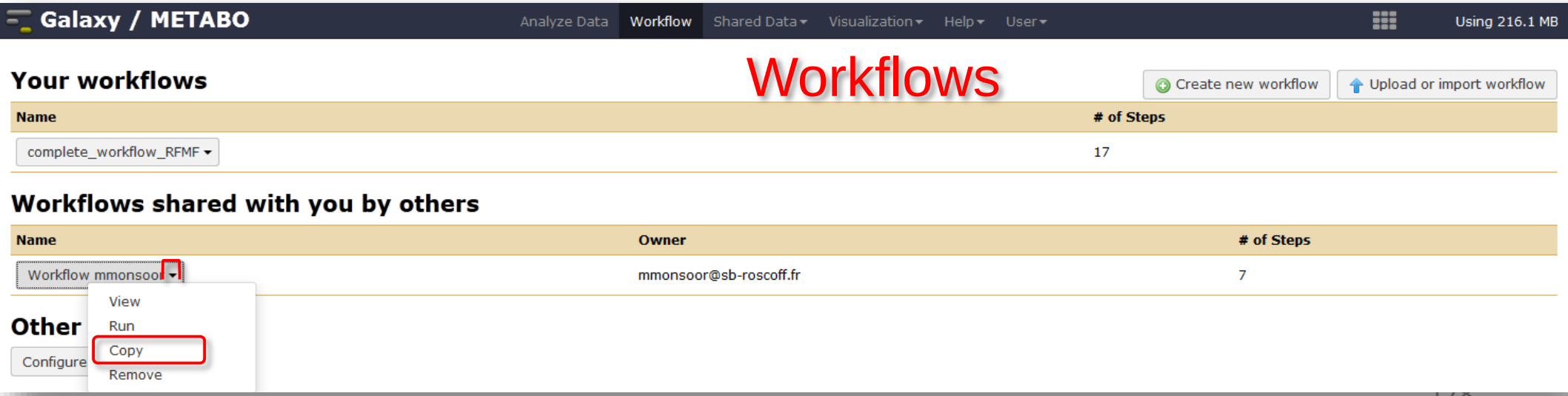




Share

- Import shared





Share

Level 5

- Share tools and descriptions in the ToolShed

Level 4

- Launch tools autonomously
- Use advanced parameters
- Use the Galaxy API
- Provide workflow for colleagues Level 1-3

Level 3

- Launch tools autonomously
- Use workflow more or less preset

Level 2

- Use preset workflow

Level 1

- Share his data to collegues Level 2-5

Hands-on

WORKFLOW

Workflow – Hands-on

Which coding exon has the highest number of ~~single nucleotide polymorphisms~~ on chromosome 22?

repeats

1. Extract a workflow from your history
2. Edit the workflow (hide intermediate steps, rename inputs/outputs, set parameters at runtime, save)
3. Create a new history with the input data
4. Run the workflow
5. Share your history/workflow with your neighbour

Collection

DATASET

Dataset collection

- Problematic: you have a large numbers of datasets to send through the same analysis

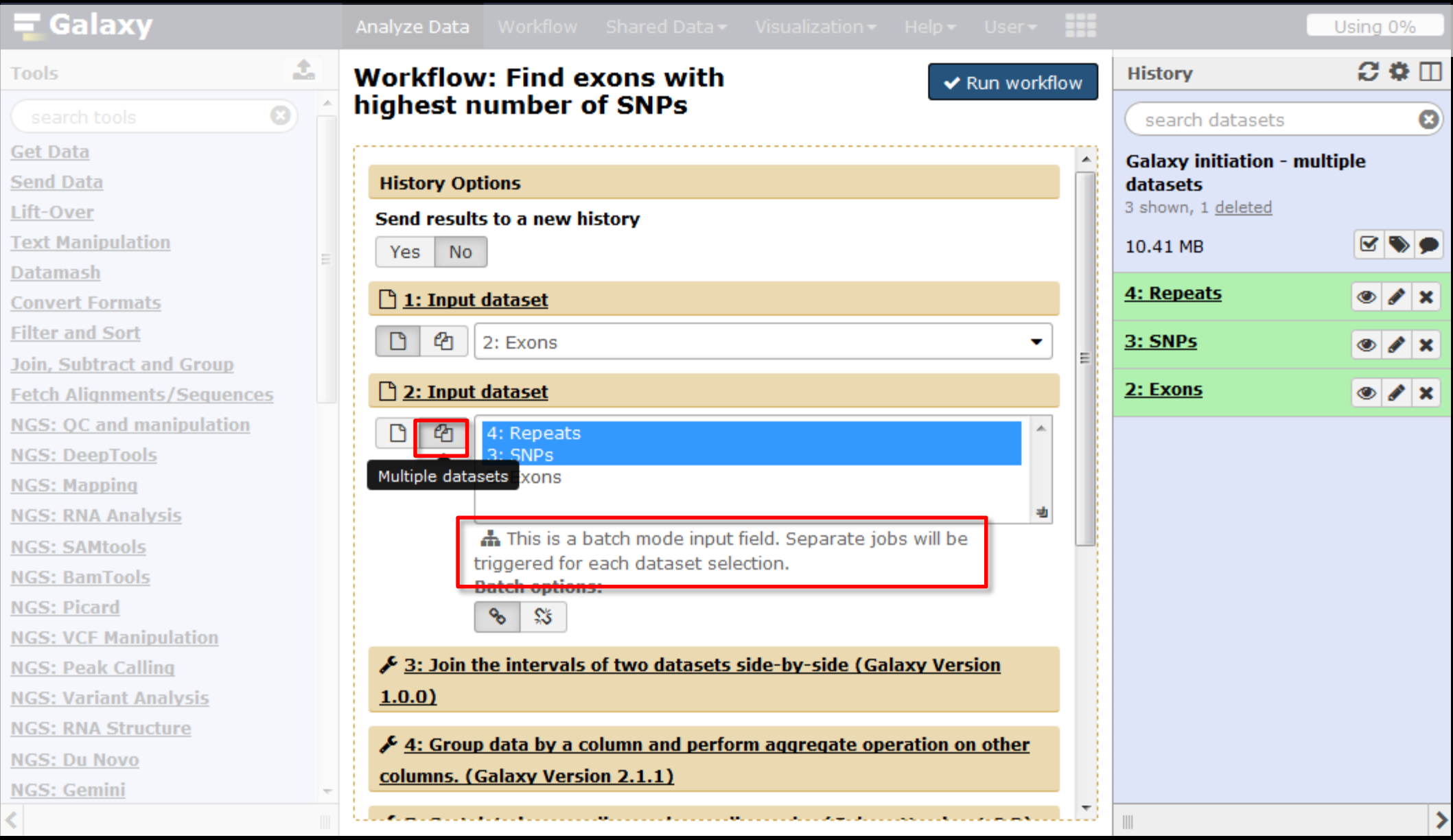
Dataset collection

- Problematic: you have a large numbers of datasets to send through the same analysis

- Solution 1: select multiple datasets as input

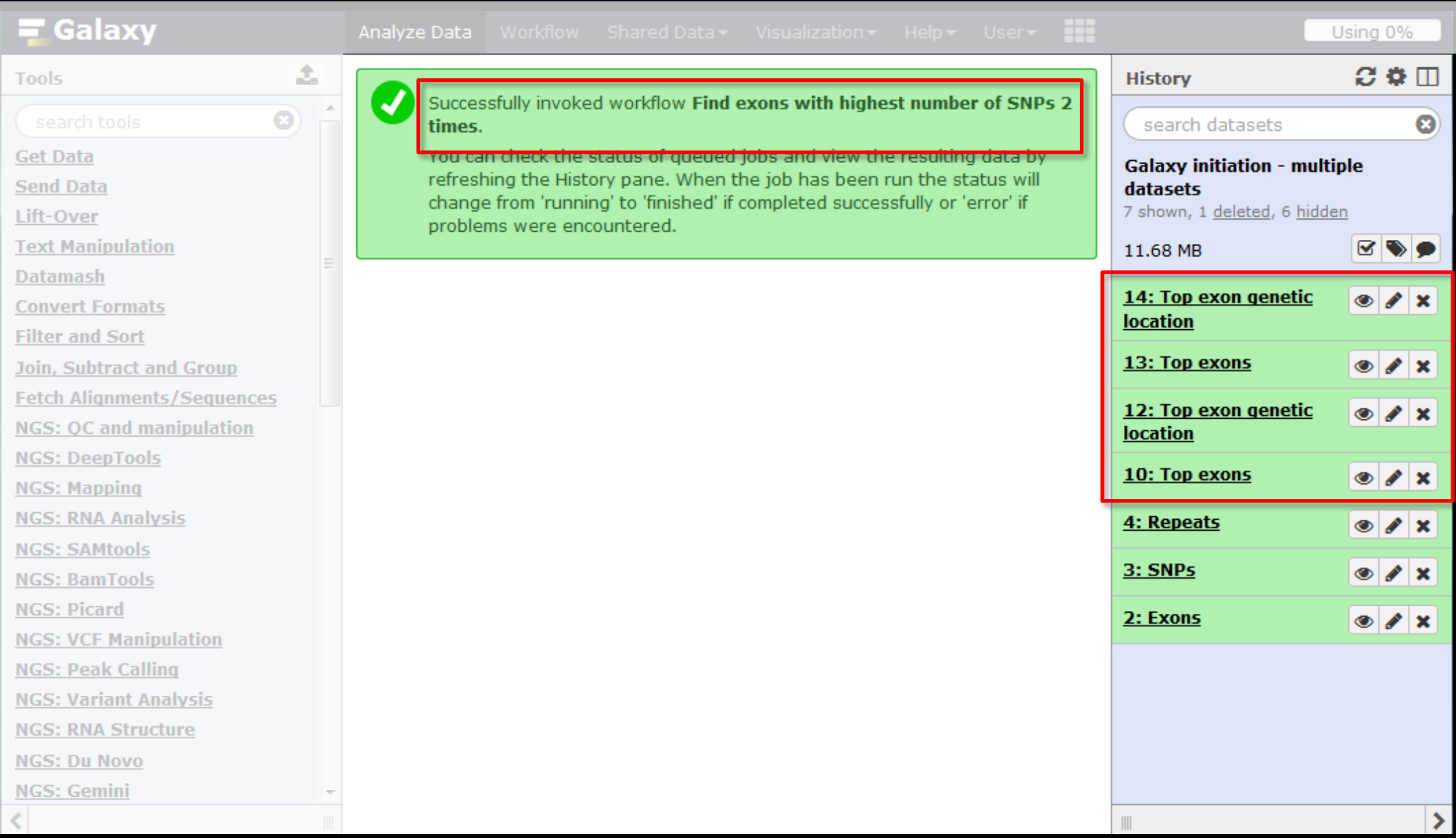
Dataset collection

Select multiple datasets as input



Dataset collection

Select multiple datasets as input



Dataset collection

- Problematic: you have a large numbers of datasets to send through the same analysis

- Solution 1: select multiple datasets as input
- Solution 2: create a dataset collection

Dataset collection

- Problematic: you have a large numbers of datasets to send through the same analysis

- Solution 1: select multiple datasets as input
- Solution 2: create a dataset collection
  - Dataset list: set of files of the same type
  - Dataset pairs: pairs of read files (forward, reverse)
  - List of dataset pairs

Dataset collection

- Problematic: you have a large numbers of datasets to send through the same analysis

- Solution 1: select multiple datasets as input
- Solution 2: create a dataset collection
  - Dataset list: set of files of the same type
  - Dataset pairs: pairs of read files (forward, reverse)
  - List of dataset pairs

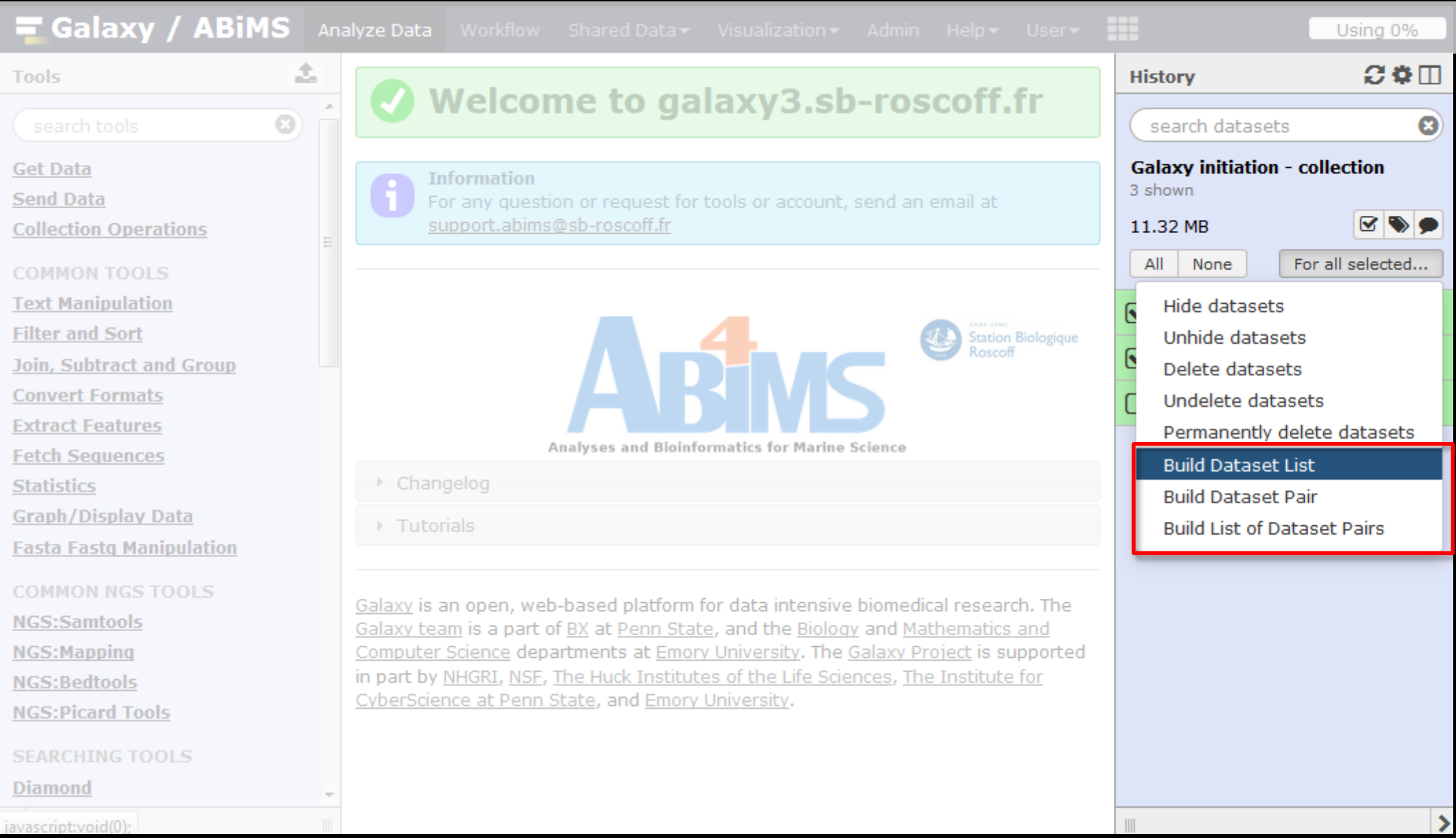
- Galaxy runs the tool automatically on each dataset in the collection using the same settings

Dataset collection

Create a dataset collection

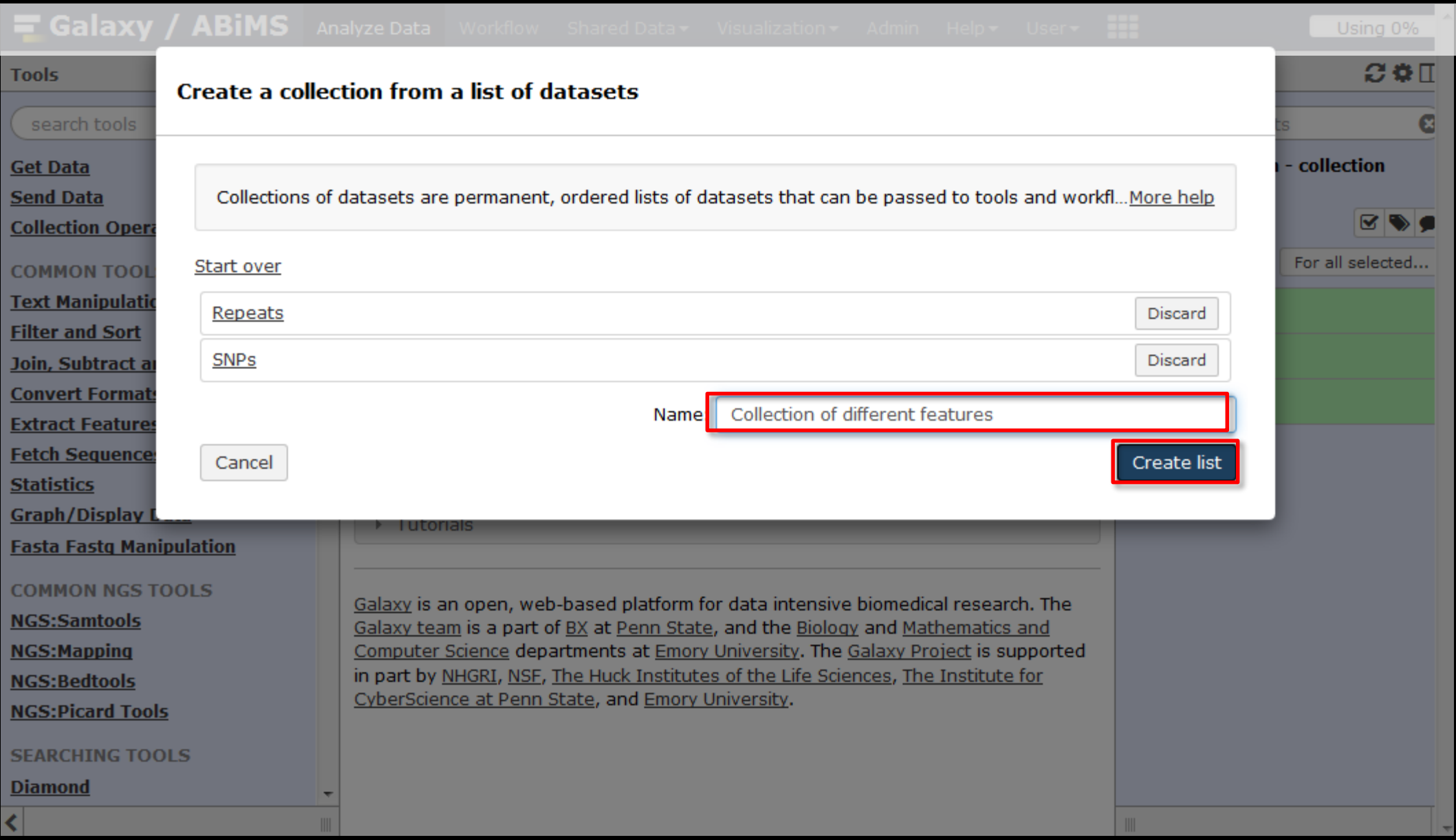
Dataset collection

Create a dataset collection



Dataset collection

Create a dataset collection



Dataset collection

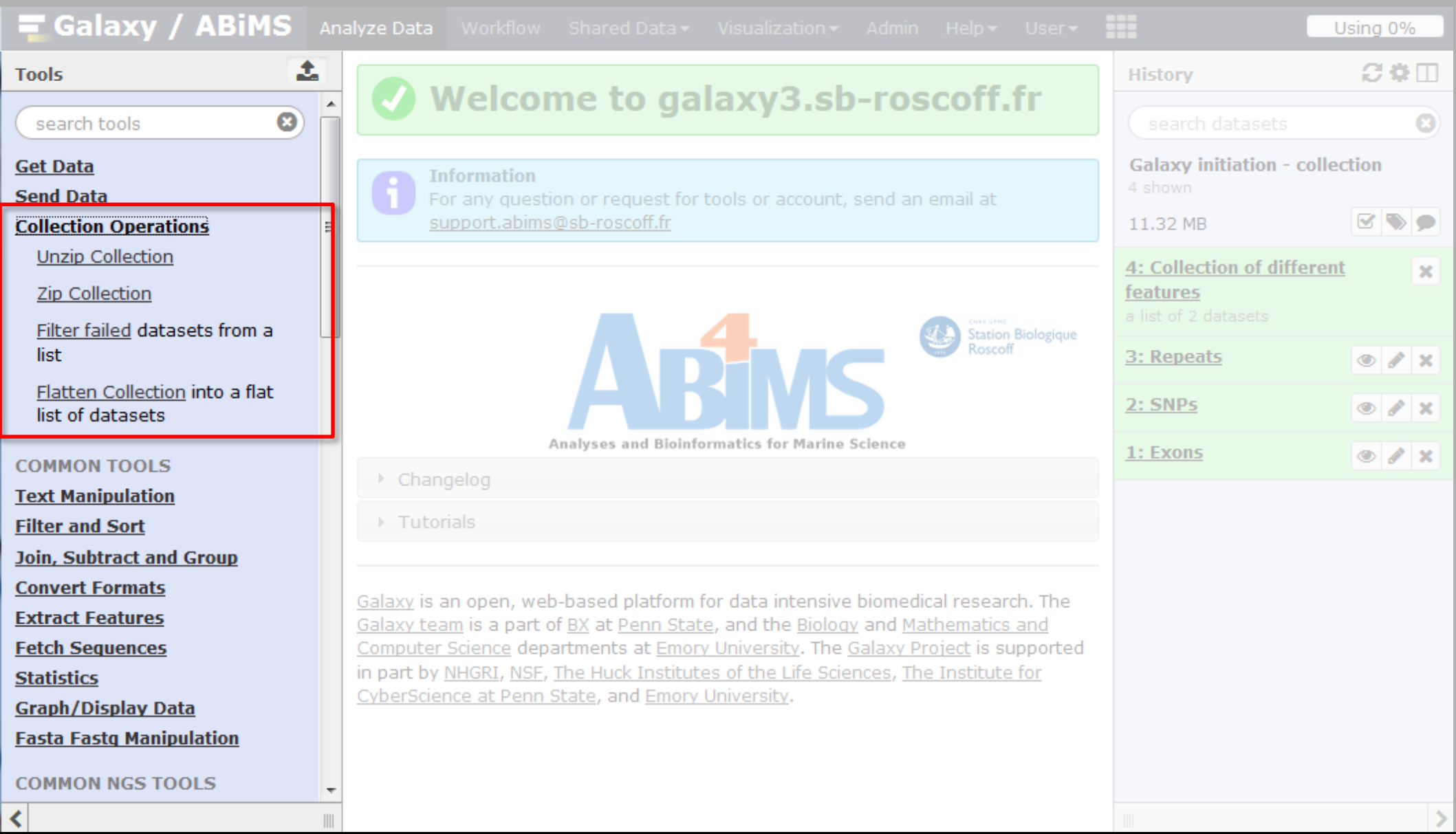
Create a dataset collection

Dataset collection

Create a dataset collection

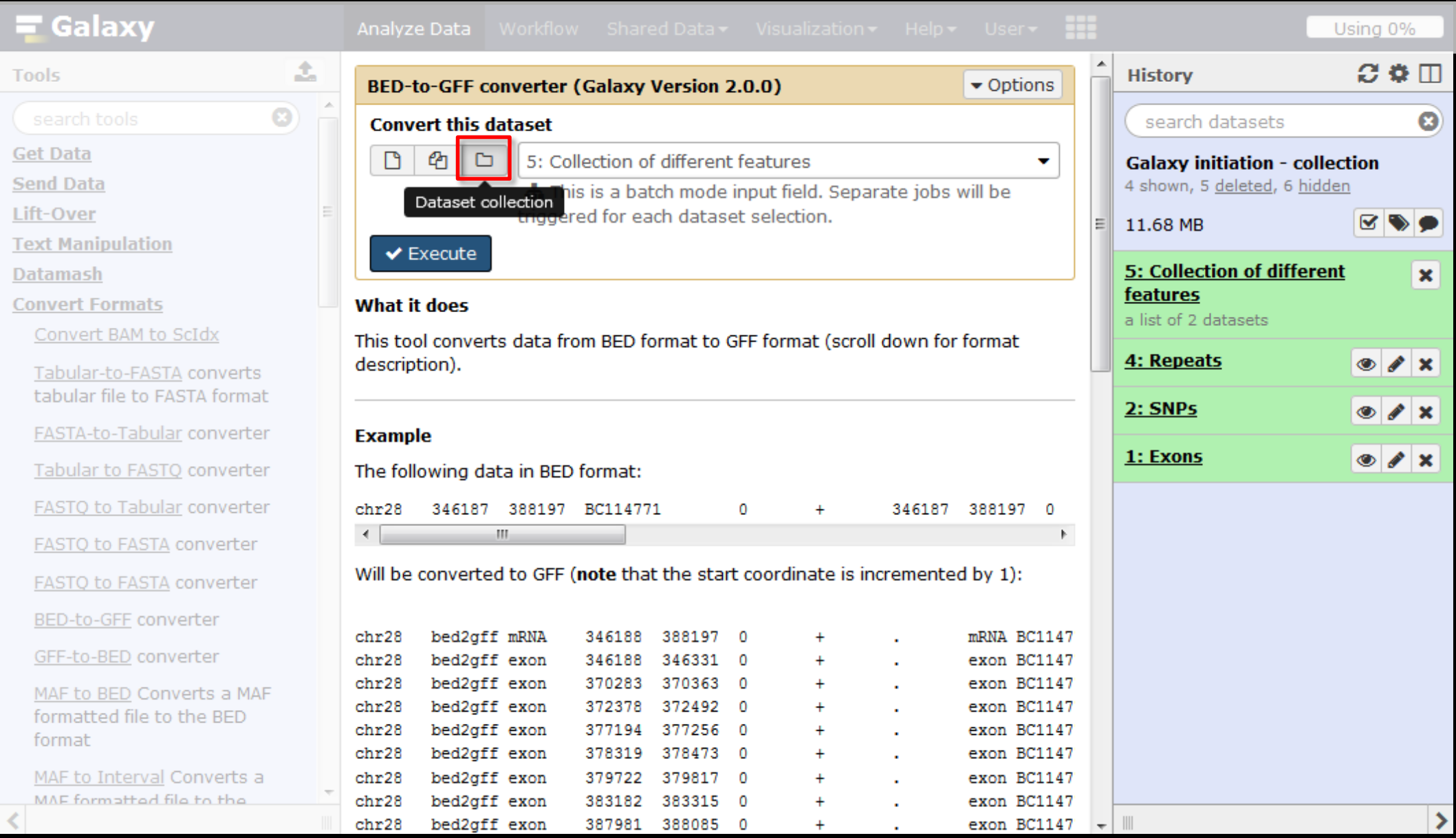
Dataset collection

Tools for collection operations



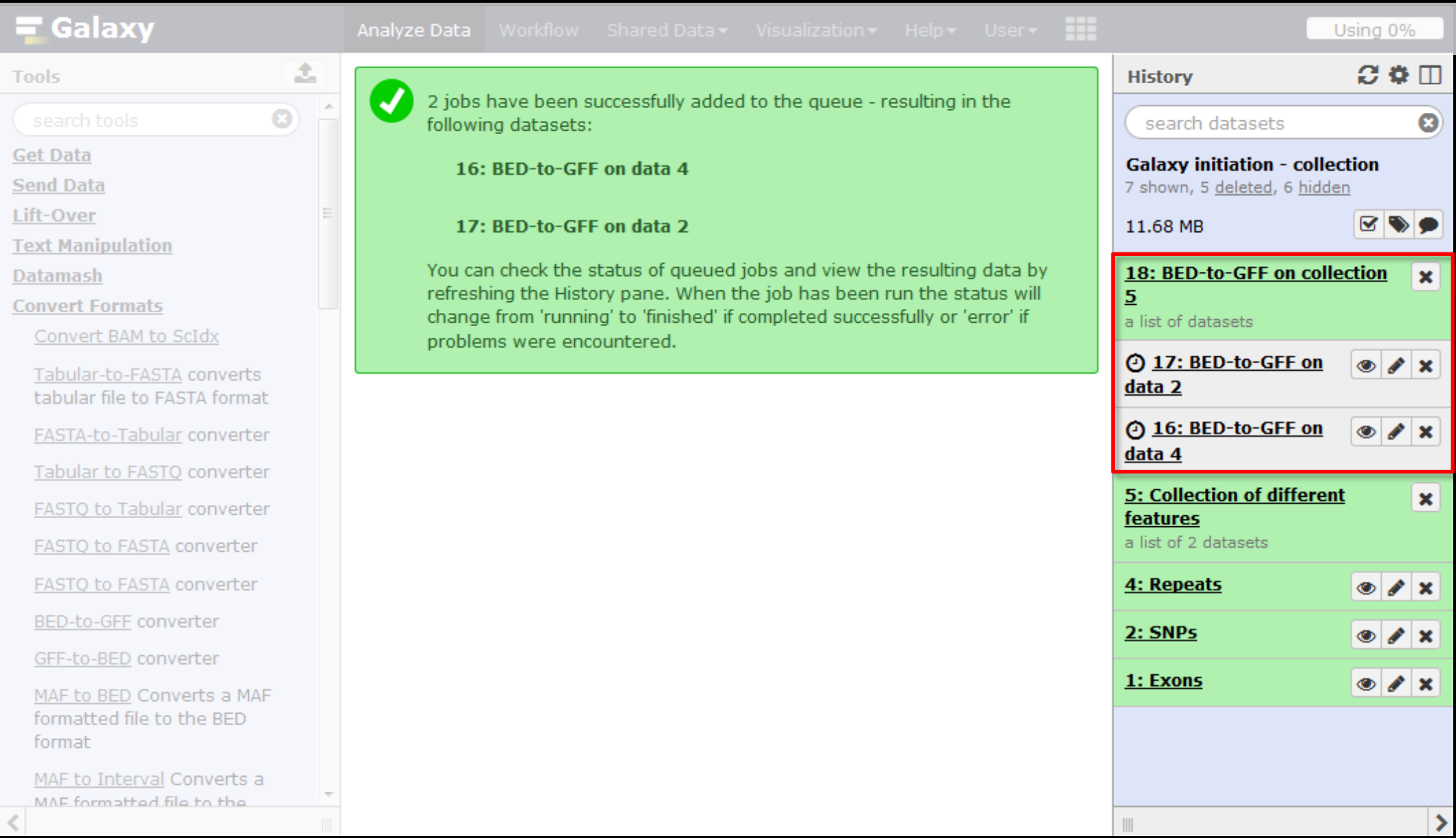
Dataset collection

Use a collection as input



Dataset collection

Use a collection as input



Hands-on

COLLECTION

Collection – Hands-on

1. Create a dataset list with SNPs, repeats and exons
2. Run tool “Convert Formats -> BED-to-GFF converter” on the dataset list

END

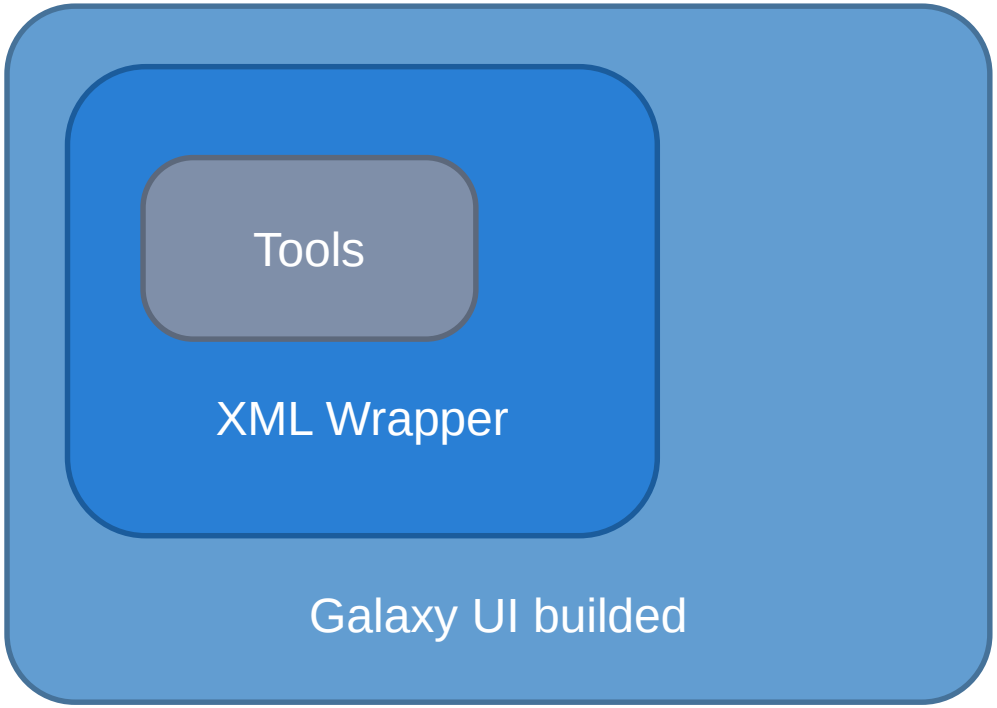
BONUS

How are tools born?

BONUS

Bonus

- How to import a tool in Galaxy?



Bonus

- How to import a tool in Galaxy?

```
[lecorguille@n0 ~]$ e-PCR --help
e-PCR: invalid option -- -
usage: [-hV] [posix-options] stsfile [fasta ...]
[compat-options]
where posix-options are:
        -m ##    Margin (default 50)
        -w ##    Wordsize  (default 7)
        -n ##    Max mismatches allowed (default 0
        -g ##    Max indels allowed (default 0)
        -f ##    Use ## discontiguos words, slow i
                        ##>1
        -o ##    Set output file
        -t ##    Set output format:
                 1 - classic, range (pos1..pos2)
                 2 - classic, midpoint
                 3 - tabular
                 4 - tabular with alignment in com
                        (slow)
        -d##-##  Set default size range
                        (default 100-350)
        -p +-    Turn hits postprocess on/off
        -v ##    Verbosity flags
        -a a|f   Use presize alignmens (only if
                        gaps>0), slow
                  a - Allways or f - as Fallback
        -x +-    Use 5'-end lowercase masking of
                        primers (default -)
        -u +-    Uppercase all primers (default -)

[...]
```

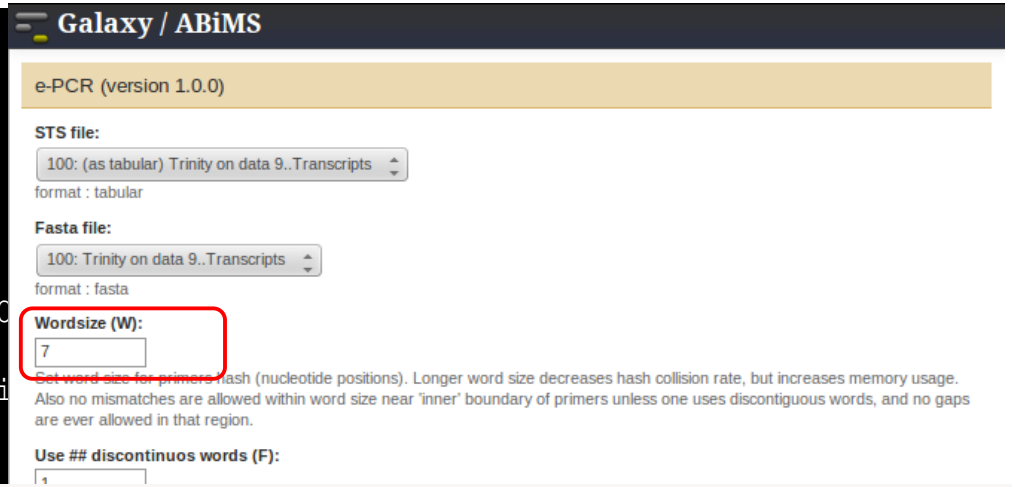
Bonus

- How to import a tool in Galaxy?

```
<tool id="abims_epcr" name="e-PCR">
    <!-- author : lecorguille@sb-roscoff.fr -->
    <!-- date : 11-05-12 -->
    <description>e-PCR parses stsfile in unists format, then reads nucleotide sequence data in FASTA format from files listed in commandline if any, or from stdin otherwise. For input sequences e-PCR finds matches and prints output in one of three formats.</description>

    <command>e-PCR -w $wordsize -f $wordcnt -m $margin -d$sts_size_lo-$sts_size_hi -n $max_mismatch -g $max_gap -t $output_format $infile_stsfile $infile_fasta > $output</command>

    <inputs>
        <param name="infile_stsfile" type="data" label="STS file" format="tabular" help="format : tabular" />
        <param name="infile_fasta" type="data" label="Fasta file" format="fasta" help="format : fasta" />
        <param name="wordsize" type="integer" label="Wordsize (W)" value="7" help="Set word size for primers hash (nucleotide positions). Longer word size decreases hash collision rate, but increases memory usage. Also no mismatches are allowed within word size near 'inner' boundary of primers unless one uses discontiguous words, and no gaps are ever allowed in that region." />
        <param name="wordcnt" type="integer" label="Use ## discontinuos words (F)" value="1" help="Set discontiguous word count for primers hash (1 means 'use contiguous words'). Discontiguous words increase number of hash tables and decrease 'effective' word size (thus increasing hash collision rate), so make search significantly slower, but increase sencitivity by allowing mismatches within word size. Reasonable values are 1 (contiguous words) and 3." />
        <param name="margin" type="integer" label="Margin (M)" value="50" help="Set maximal allowed deviation of hit product size from expected STS size." />
        <param name="sts_size_lo" type="integer" label="Set default sts lower size (D)" value="100" help="Set ddefault STS size range - values used for STSs that have no size associated in file." />
        <param name="sts_size_hi" type="integer" label="Set default sts higher size (D)" value="400" help="Set ddefault STS size range - values used for STSs that have no size associated in file." />
        <param name="max_mismatch" type="integer" label="Max mismatches allowed (N)" value="0" help="Set maximal number of mismatches allowed in primer-to-sequence alignment (per primer!)." />
        <param name="max_gap" type="integer" label="Max indels allowed (G)" value="0" help="Set maximal number of gaps allowed in primer-to-sequence alignment (per primer!)." />
        <param name="output_format" type="select" help="Output formats">
            <label>Set output format (T)</label>
            <option value="1">classic, range (pos1..pos2)</option>
            <option value="2">classic, midpoint</option>
            <option value="3" selected="true">tabular</option>
            <option value="4">tabular with alignment in comments (slow)</option>
        </param>
    </inputs>

    <outputs>
        <data name="output" format="tabular" />
    </outputs>
```

- How to import a tool in Galaxy?

```
[lecorguille@n0 ~]$ e-PCR --help
e-PCR: invalid option -- -
usage: [-hV] [posix-options] stsfile [fasta ...]
[compat-options]
where posix-options are:
        -m ##    Margin (default 50)
        -w ##    Wordsize  (default 7)
        -n ##    Max mismatches allowed (default 0
        -g ##    Max indels allowed (default 0)
        -f ##    Use ## discontiguos words, slow i
                          ##>1
        -o ##    Set output file
```




```
<tool id="abims_epcr" name="e-PCR">
    <!-- author : lecorguille@sb-roscoff.fr -->
    <!-- date : 11-05-12 -->
    <description>e-PCR parses stsfile in unists format, then reads nucleotide sequence data in FASTA format from files listed in commandline if
    any, or from stdin otherwise. For input sequences e-PCR finds matches and prints output in one of three formats.</description>
    <command>e-PCR -w $wordsize -f $wordcnt -m $margin -d$sts_size_lo-$sts_size_hi -n $max_mismatch -g $max_gap -t $output_format $infile_stsfile
    $infile_fasta > $output</command>
    <inputs>
        <param name="infile_stsfile" type="data" label="STS file" format="tabular" help="format : tabular" />
        <param name="infile_fasta" type="data" label="Fasta file" format="fasta" help="format : fasta" />
        <param name="wordsize" type="integer" label="Wordsize (W)" value="7" help="Set word size for primers hash (nucleotide positions).
        Longer word size decreases hash collision rate, but increases memory usage. Also no mismatches are allowed within word size near
        'inner' boundary of primers unless one uses discontiguous words, and no gaps are ever allowed in that region." />
        <param name="wordcnt" type="integer" label="Use ## discontinuos words (F)" value="1" help="Set discontiguous word count for primers
        hash (1 means 'use contiguous words'). Discontiguous words increase number of hash tables and decrease 'effective' word size (thus
        increasing hash collision rate), so make search significantly slower, but increase sencitivity by allowing mismatches within word
        size. Reasonable values are 1 (contiguous words) and 3." />
        <param name="margin" type="integer" label="Margin (M)" value="50" help="Set maximal allowed deviation of hit product size from
        expected STS size." />
        <param name="sts_size_lo" type="integer" label="Set default sts lower size (D)" value="100" help="Set ddefault STS size range - values
        used for STSs that have no size associated in file." />
        <param name="sts_size_hi" type="integer" label="Set default sts higher size (D)" value="400" help="Set ddefault STS size range -
        values used for STSs that have no size associated in file." />
        <param name="max_mismatch" type="integer" label="Max mismatches allowed (N)" value="0" help="Set maximal number of mismatches allowed
```