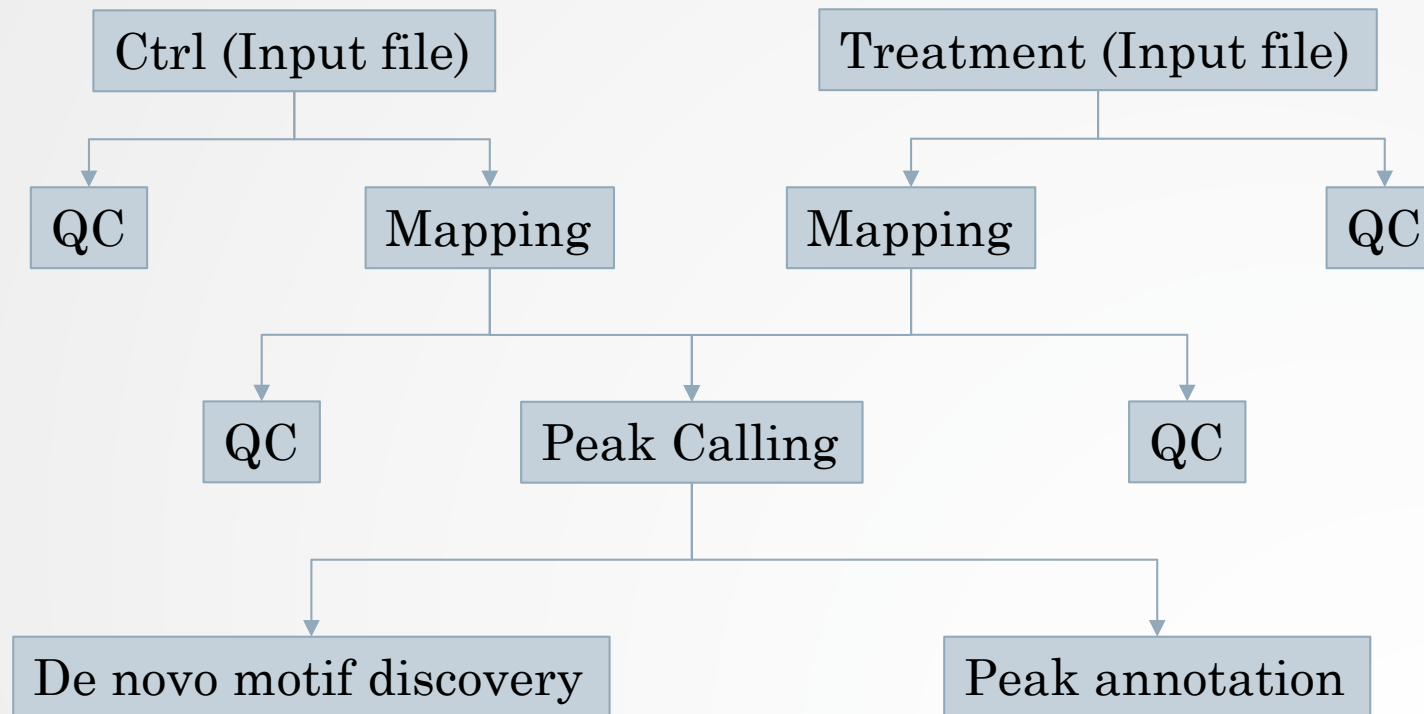


NGS analysis automatization: Galaxy workflows (answers to questions)

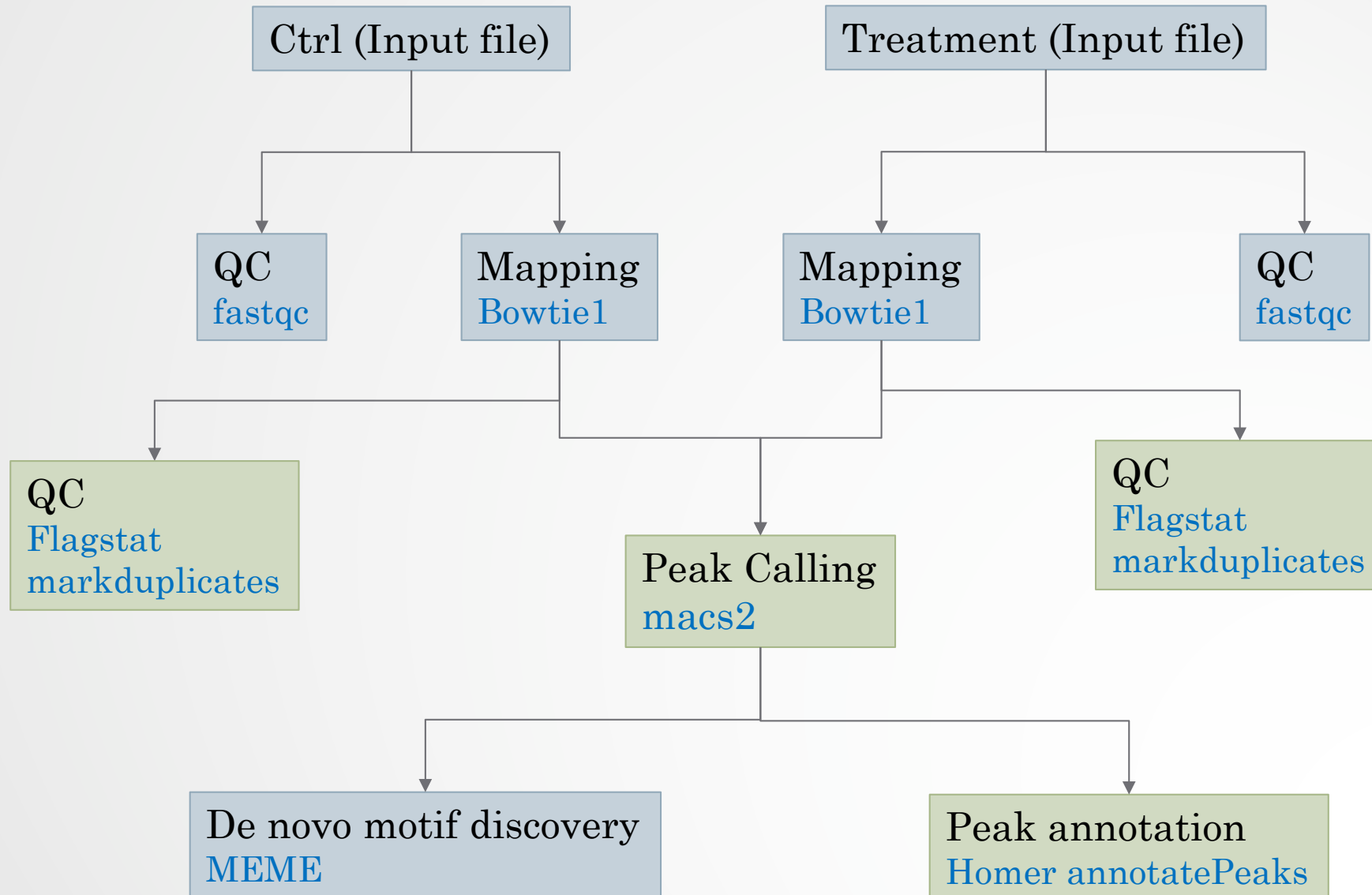
Stéphanie Le Gras
(slegras@igbmc.fr)

Exercise: your workflows for NGS data analysis



1.



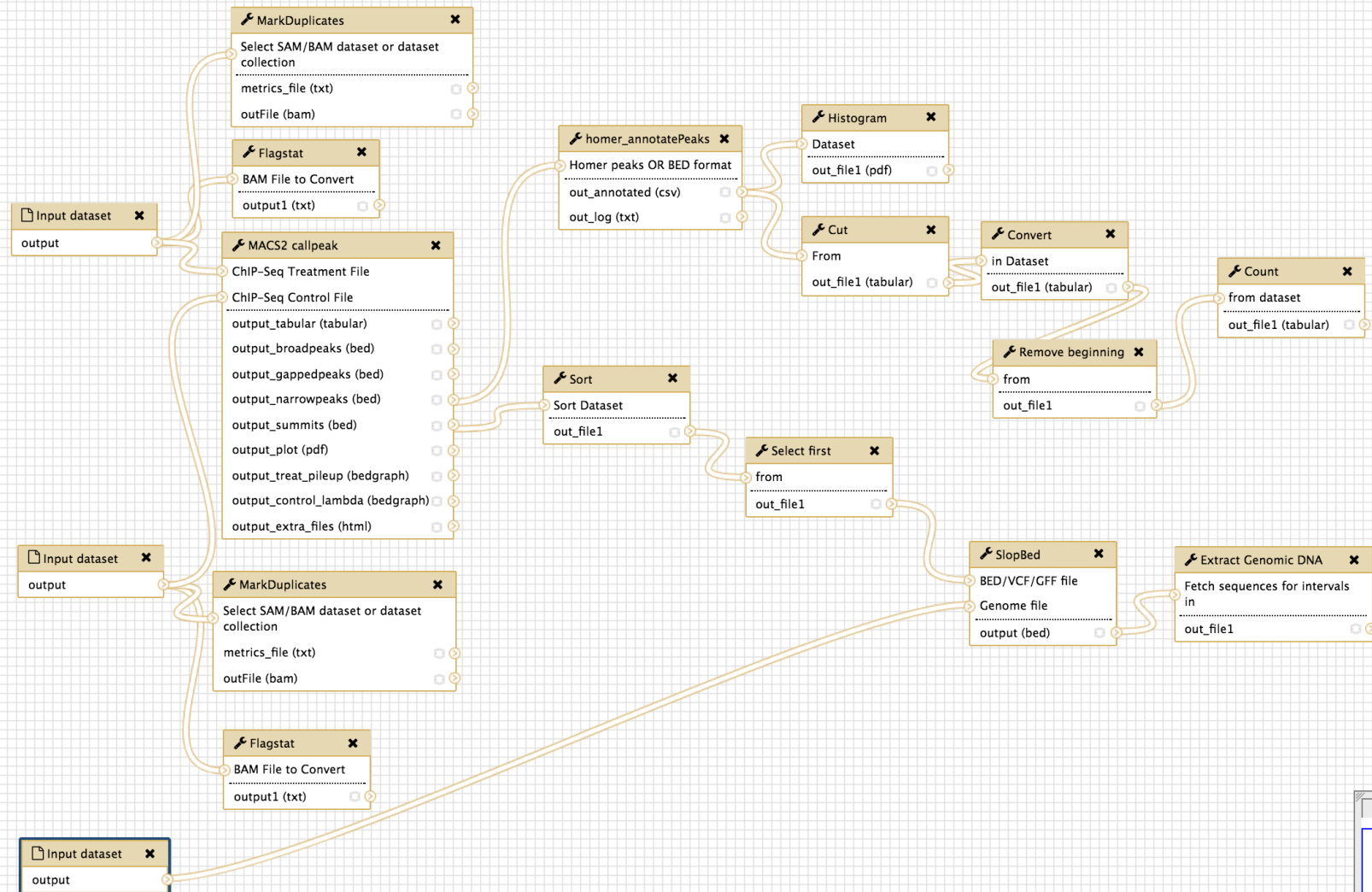
Exercise: your workflows for NGS data analysis



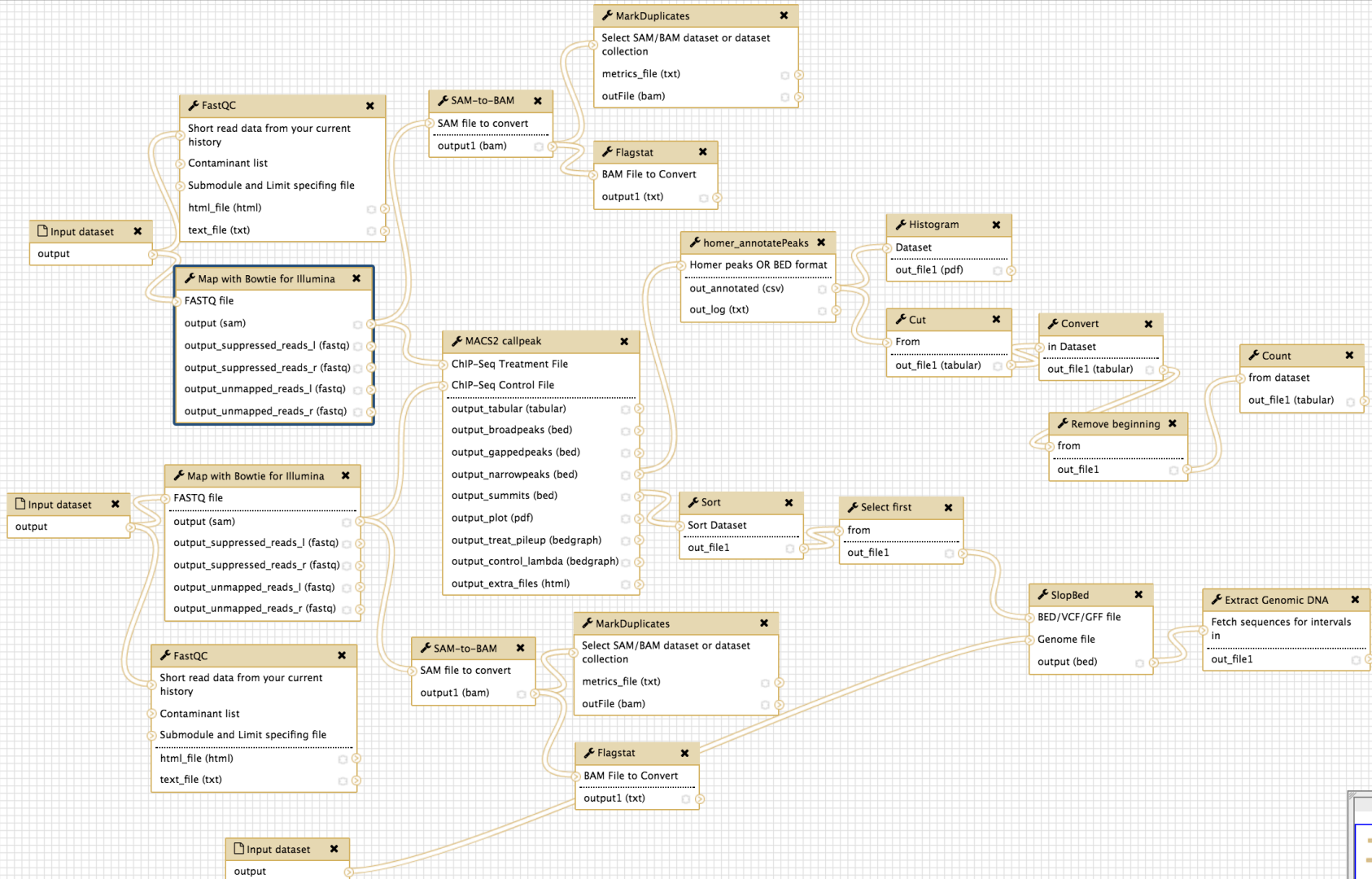
Exercise: your workflows for NGS data analysis

- 2.
 - Go to the history in which you analyzed chIP-seq data (history named “ChIP-seq data analysis”)
 - Click on 
 - Select Extract Workflow
 - Enter the workflow name: “ChIP-seq data analysis”
 - Adapt the workflow steps to the needs:
 - Keep the second MACS2 run
 - Go back to top and click on 
 - Edit the workflow:
 - Click on Workflow (top menu)
 - Click on ChIP-seq data analysis > Edit

Exercise



Exercise



Exercise: your workflows for NGS data analysis

- Name the input datasets
 - Click on the input boxes and enter “Treatment” in the treatment input box and “Control” in the control input box (After having clicked on the input box, enter text in the right panel).

Input dataset

Name:
Treatment


Edit Step Attributes

Annotation / Notes:

Add an annotation or notes to this step; annotations are available when a workflow is viewed.

- Bowtie 1 parameters:
 - Select a reference genome : **set at runtime** **Select a reference genome**
 - Bowtie settings to use: **Full parameters list**
 - Whether or not to make Bowtie guarantee that reported singleton alignments are 'best' in terms of stratum and in terms of the quality values at the mismatched positions (--best): **Use best**
 - Whether or not to report only those alignments that fall in the best stratum if many valid alignments exist and are reportable (--strata): **Use strata**
 - Suppress all alignments for a read if more than n reportable alignments exist (-m): **1**

Exercise: your workflows for NGS data analysis

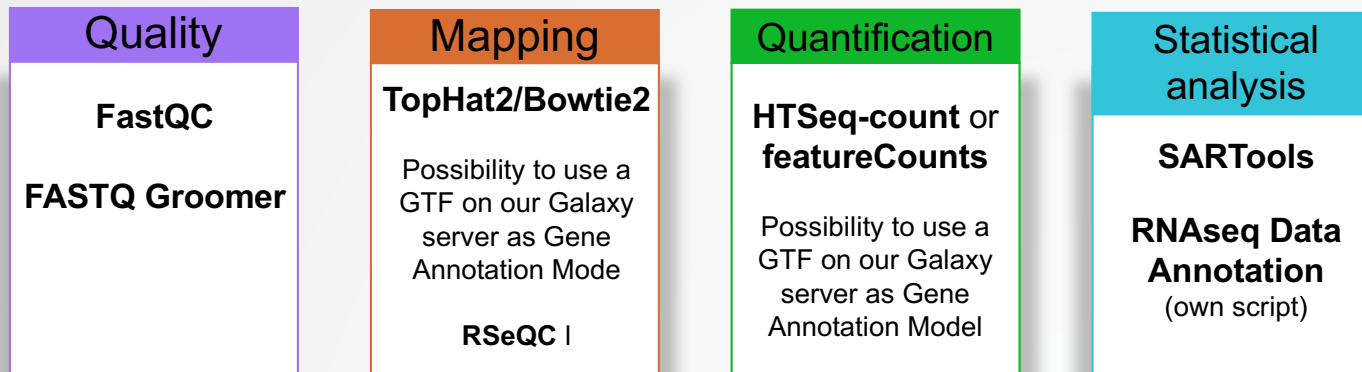
- SAM-to-BAM
 - Reference Genome: set at run time
- MACS2
 - Build Model: **Build the shifting model**
- Select the box of the tool **cut**
 - Click on **configure Output: out_file1**
 - Change datatype: **interval**
- MEME parameters:
 - Options Configuration: **Advanced**
 - Number of different motifs to search: **2**
 - Min width of motif to search: **6**
 - Max width of motif to search: **12**
 - E-value to stop looking for motifs : **1**
 - I certify that I am not using this tool for commercial purposes:
Yes
- Click on  and Save

Exercise: your workflows for NGS data analysis

- 3.
 - Click on Analyze Data (top menu)
 - Go to Shared data > Data Libraries > CNRS training > ChIPseq > workflow and add the two datasets to your history.
 - Tick the two boxes beside the sample names
 - Click on the button To history
 - Add the datasets to the new history “ChIP-seq test workflow”
 - Click on Workflow (top menu)
 - Click on the workflow “ChIP-seq data analysis” and select Run.
 - Treatment: chr10_mitf_2.fastq
 - Control: chr10_ctrl2_1.fastq
 - Len: hg38.len
 - Step 4: Map with Bowtie for Illumina:
 - Select a reference genome: hg38
 - Step 5: Map with Bowtie for Illumina
 - Select a reference genome: hg38
 - Click on Run workflow

Exercise: your workflows for NGS data analysis

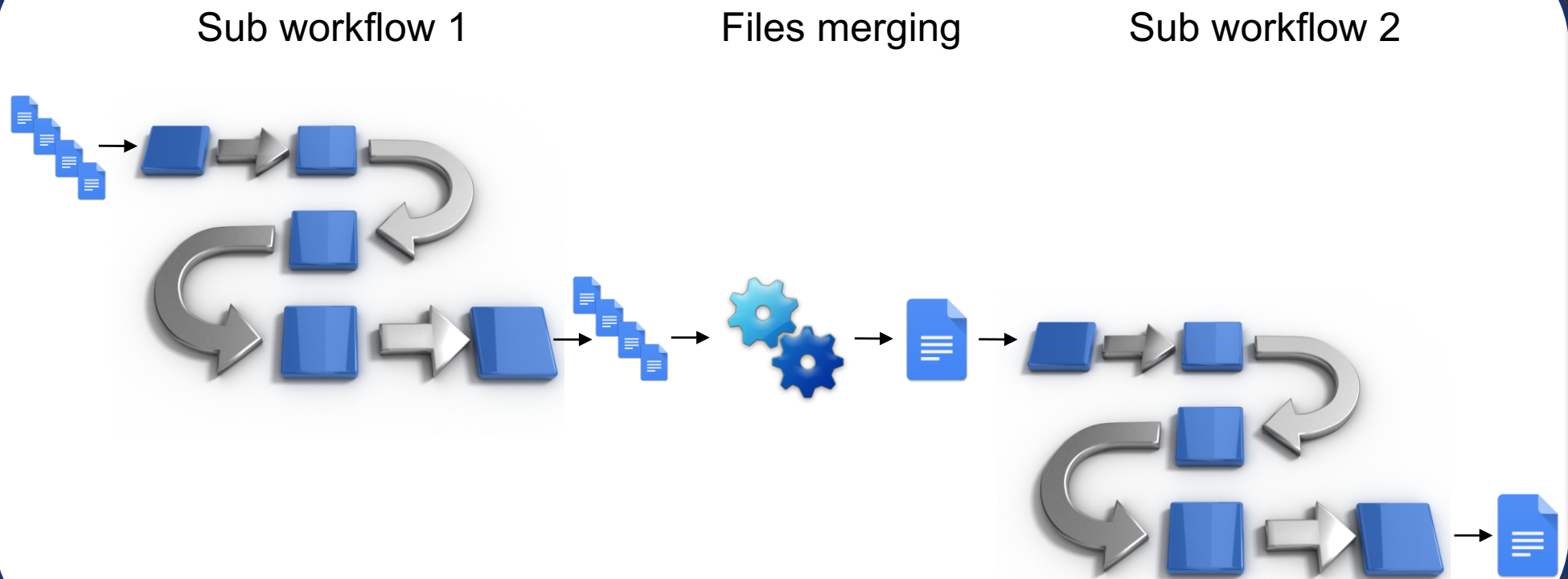
• 4.



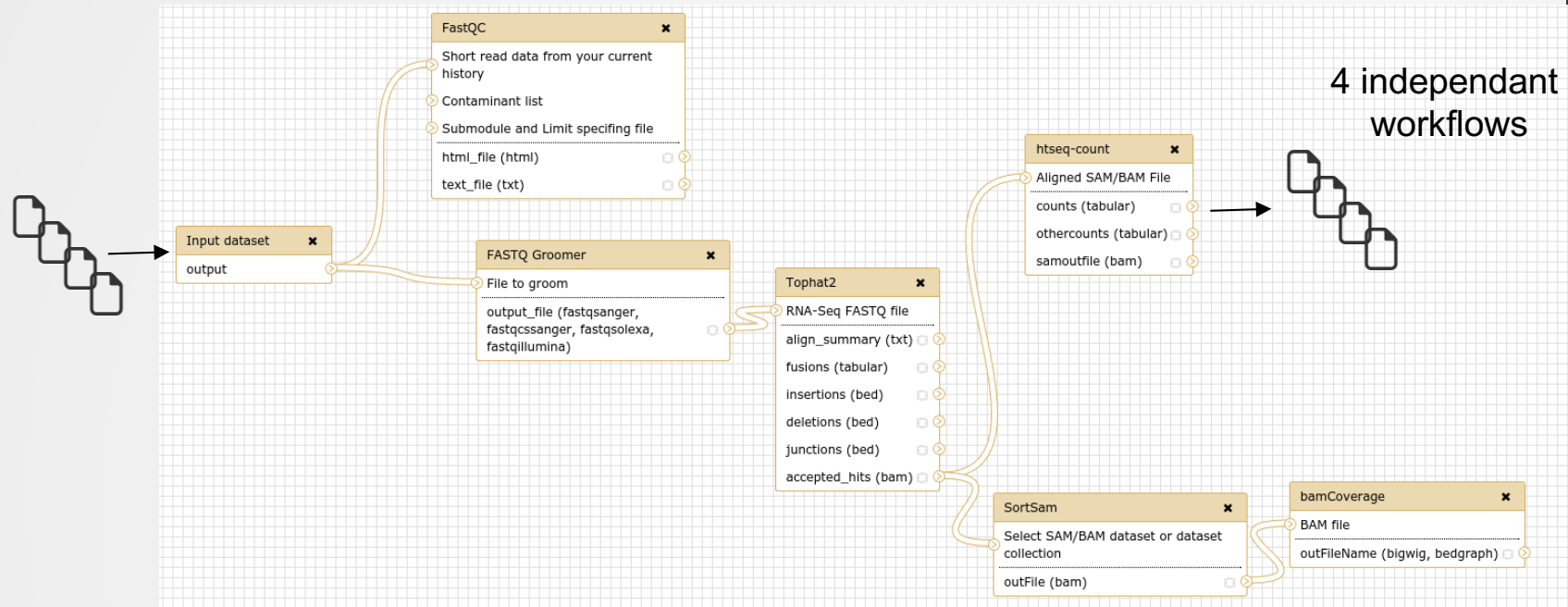
Problem : all steps can't be in a same workflow

RNAseq workflow : limits

Main workflow



RNAseq workflow : limits



HTSeq-count outputs



Merge



SARTools