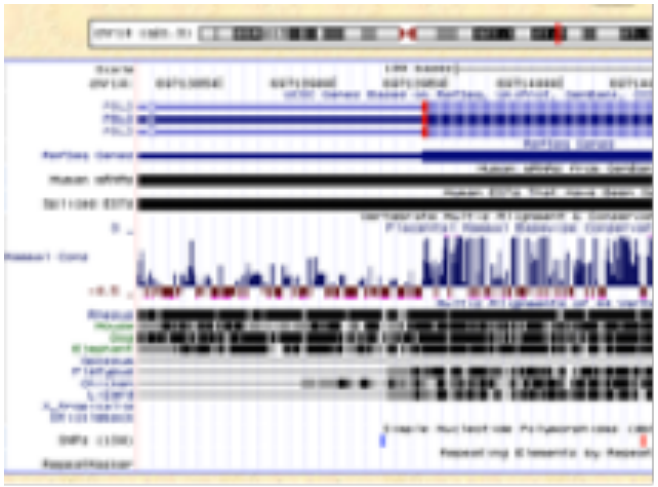
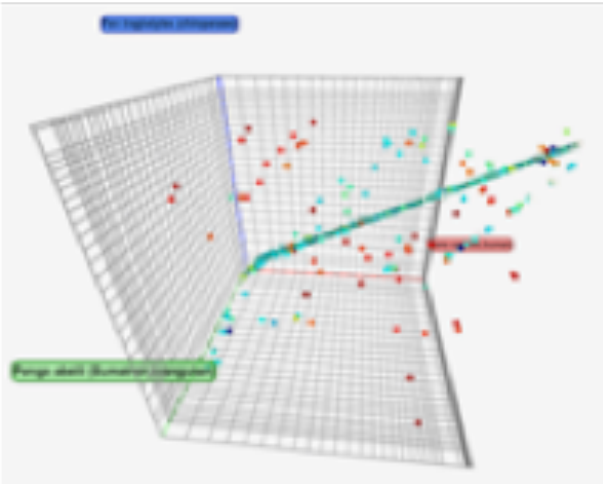
Computational Genomics

Computational Arithmetics I





Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Retrieve all Gencode Exons from Human Genome (hg38)

HPRC Kaiser Galaxy | Analyze Data | Workflow | Visualize | Shared Data | Admin | Help | User | Using 2%

Tools

search tools

- HPRC
- Get Data
- Send Data
- Collection Operations
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- FASTQ Quality Control
- Mapping
- NCBI BLAST+
- Assembly
- Annotation
- FASTA/FASTQ
- Datamash
- EMBOSS
- Multiple Alignments
- Picard
- BED
- SAM/BAM
- Proteomics

Best Practices for Kaiser Galaxy

- Kaiser Galaxy (docs, slides) is configured for teaching purposes so all users have a file quota of 1TB. How to permanently delete nonessential files.
- FTP uploads are removed from ftp directory 48 hours after uploading so import your ftp files into Galaxy the same day as you upload to ftp
- Only certain tools that support multi-core processing have the Job Resources Parameters option which allow you to select cores, memory and time.
- The default job resource parameters for all tools is 1 core with 2GB memory for 24 hours (24 SUs).
  - Configuring a job to use 28 cores for 1 hour requires 28 SUs. (28 cores for 168 hours = 4704 SUs).
  - Configuring a job to use 54GB memory for 1 hour requires 28 SUs. (54GB memory for 168 hours = 4704 SUs).
  - If a tool you used failed because it needs the Job Resource Parameters option added, contact the HPRC helpdesk.

COVID-19 related research on Galaxy: **training**, **tutorials**, **documents**

Current known issues: Some tools cannot be removed from favorites. Choose your favorites wisely.

Uploading data from your computer
from Galaxy Project

History

search datasets

Unnamed history

(empty)

This history is empty. You can **load your own data** or **get data from an external source**

Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Retrieve all Gencode Exons from Human Genome (hg38)

HPRC Kaiser Galaxy | Analyze Data | Workflow | Visualize | Shared Data | Admin | Help | User | Using 2%

Tools

search tools

- HPRC
- Get Data
- Send Data
- Collection Operations
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- FASTQ Quality Control
- Mapping
- NCBI BLAST+
- Assembly
- Annotation
- FASTA/FASTQ
- Datamash
- EMBOSS
- Multiple Alignments
- Picard
- BED
- SAM/BAM
- Proteomics

Best Practices for Kaiser Galaxy

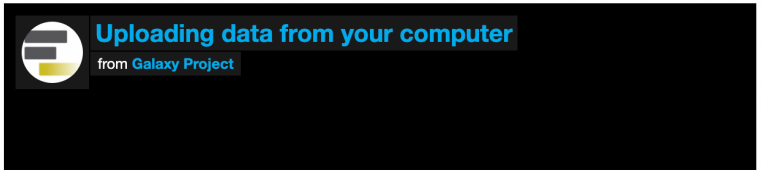
Name Your History!

time.

- The default job resource parameters for all tools is 1 core with 2GB memory for 24 hours (24 SUs).
  - Configuring a job to use 28 cores for 1 hour requires 28 SUs. (28 cores for 168 hours = 4704 SUs).
  - Configuring a job to use 54GB memory for 1 hour requires 28 SUs. (54GB memory for 168 hours = 4704 SUs).
  - If a tool you used failed because it needs the Job Resource Parameters option added, contact the HPRC helpdesk.

COVID-19 related research on Galaxy: **training**, **tutorials**, **documents**

Current known issues: Some tools cannot be removed from favorites. Choose your favorites wisely.



History

search datasets

JSG_History

(empty)

This history is empty. You can **load your own data** or **get data from an external source**

Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Retrieve all Gencode Exons from Human Genome (hg38)

HPRC Kaiser Galaxy | Analyze Data | Workflow | Visualize | Shared Data | Admin | Help | User | Using 2%

Tools

search tools

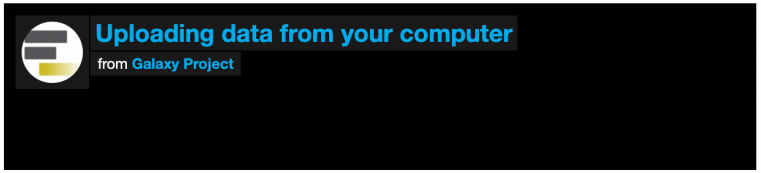
- HPRC
- Get Data
  - **Faster Download and Extract Reads in FASTQ** format from NCBI SRA
  - **Download and Extract Reads in FASTA/Q** format from NCBI SRA
  - **Download and Extract Reads in BAM** format from NCBI SRA
  - **Upload File** from your computer
  - **UCSC Main** table browser
  - **UCSC Archaea** table browser
  - **SRA** server
  - **EBI SRA** ENA SRA
  - **modENCODE fly** server
  - **InterMine** server
  - **Flymine** server
  - **modENCODE modMine** server
  - **MouseMine** server
  - **Ratmine** server
  - **YeastMine** server
  - **modENCODE worm** server
  - **WormBase** server
  - **ZebrafishMine** server
  - **EuPathDB** server
  - **HbVar** Human Hemoglobin Variants and Thalassemias
- Send Data
- Collection Operations
- Lift-Over
- Text Manipulation
- Convert Formats

Best Practices for Kaiser Galaxy

- Kaiser Galaxy (docs, slides) is configured for teaching purposes so all users have a file quota of 1TB. How to permanently delete nonessential files.
- FTP uploads are removed from ftp directory 48 hours after uploading so import your ftp files into Galaxy the same day as you upload to ftp
- Only certain tools that support multi-core processing have the Job Resources Parameters option which allow you to select cores, memory and time.
- The default job resource parameters for all tools is 1 core with 2GB memory for 24 hours (24 SUs).
  - Configuring a job to use 28 cores for 1 hour requires 28 SUs. (28 cores for 168 hours = 4704 SUs).
  - Configuring a job to use 54GB memory for 1 hour requires 28 SUs. (54GB memory for 168 hours = 4704 SUs).
  - If a tool you used failed because it needs the Job Resource Parameters option added, contact the HPRC helpdesk.

COVID-19 related research on Galaxy: **training**, **tutorials**, **documents**

Current known issues: Some tools cannot be removed from favorites. Choose your favorites wisely.



History

search datasets

JSG_History

(empty)

This history is empty. You can **load your own data** or **get data from an external source**

Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Retrieve all Gencode Exons from Human Genome (hg38)

Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | Projects | Help | About Us

Table Browser

Use this tool to retrieve and export data from the Genome Browser annotation track database. You can limit retrieval based on data attributes and intersect or merge with data from another track, or retrieve DNA sequence covered by a track. More...

Select dataset

clade: Mammal **genome:** Human **assembly:** Dec. 2013 (GRCh38/hg38)
group: Genes and Gene Predictions **track:** GENCODE V39
table: knownGene describe table schema

Define region of interest

region: genome position chrX:15,560,138-15,602,945 lookup define regions
identifiers (names/accessions): paste list upload list

Optional: Subset, combine, compare with another track

filter: create
intersection: create

Retrieve and display data

output format: BED - browser extensible data Send output to Galaxy GREAT
output filename: (leave blank to keep output in browser)
file type returned: plain text gzip compressed

get output summary/statistics

Using the Table Browser

Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Retrieve all Gencode Exons from Human Genome (hg38)

Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | Projects | Help | About Us

Output knownGene as BED

☐ **Include custom track header:**

name= tb_knownGene

description= table browser query on knownGene

visibility= pack

url=

Create one BED record per:

- ○ Whole Gene
- ○ Upstream by 200 bases
- ◉ Exons plus 0 bases at each end
- ○ Introns plus 0 bases at each end
- ○ 5' UTR Exons
- ○ Coding Exons
- ○ 3' UTR Exons
- ○ Downstream by 200 bases

Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

Send query to Galaxy

Cancel

Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Retrieve all Gencode Exons from Human Genome (hg38)

HPRC Kaiser Galaxy | Analyze Data | Workflow | Visualize | Shared Data | Admin | Help | User | Using 2%

Tools

search tools

- HPRC
- Get Data
- Send Data
- Collection Operations
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- FASTQ Quality Control
- Mapping
- NCBI BLAST+
- Assembly
- Annotation
- FASTA/FASTQ
- Datamash
- EMBOSS
- Multiple Alignments
- Picard
- BED
- SAM/BAM
- Proteomics

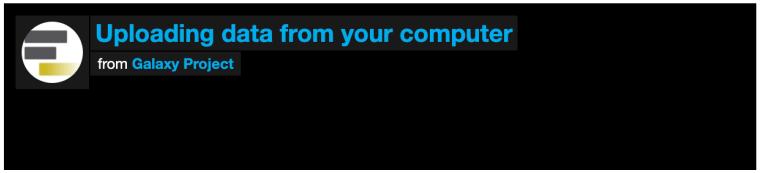
Best Practices for Kaiser Galaxy

- Kaiser Galaxy (docs, slides) is configured for teaching purposes so all users have a file quota of 1TB. How to permanently delete nonessential files.
- FTP uploads are removed from ftp directory 48 hours after uploading so import your ftp files into Galaxy the same day as you upload to ftp
- Only certain tools that support multi-core processing have the Job Resources Parameters option which allow you to select cores, memory and time.
- The default job resource parameters for all tools is 1 core with 2GB memory for 24 hours (24 SUs).
  - Configuring a job to use 28 cores for 1 hour requires 28 SUs. (28 cores for 168 hours = 4704 SUs).
  - Configuring a job to use 54GB memory for 1 hour requires 28 SUs. (54GB memory for 168 hours = 4704 SUs).
  - If a tool you used failed because it needs the Job Resource Parameters option added, contact the HPRC helpdesk.

Running

COVID-19 related research on Galaxy: **training**, **tutorials**, **documents**

Current known issues: Some tools cannot be removed from favorites. Choose your favorites wisely.



History

search datasets

JSG_History

1 shown

(empty)

1: UCSC Main on Human: knownGene (genome)

Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Retrieve all Gencode Exons from Human Genome (hg38)

HPRC Kaiser Galaxy | Analyze Data | Workflow | Visualize | Shared Data | Admin | Help | User | Using 2%

Tools

search tools

- HPRC
- Get Data
- Send Data
- Collection Operations
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- FASTQ Quality Control
- Mapping
- NCBI BLAST+
- Assembly
- Annotation
- FASTA/FASTQ
- Datamash
- EMBOSS
- Multiple Alignments
- Picard
- BED
- SAM/BAM
- Proteomics

Best Practices for Kaiser Galaxy

- Kaiser Galaxy (docs, slides) is configured for teaching purposes so all users have a file quota of 1TB. How to permanently delete nonessential files.
- FTP uploads are removed from ftp directory 48 hours after uploading so import your ftp files into Galaxy the same day as you upload to ftp
- Only certain tools that support multi-core processing have the Job Resources Parameters option which allow you to select cores, memory and time.
- The default job resource parameters for all tools is 1 core with 2GB memory for 24 hours (24 SUs).
  - Configuring a job to use 28 cores for 1 hour requires 28 SUs. (28 cores for 168 hours = 4704 SUs).
  - Configuring a job to use 54GB memory for 1 hour requires 28 SUs. (54GB memory for 168 hours = 4704 SUs).
  - If a tool you used failed because it needs the Job Resource Parameters option added, contact the HPRC helpdesk.

Done!

History

search datasets

JSG_History

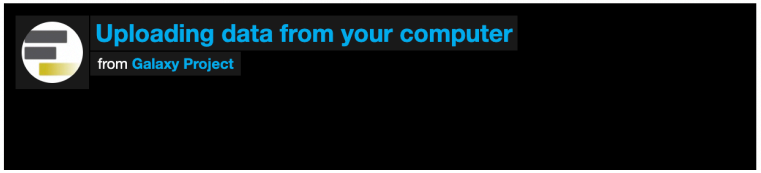
1 shown

113.62 MB

1: UCSC Main on Human: knownGene (genome)

COVID-19 related research on Galaxy: **training**, **tutorials**, **documents**

Current known issues: Some tools cannot be removed from favorites. Choose your favorites wisely.



Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Determine their lengths (using Compute an expression)

HPRC Kaiser Galaxy | Analyze Data | Workflow | Visualize | Shared Data | Admin | Help | User | Using 2%

Tools

compute an expression

Show Sections

Replace parts of text

Mummer Align two or more sequences

Search in textfiles (grep)

Regex Find And Replace

Text transformation with sed

Column Regex Find And Replace

Compute an expression on every row

Text reformatting with awk

tac reverse a file (reverse cat)

Select lines that match an expression

Summary Statistics for any numerical column

Replace Text in a specific column

Replace Text in entire line

bedtools Compute both the depth and breadth of coverage of features in file B on the features in file A (bedtools coverage)

MarkDuplicates examine aligned records in BAM datasets to locate duplicate molecules

Convert SAM to interval

MarkDuplicatesWithMateCigar examine aligned records in BAM datasets to locate duplicate molecules

EstimateLibraryComplexity assess sequence library complexity from read sequences

Filter GFF data by feature count using simple expressions

Filter GFF data by attribute using simple expressions

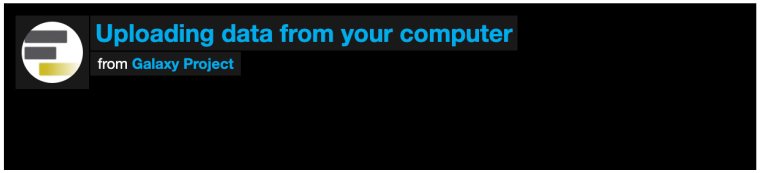
WORKFLOWS

Best Practices for Kaiser Galaxy

- Kaiser Galaxy (docs, slides) is configured for teaching purposes so all users have a file quota of 1TB. How to permanently delete nonessential files.
- FTP uploads are removed from ftp directory 48 hours after uploading so import your ftp files into Galaxy the same day as you upload to ftp
- Only certain tools that support multi-core processing have the Job Resources Parameters option which allow you to select cores, memory and time.
- The default job resource parameters for all tools is 1 core with 2GB memory for 24 hours (24 SUs).
  - Configuring a job to use 28 cores for 1 hour requires 28 SUs. (28 cores for 168 hours = 4704 SUs).
  - Configuring a job to use 54GB memory for 1 hour requires 28 SUs. (54GB memory for 168 hours = 4704 SUs).
  - If a tool you used failed because it needs the Job Resource Parameters option added, contact the HPRC helpdesk.

COVID-19 related research on Galaxy: **training**, **tutorials**, **documents**

Current known issues: Some tools cannot be removed from favorites. Choose your favorites wisely.



History

search datasets

JSG_History

1 shown

113.62 MB

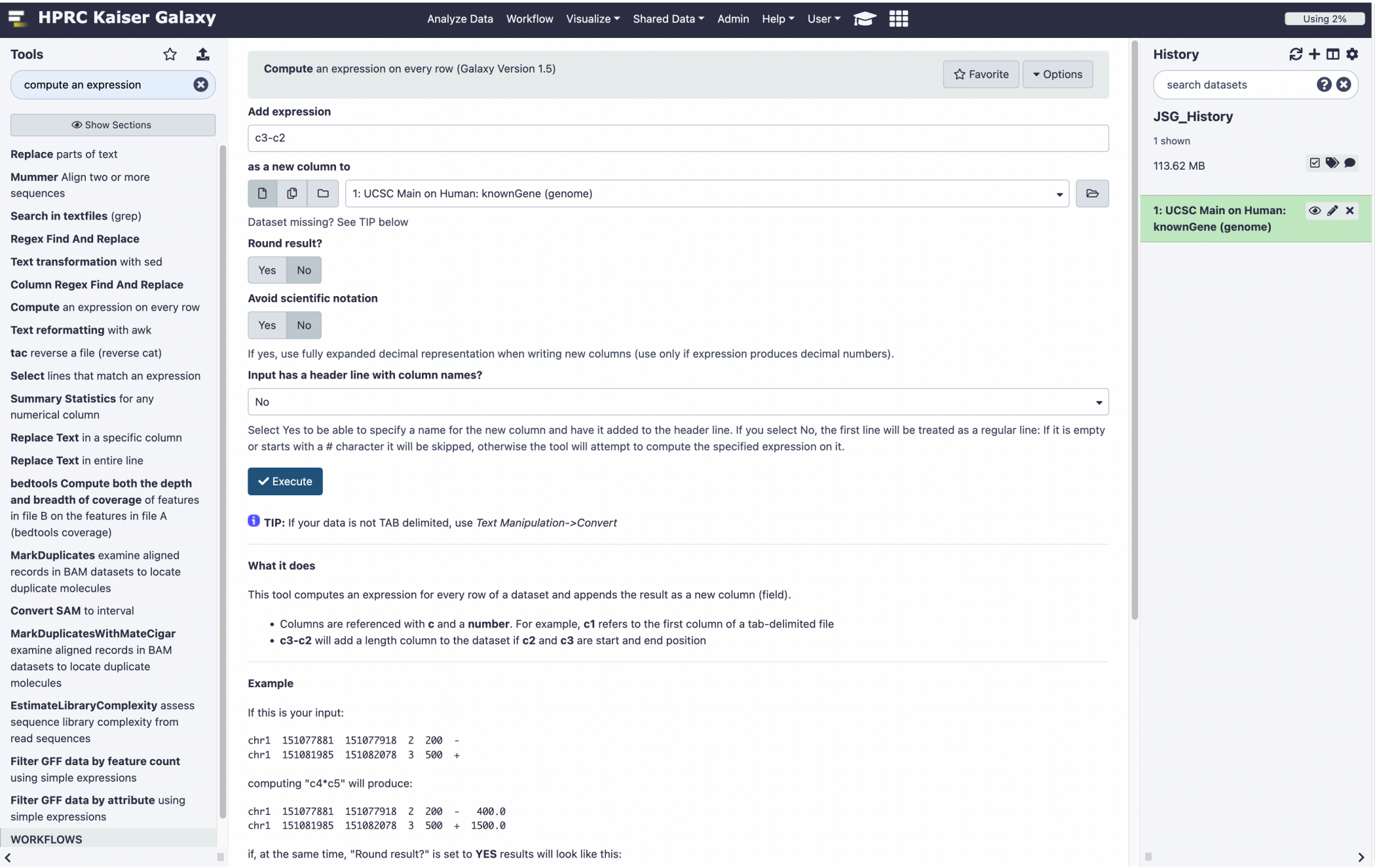
1: UCSC Main on Human: knownGene (genome)

Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Determine their lengths (using Compute an expression)



Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Determine their lengths (using Compute an expression)

Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Determine their lengths (using Compute an expression)

Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

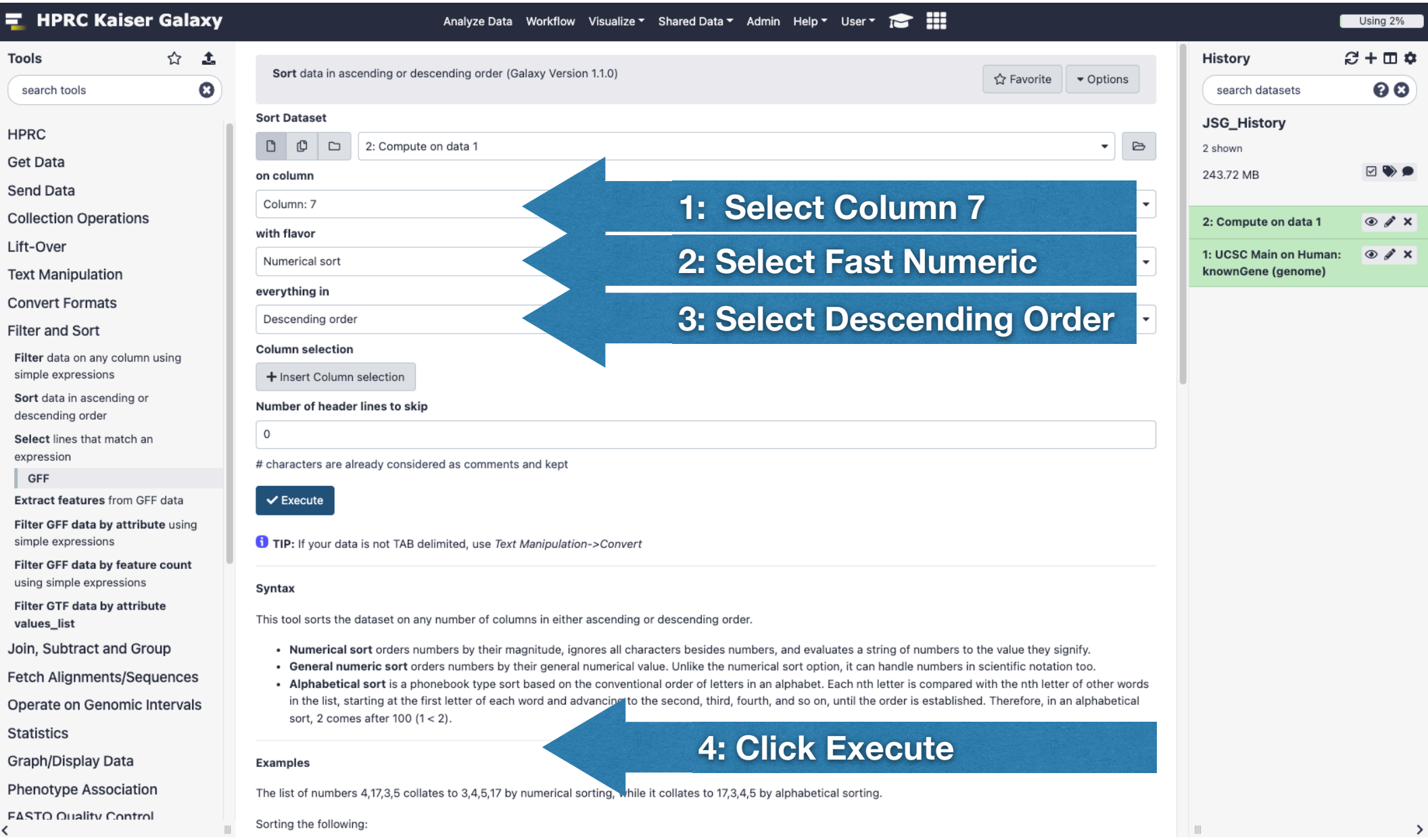
Sort numeric descending on c7 (using Sort data in ascending or descending order)

Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Sort numeric descending on c7 (using Sort data in ascending or descending order)



Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Sort numeric descending on c7 (using Sort data in ascending or descending order)

Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

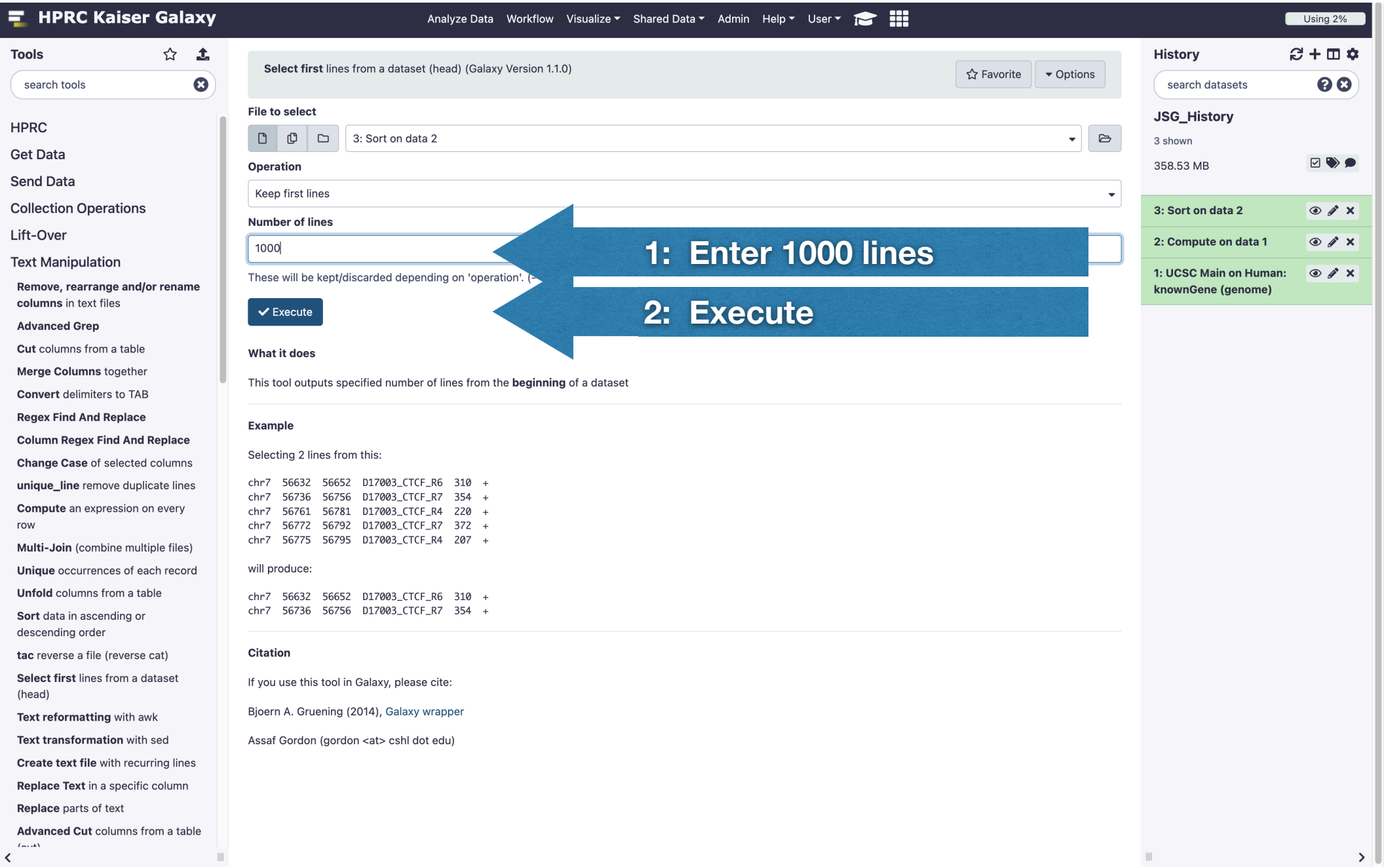
Select first 1000 lines (using Select first lines from a dataset)

Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Select first 1000 lines (using Select first lines from a dataset)

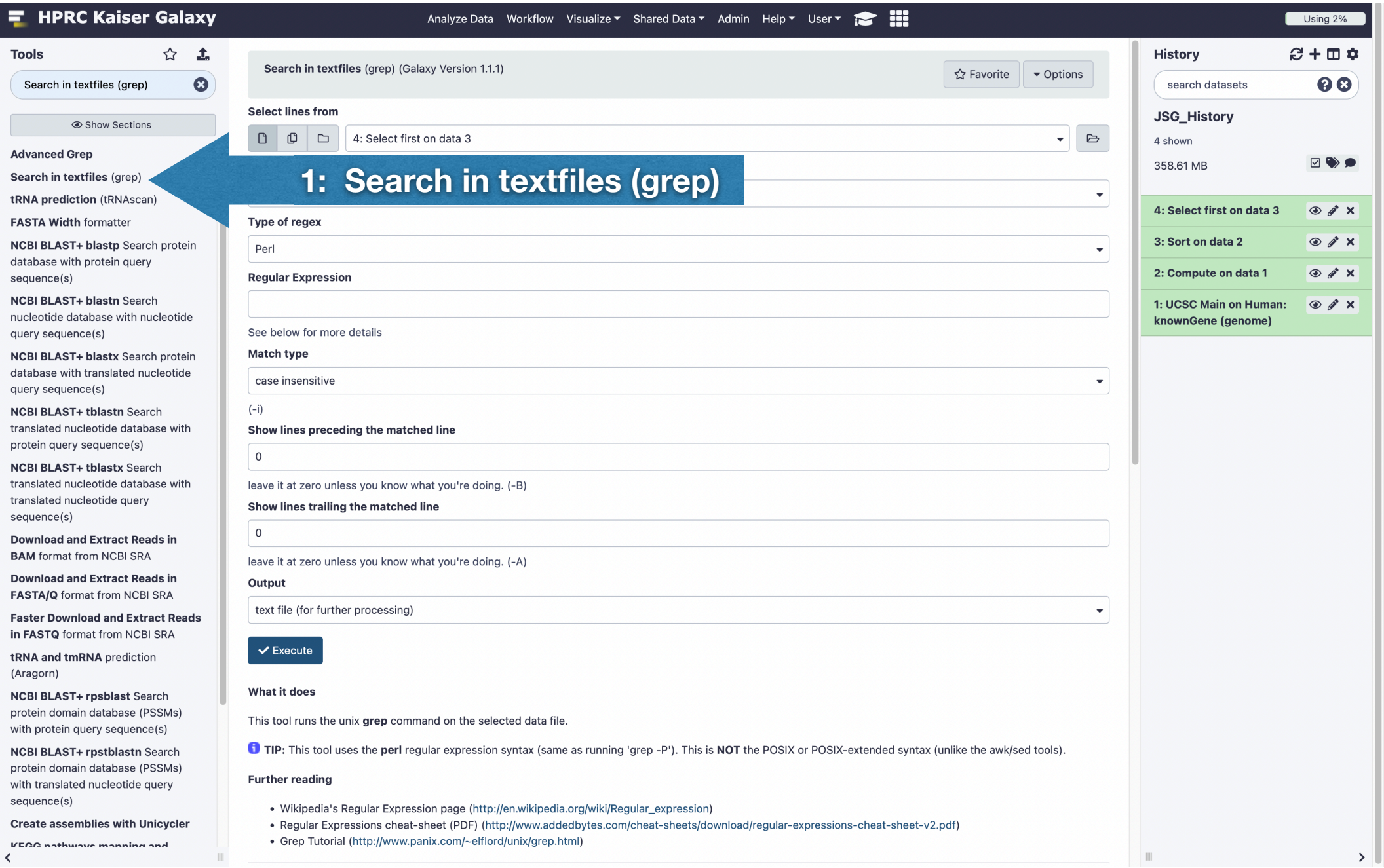


Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Search in textfiles (grep)

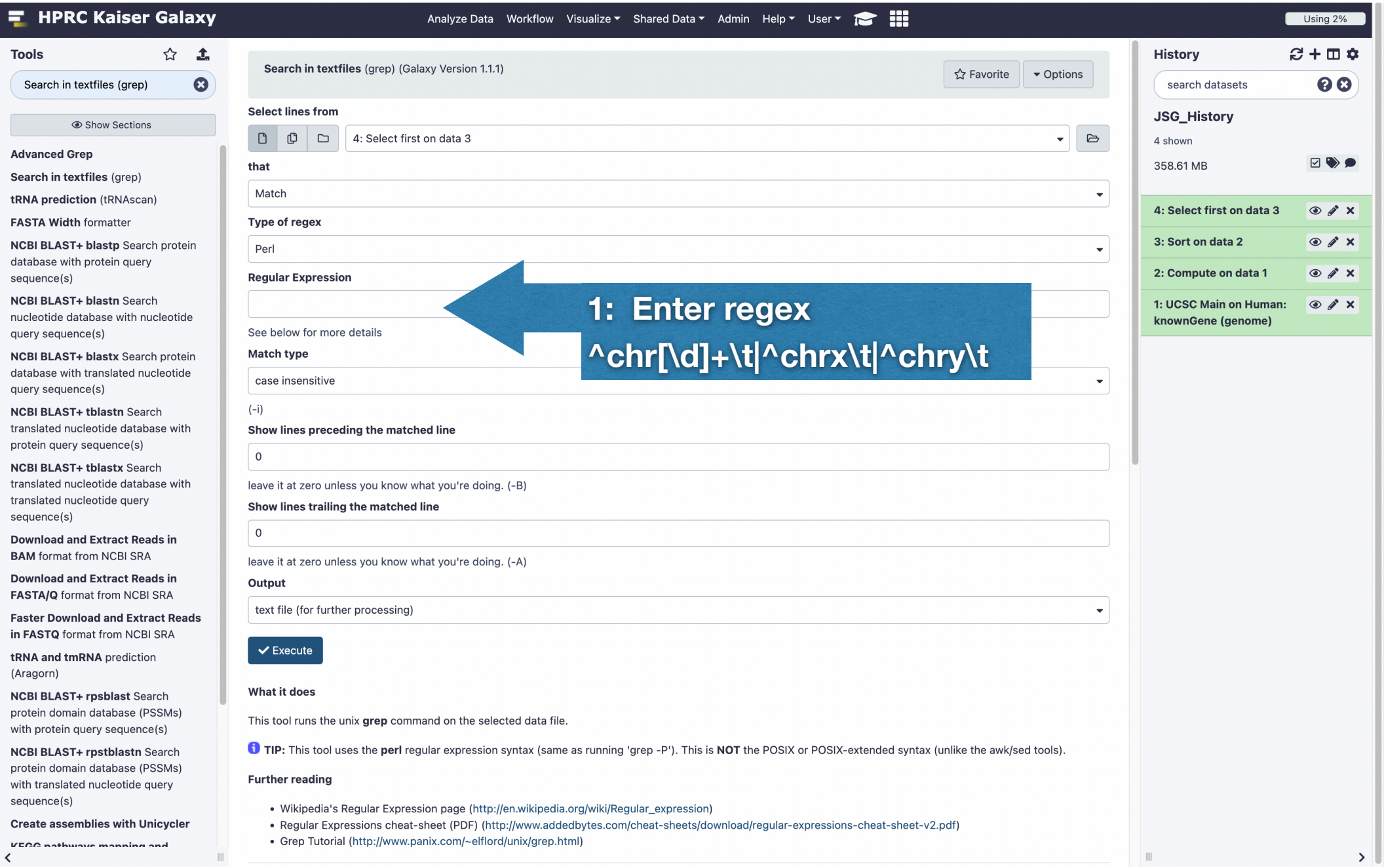


Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Search in textfiles (grep)

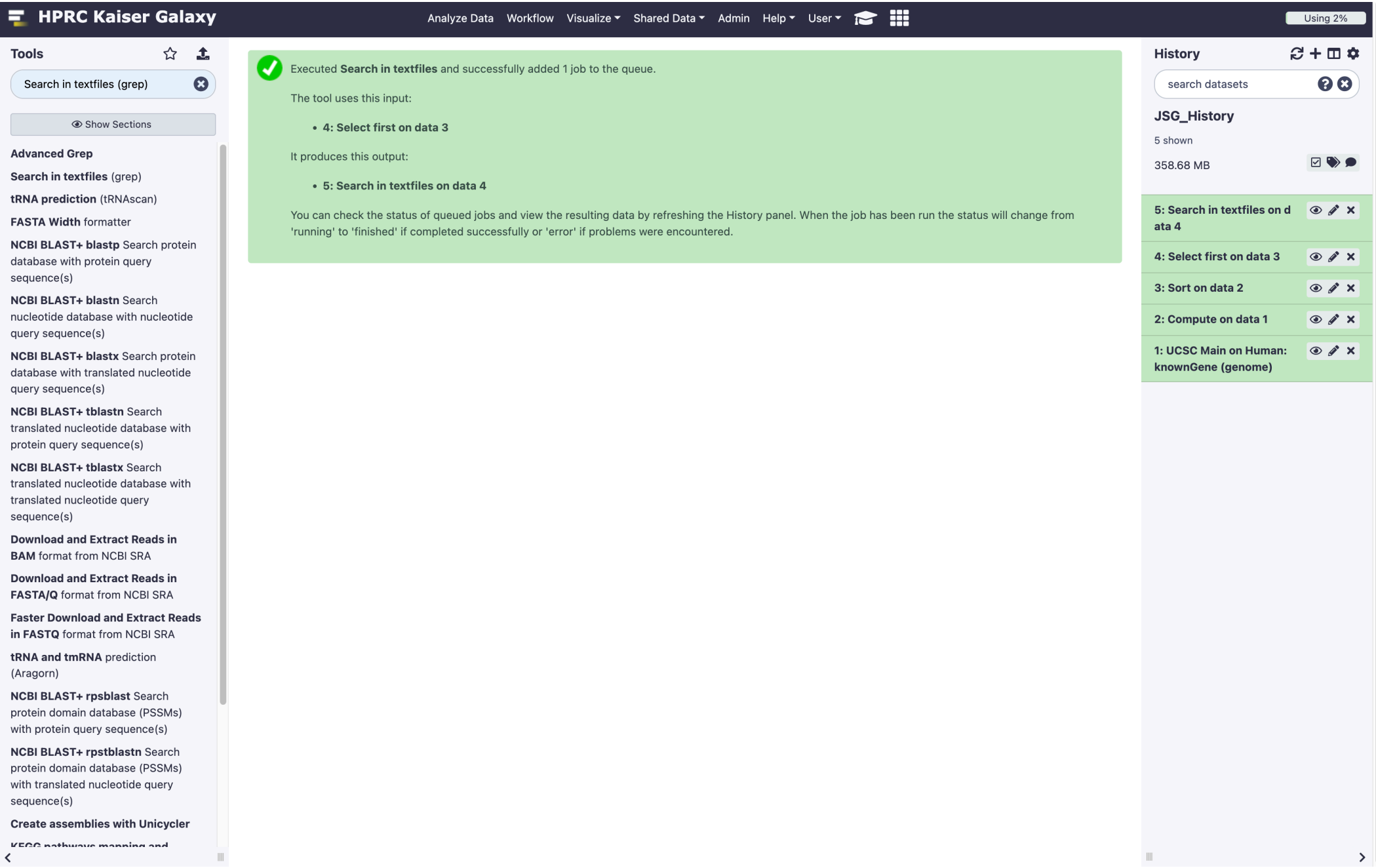


Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Search in textfiles (grep)

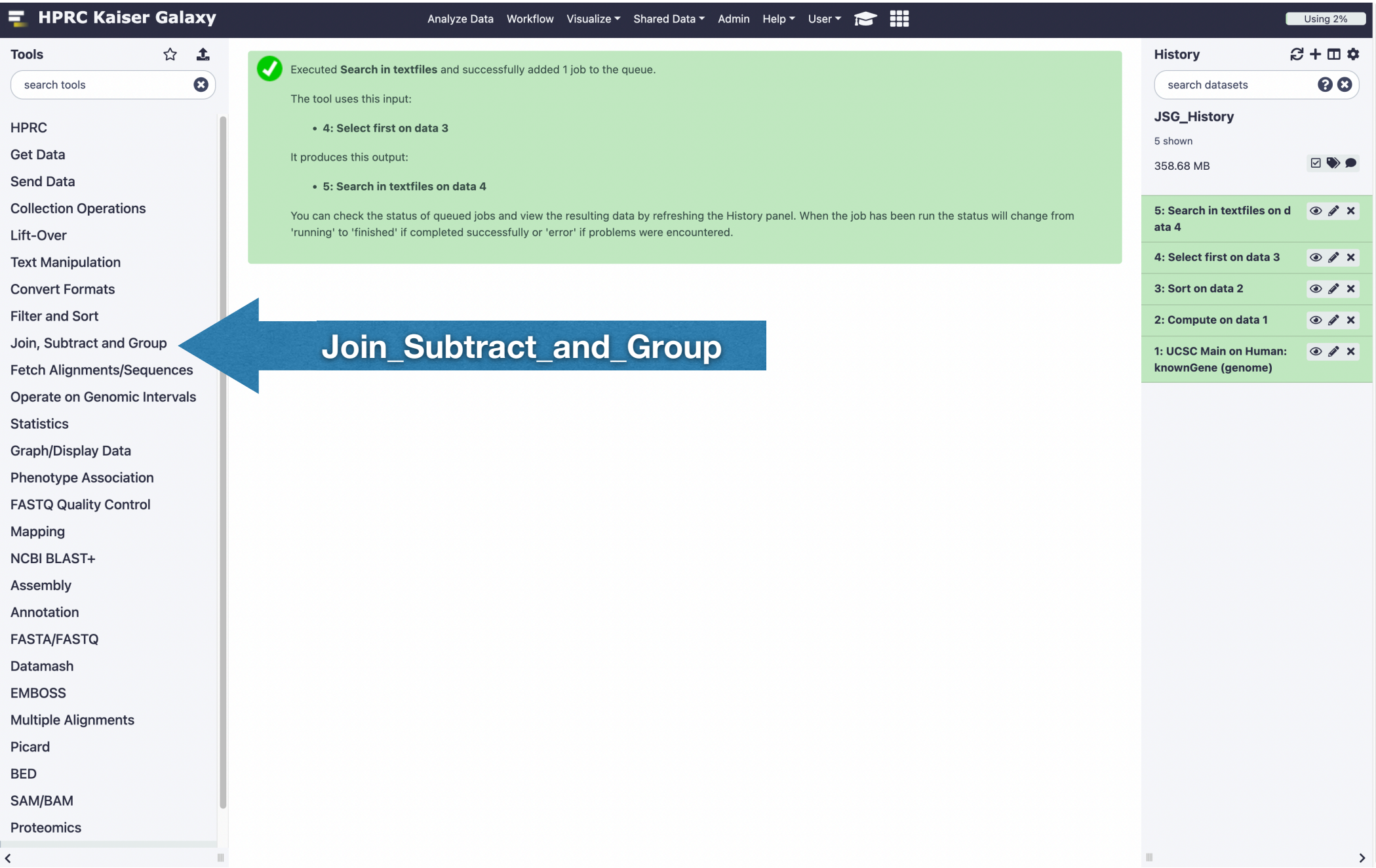


Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Go to: 'Join_Subtract_and_Group' => Group



Computational Arithmetics
Join, Subtract and Group

About GROUPING

For the following input:

```
chr22  1000  1003  TTT
chr22  2000  2003  aaa
chr10  2200  2203  TTT
chr10  1200  1203  ttt
chr22  1600  1603  AAA
```

- **Grouping on column 4 while ignoring case, and performing operation Count on column 1 will return:**

```
AAA    2
TTT    3
```

- **Grouping on column 4 while not ignoring case, and performing operation Count on column 1 will return:**

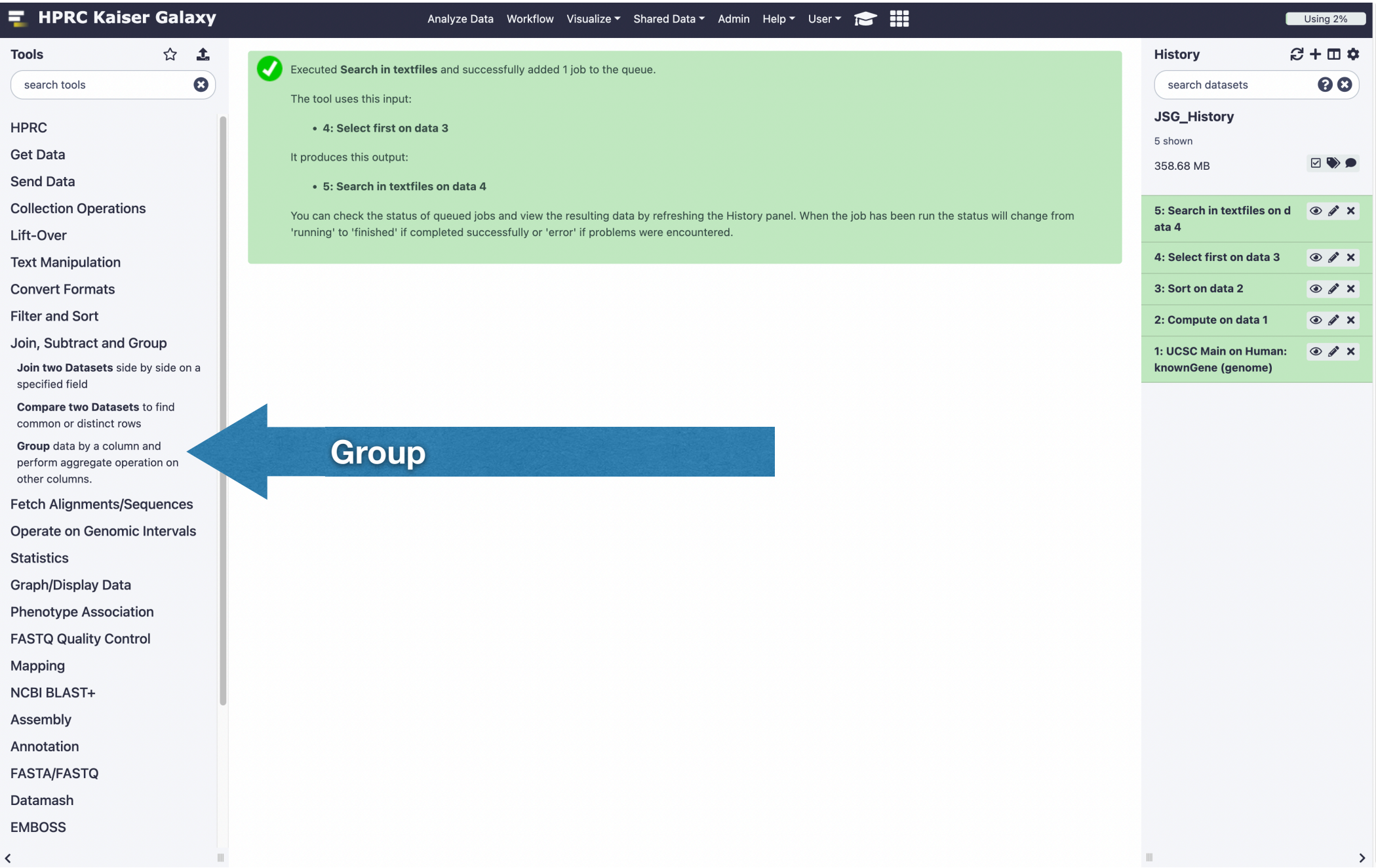
```
aaa    1
AAA    1
ttt    1
TTT    2
```

Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Go to: 'Join_Subtract_and_Group' => Group

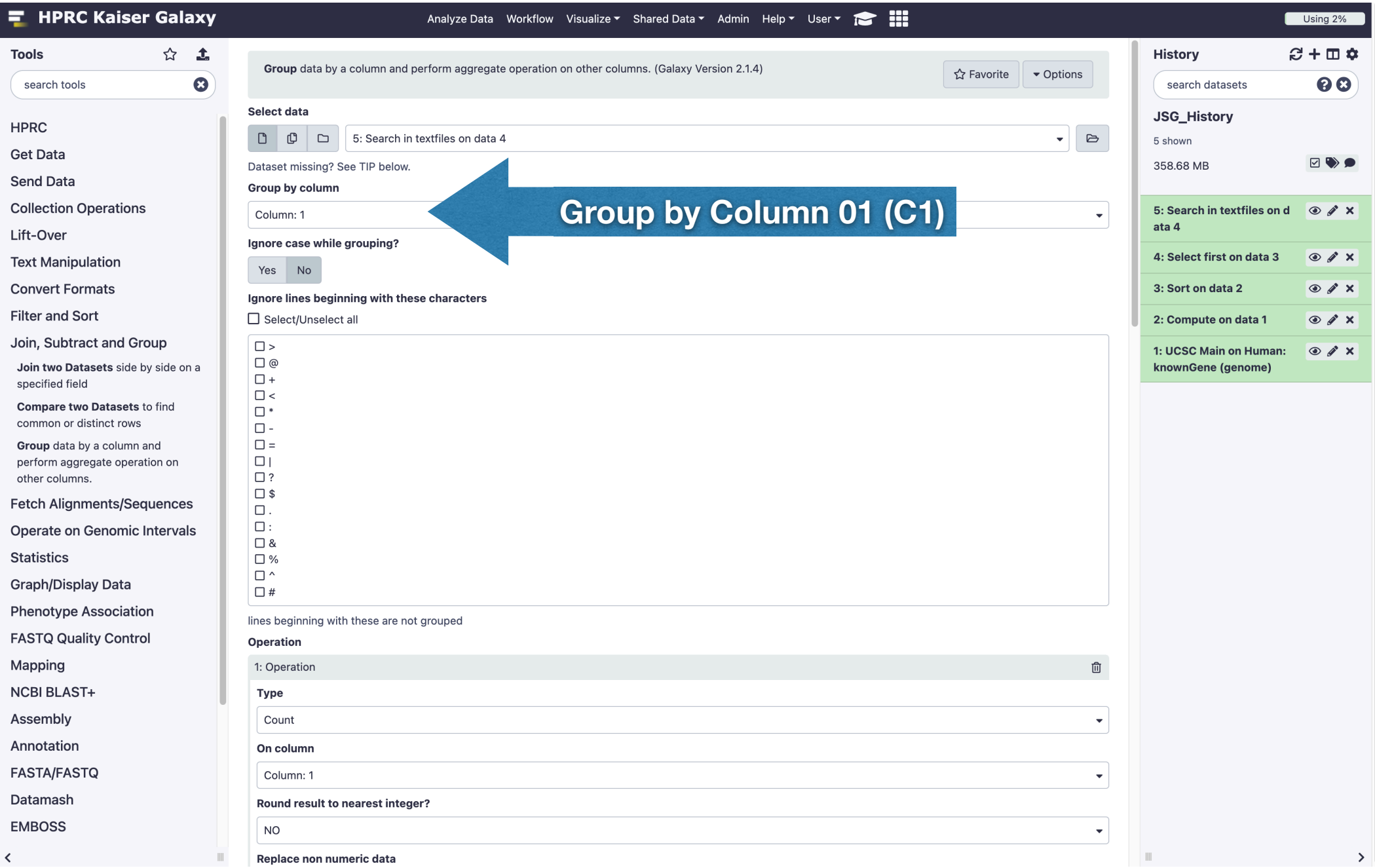


Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Go to: 'Join_Subtract_and_Group' => Group

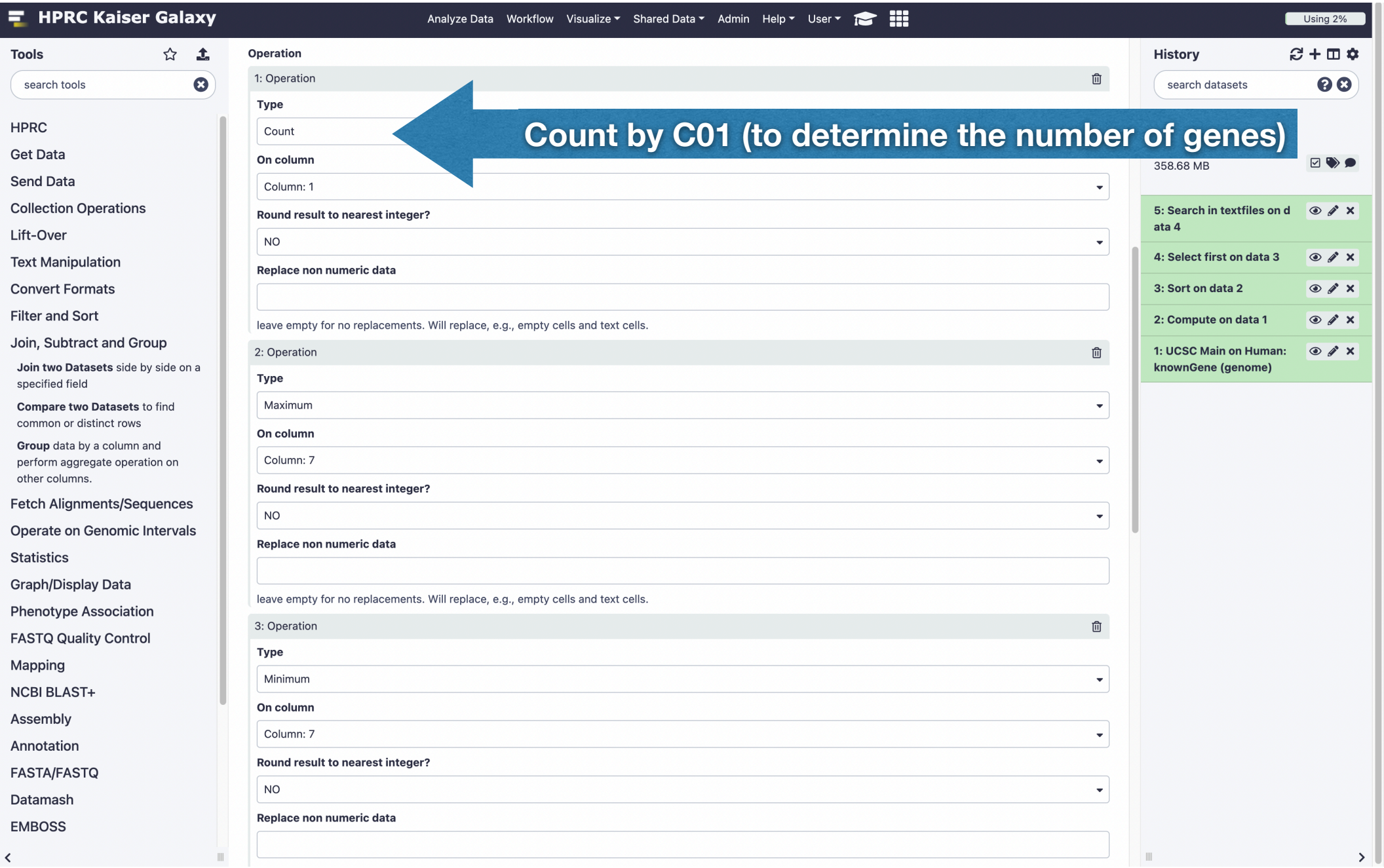


Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Go to: 'Join_Subtract_and_Group' => Group

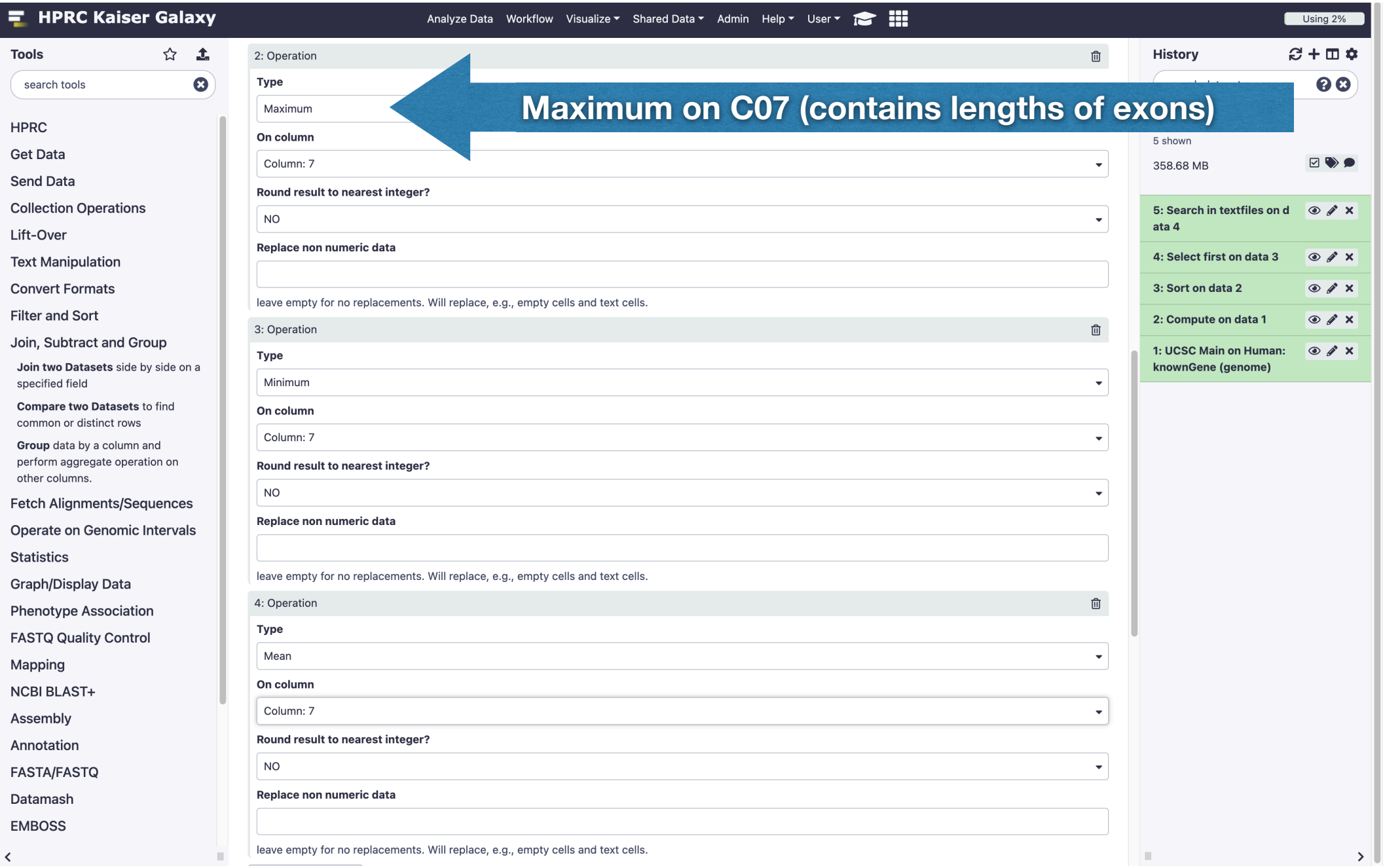


Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Go to: 'Join_Subtract_and_Group' => Group

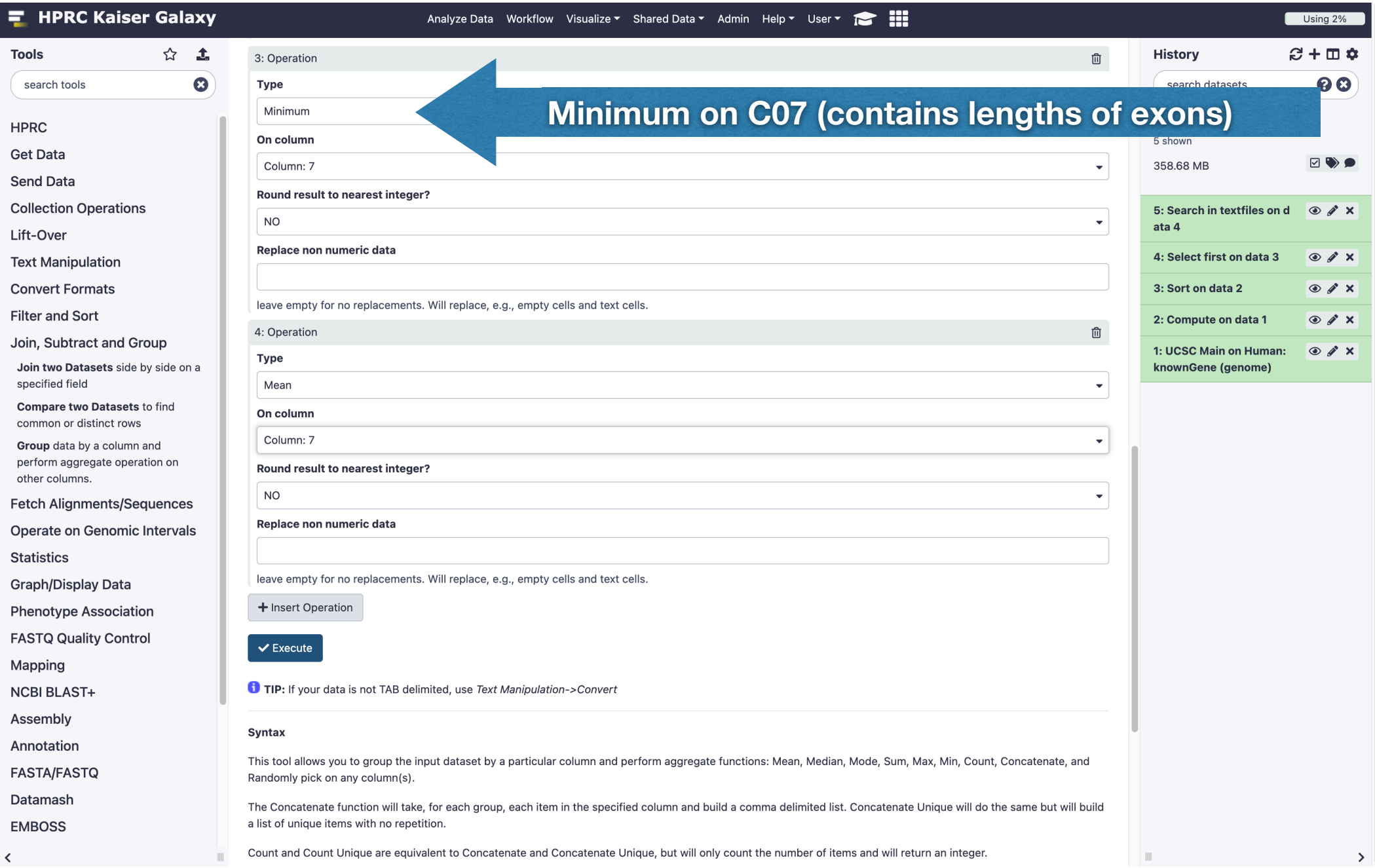


Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Go to: 'Join_Subtract_and_Group' => Group

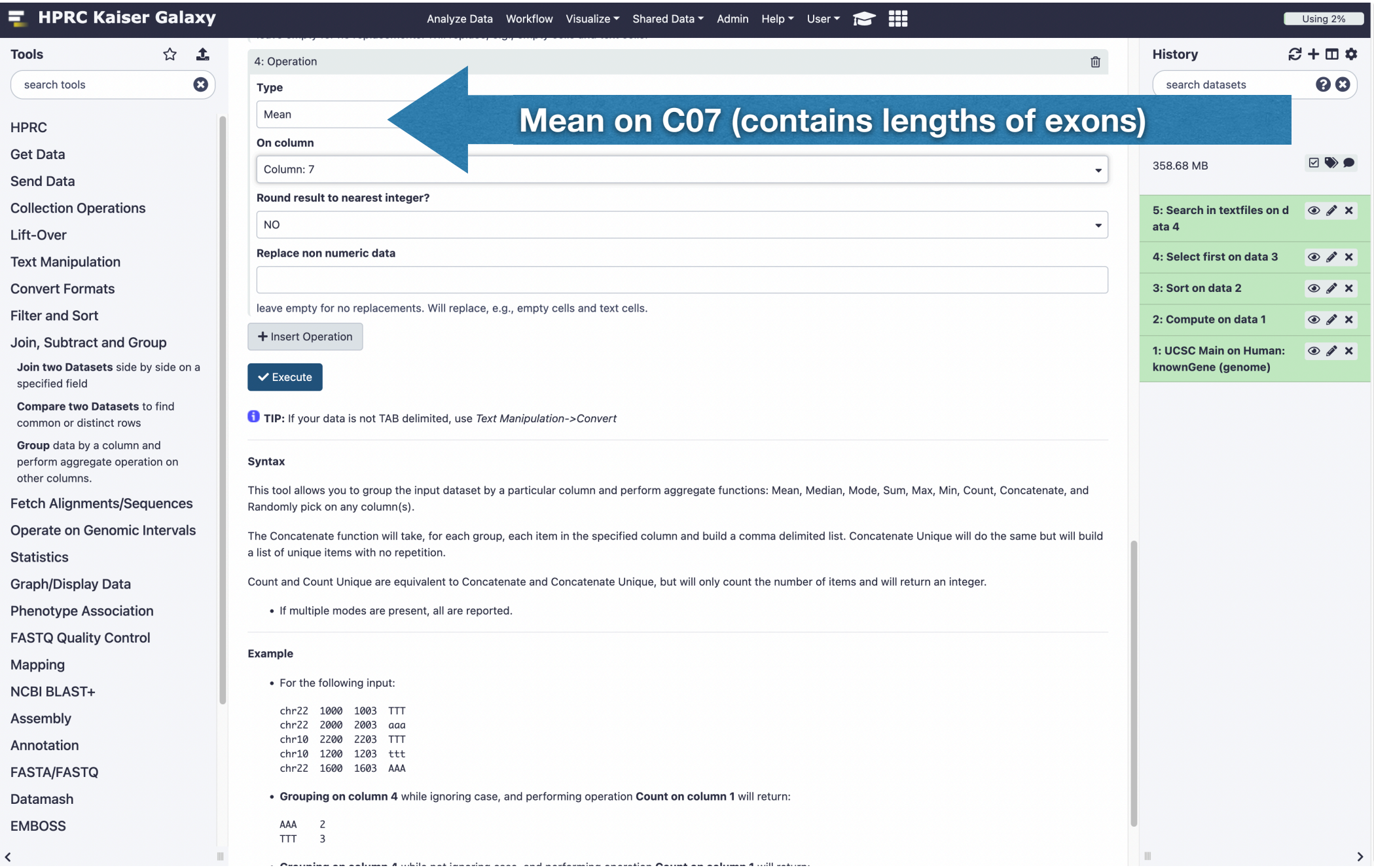


Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Go to: 'Join_Subtract_and_Group' => Group



Computational Arithmetics

Join, Subtract and Group

Running a quick statistics on the ~ first 1000 Exons of the Human Genome

Go to: 'Join_Subtract_and_Group' => Group