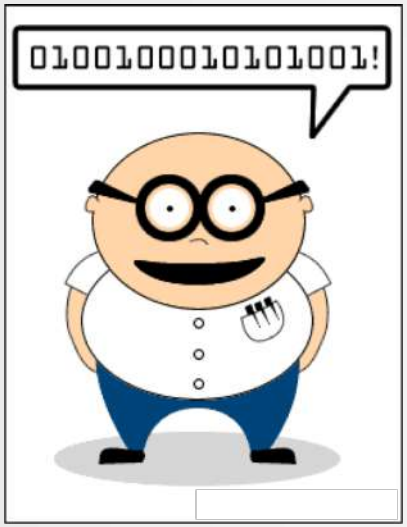
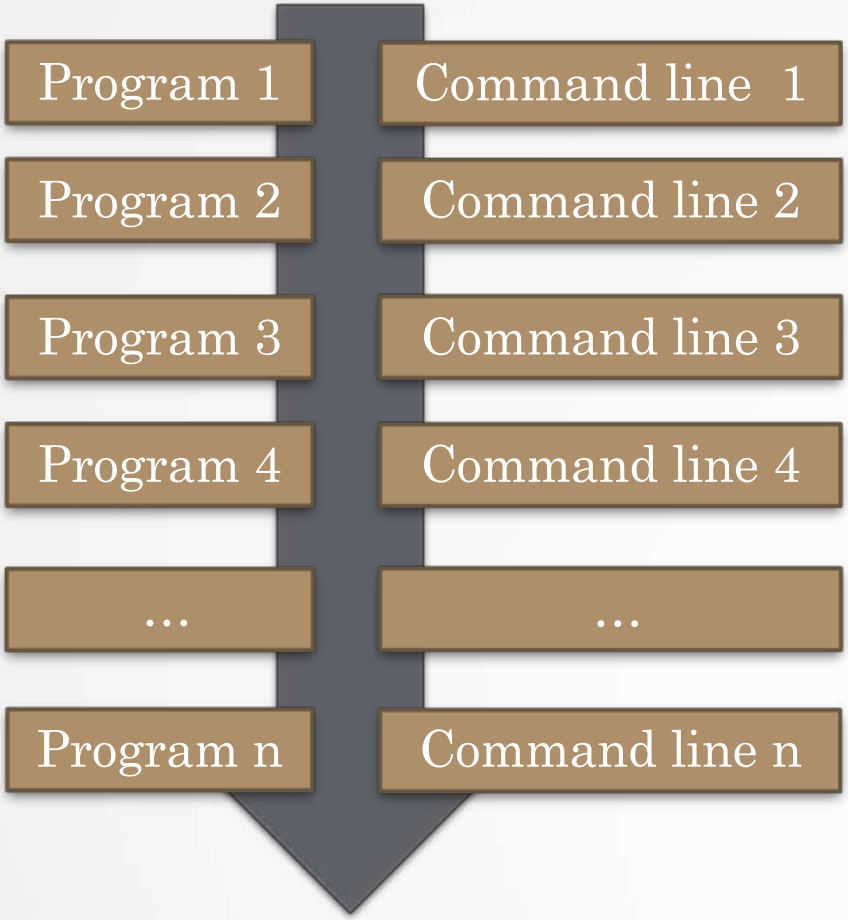
NGS analysis automatization: Galaxy workflows

Stéphanie Le Gras
(slegras@igbmc.fr)

A long time ago…

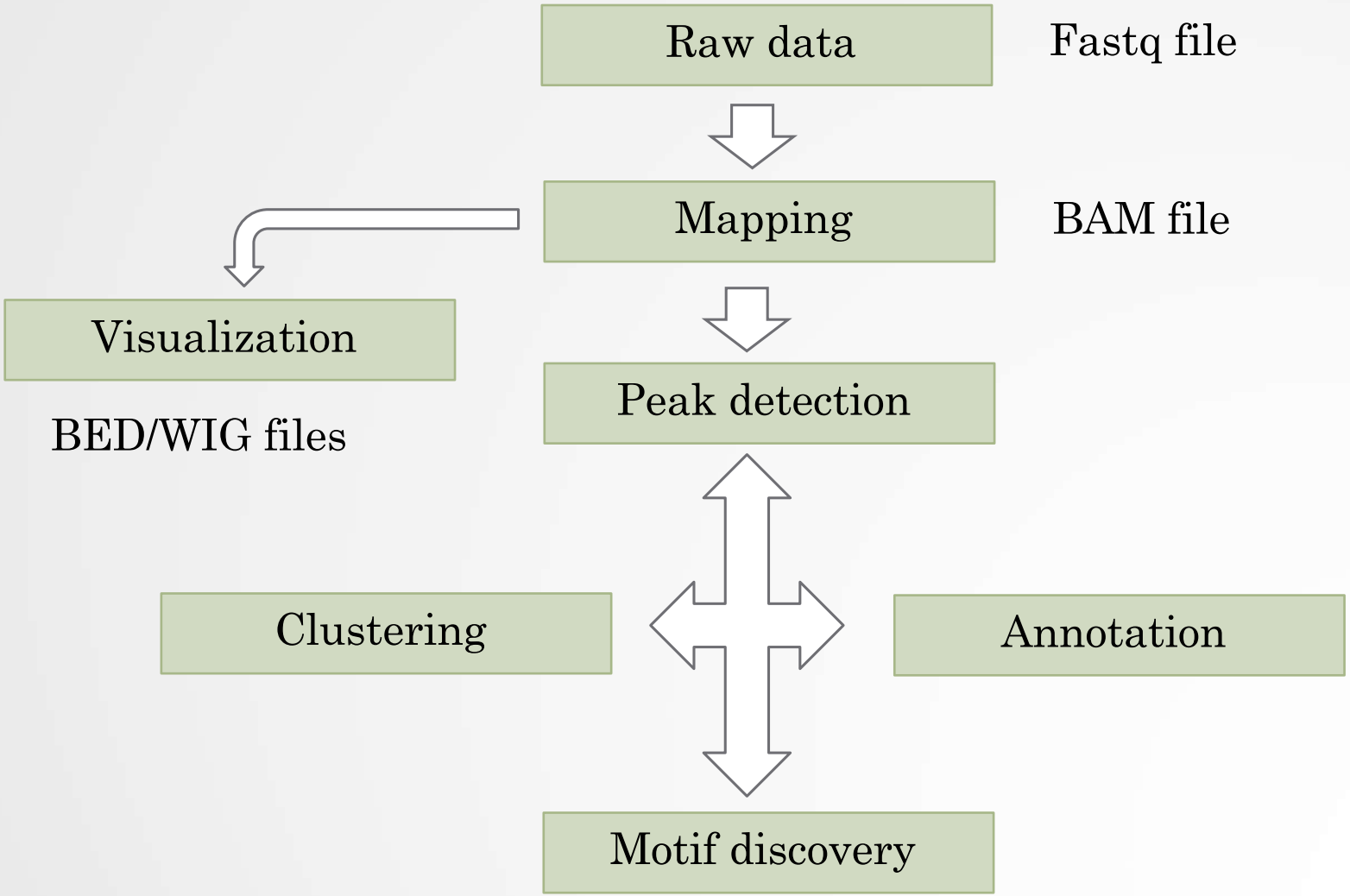


Input data



PIPELINE/
WORKFLOW

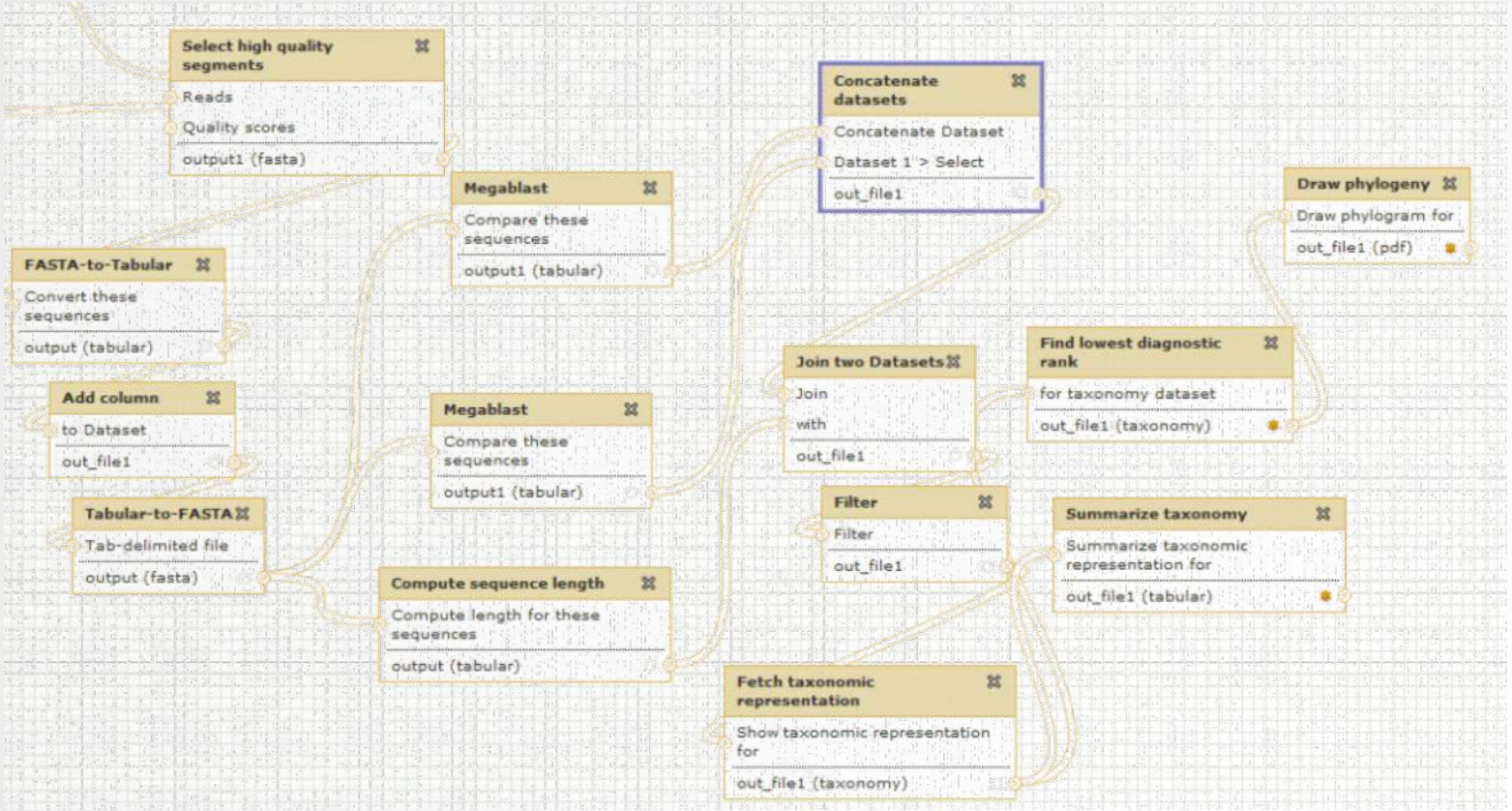
More recently…



During the entire training session..

What if we'd mix all together

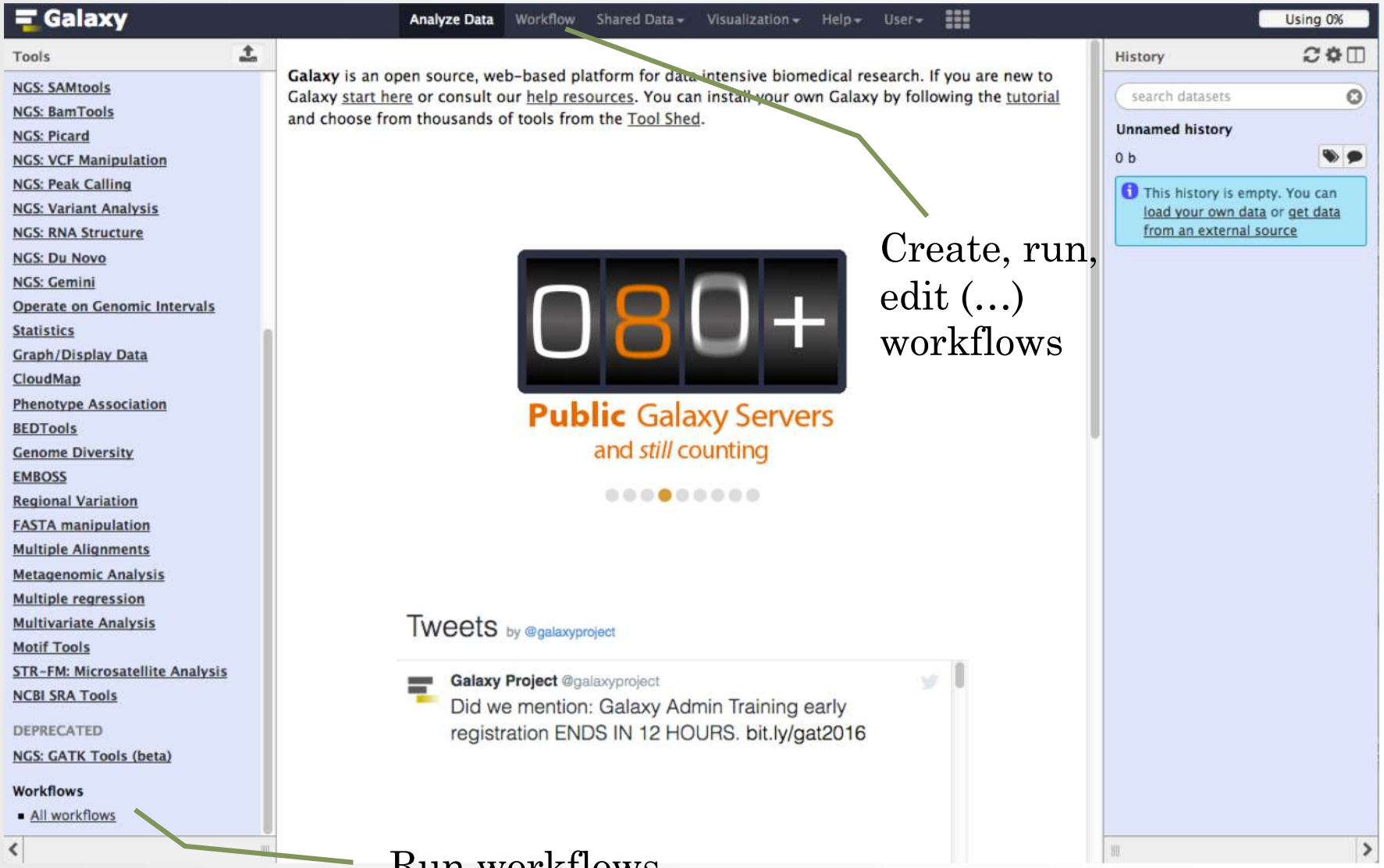
Galaxy workflow



Galaxy workflows

- Workflow:
  - Analysis protocol with several steps (tools)
  - The output of a step is used as the input of the next next so file formats between two steps should be compatible!
- Workflows are often made general so that they can be run on various datasets
- Some of the parameters are pre-defined while others are set at runtime

Workflows

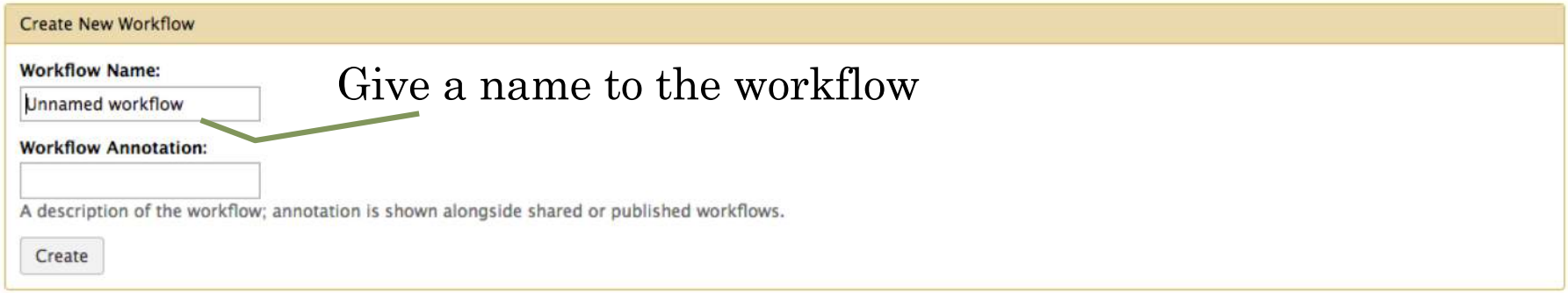


Create, run, edit (…) workflows

Run workflows

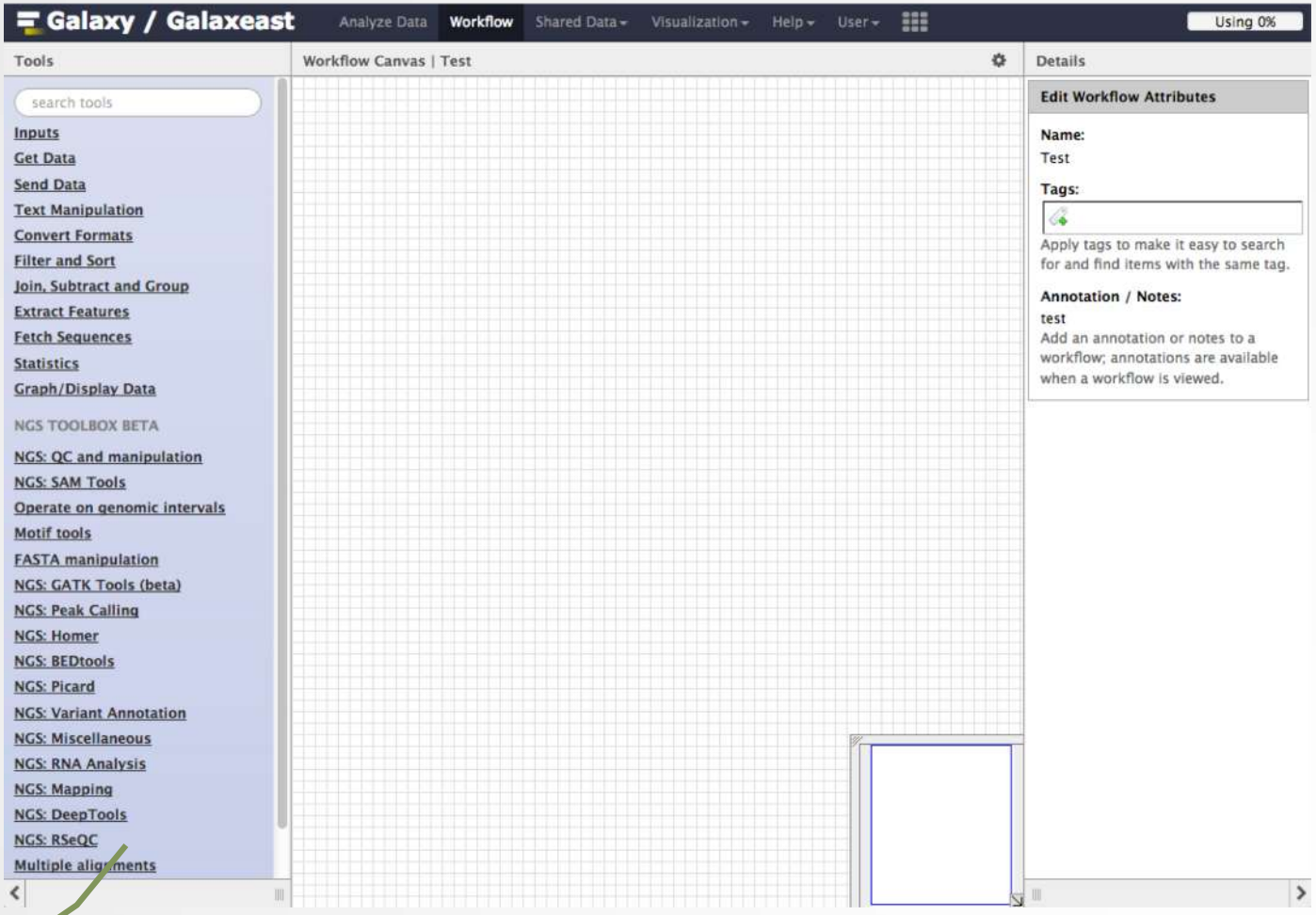
Workflows

Create workflows



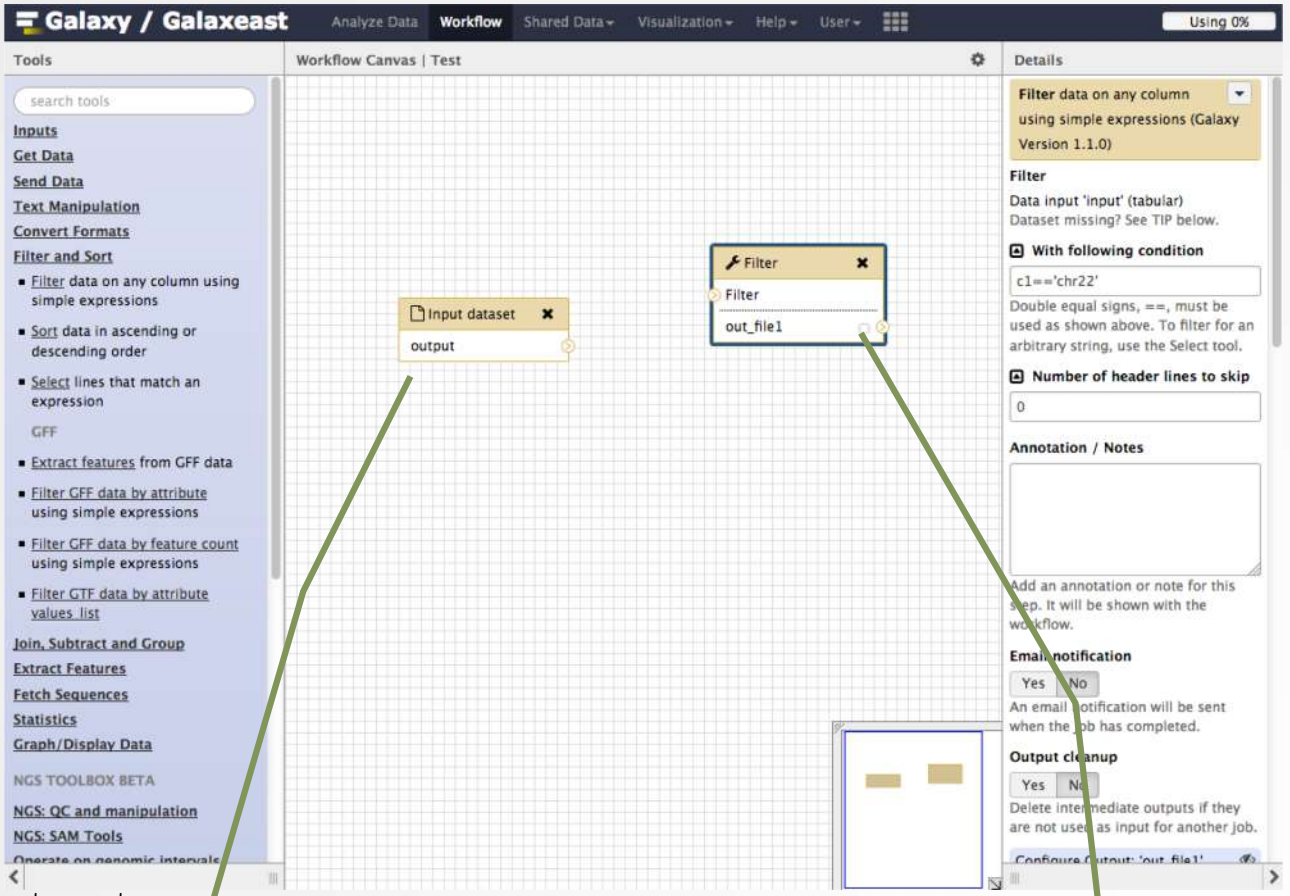
Give a name to the workflow

Workflow creation



Add tools or input datasets to the workflow

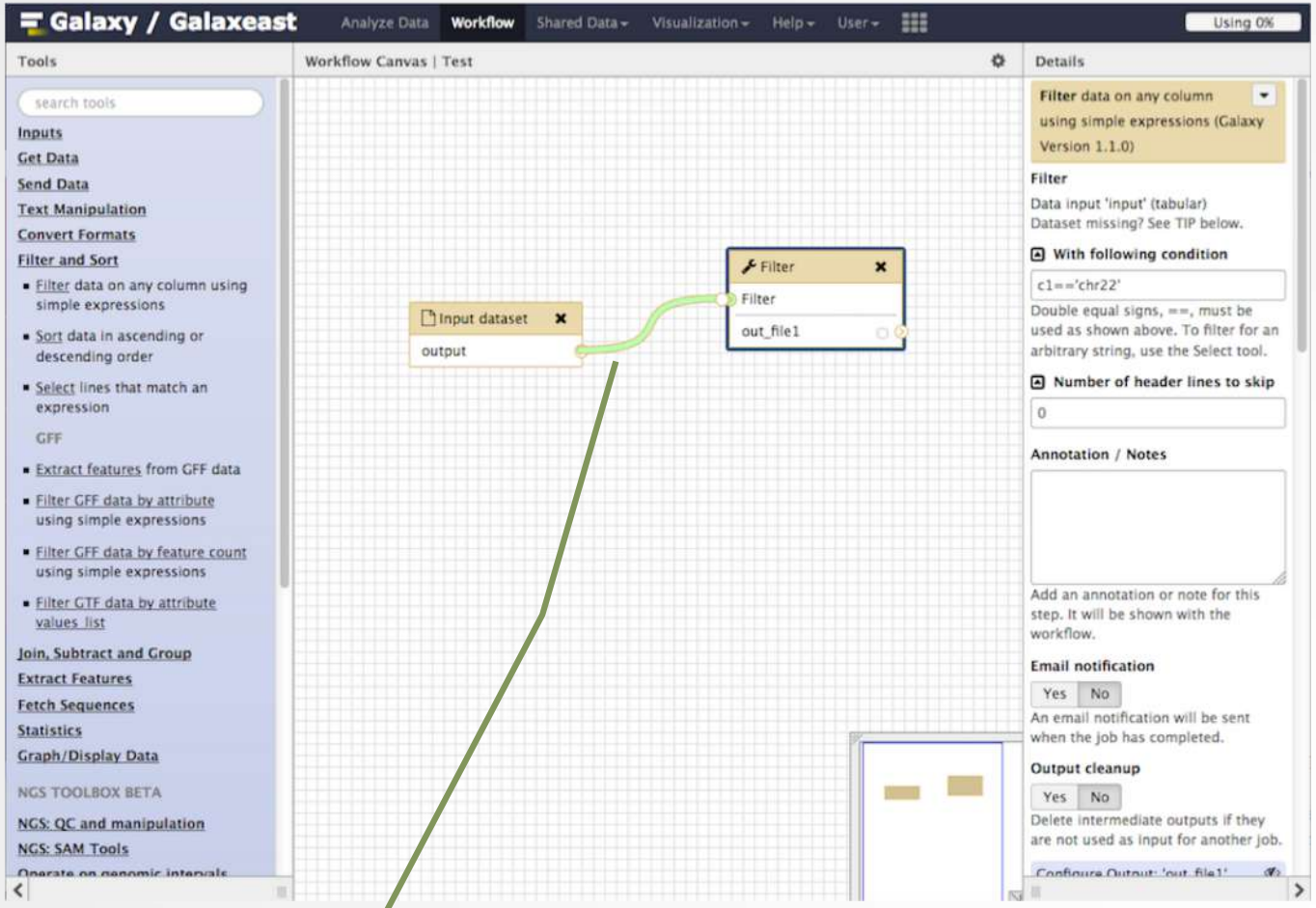
Workflow creation



Input dataset.
Most of the time, a workflow starts with an input dataset to which analyses are applied.
In Galaxy, the file format of the input dataset will be limited to the input file format of the subsequent step

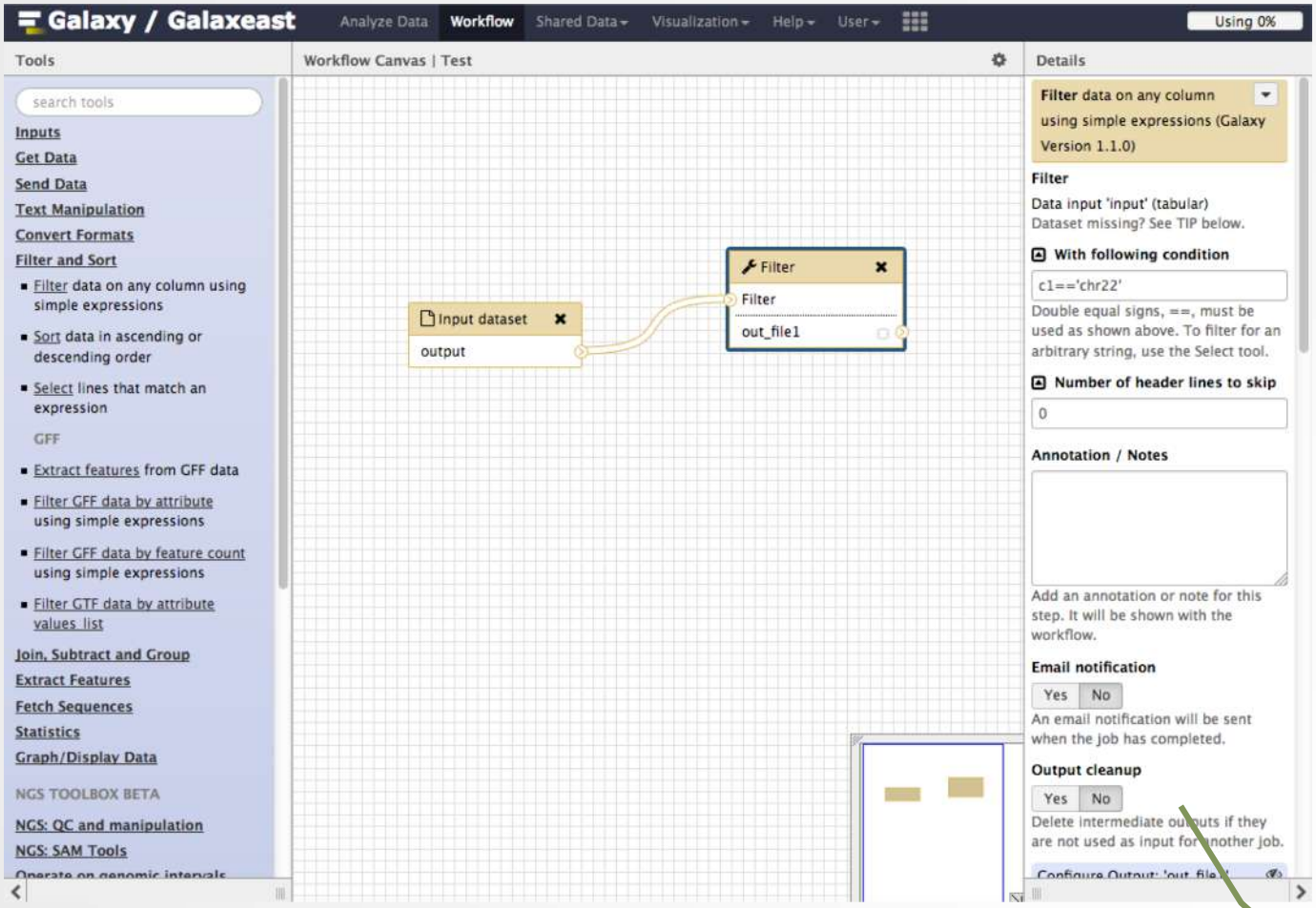
Tool to be run

Workflow creation



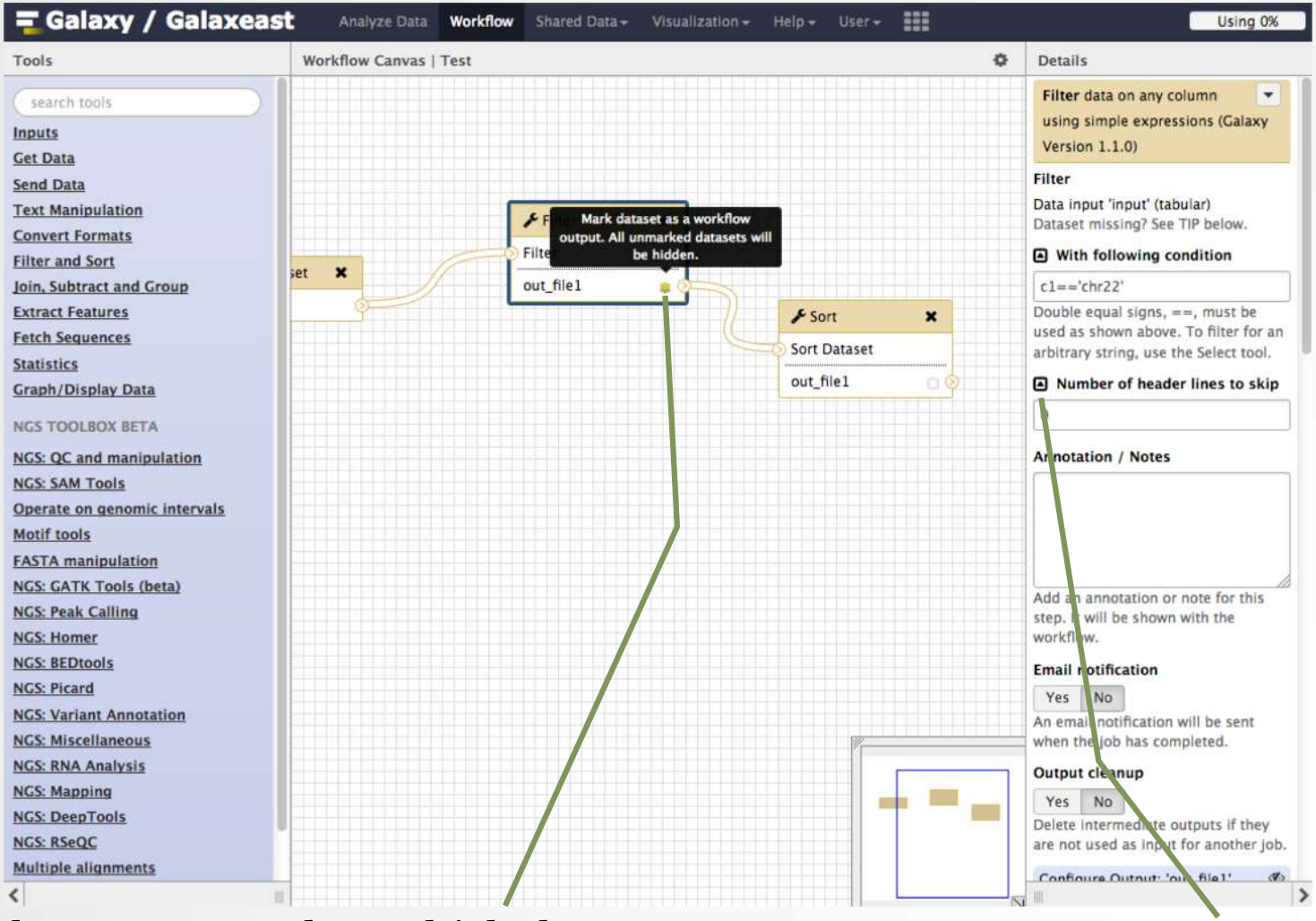
If two steps can be linked together, the link between the two boxes is green

Workflow creation



Pre-configure tool parameters and configure parameters to be set at run time

Workflow creation

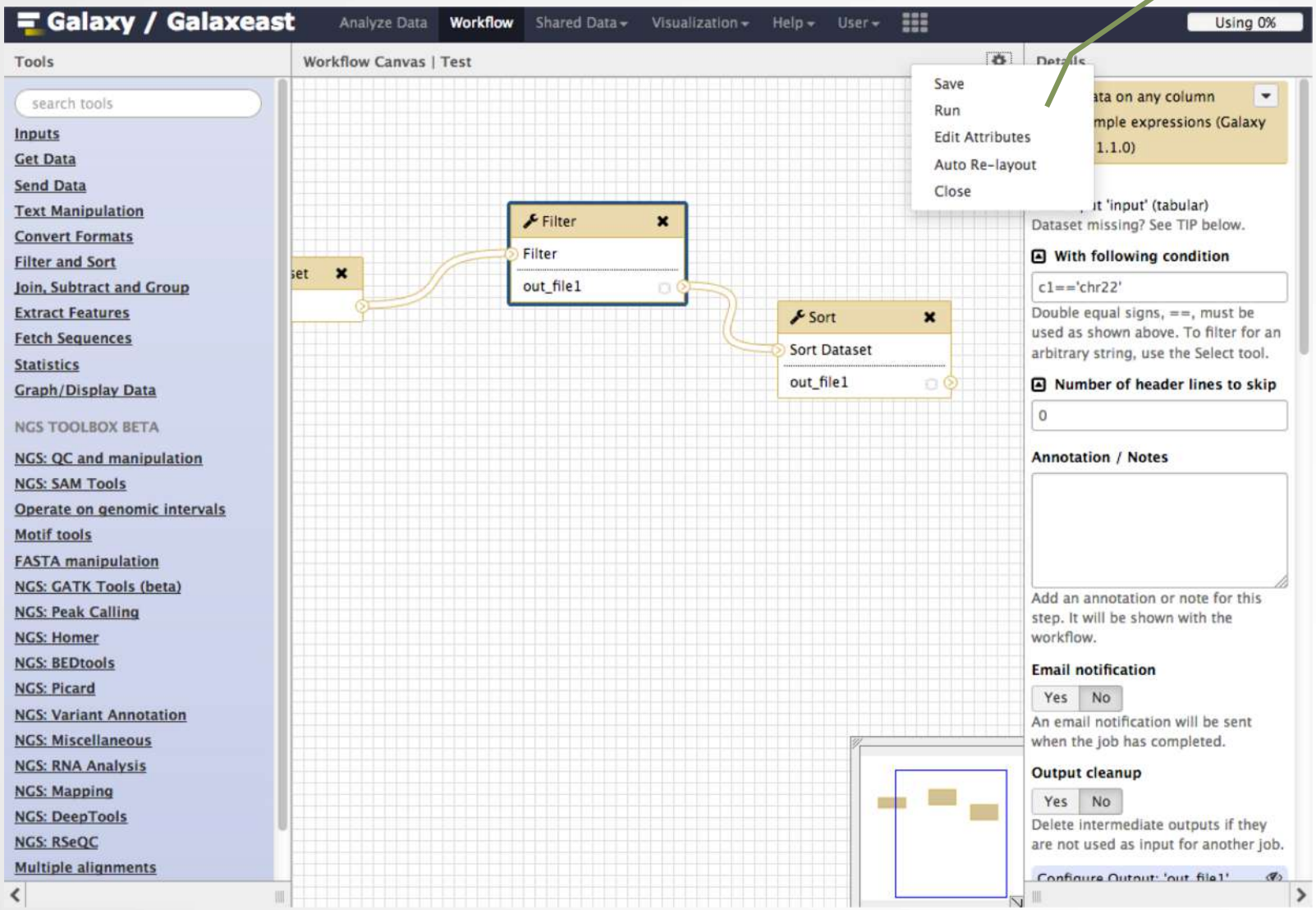


Click on star to select which datasets will be displayed in the history generated when running of the workflow

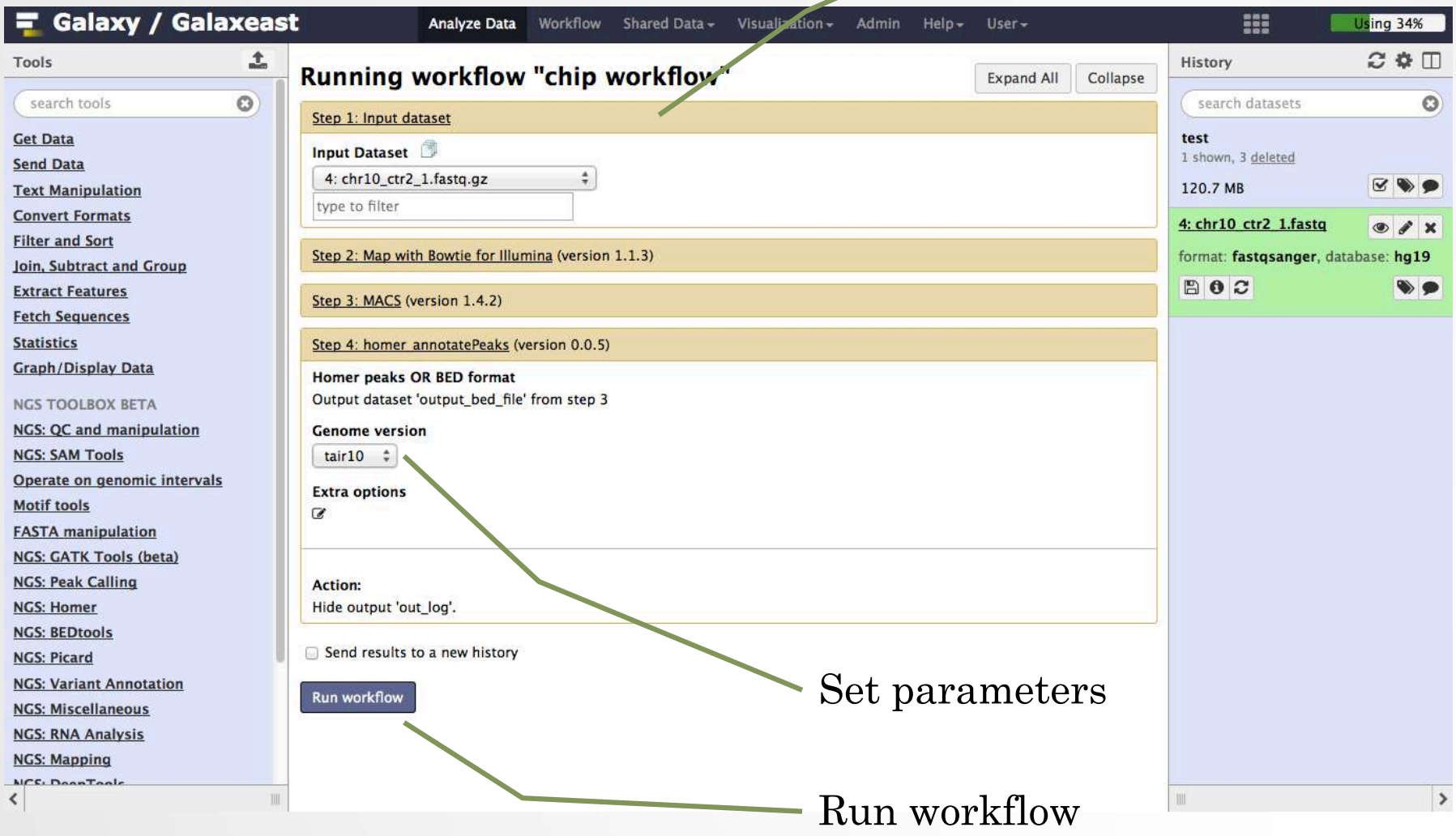
Click to get the parameter to be set at runtime

Workflow creation

Save, run workflows



Run workflows



Exercise: your workflows for NGS data analysis

We want to create a workflow to automatically analyze chIP-seq data in Galaxy.

1. Based on what you've learned during the courses, what would be the steps to implement in the workflow? The workflow must handle two input datasets: a treatment and a control (fastq files)
2. Implement the workflow into Galaxy
3. Import all datasets from the data library NGS data analysis training > ChIPseq > workflow. Run the workflow on the data

We also want to create a workflow for automatic analysis of RNA-seq data in Galaxy

4. What would be the steps, what limitation do you see in implementing RNA-seq data in Galaxy?