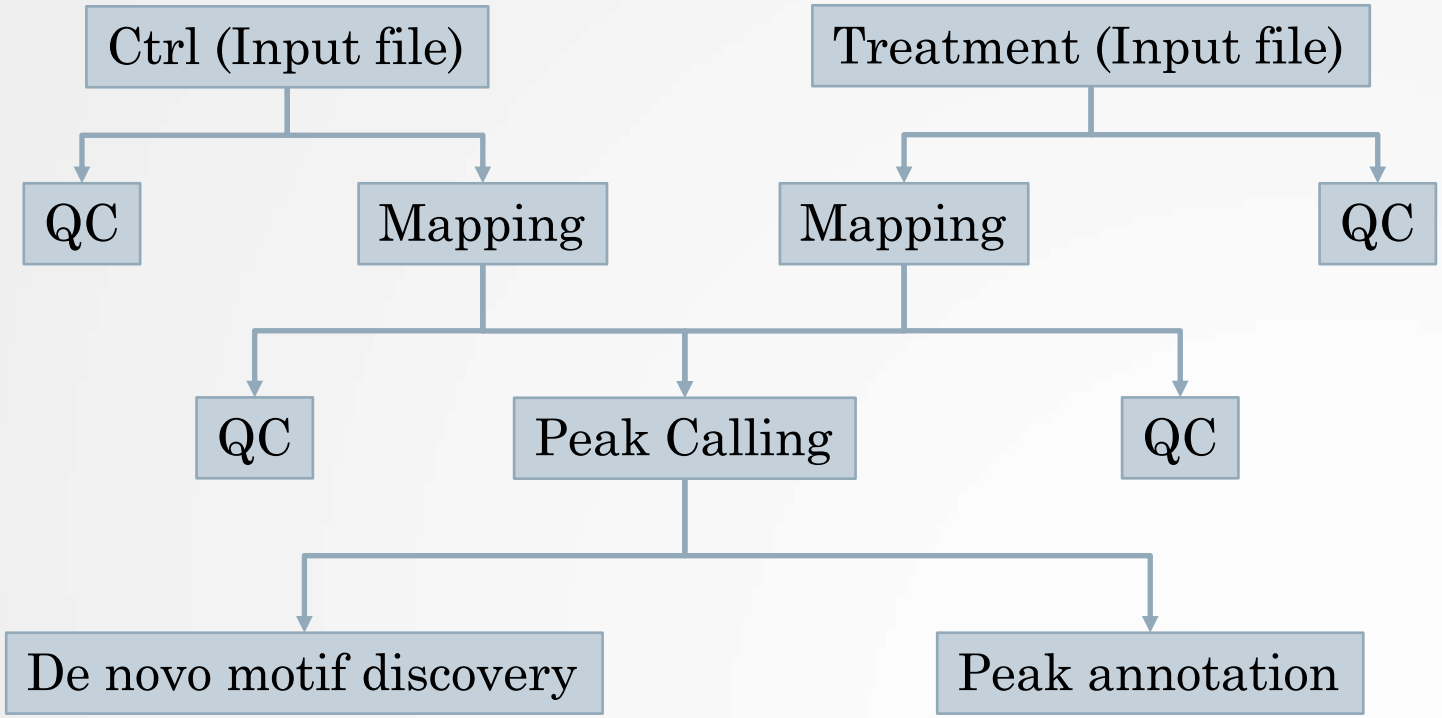
NGS analysis automatization: Galaxy workflows

(answers to questions)

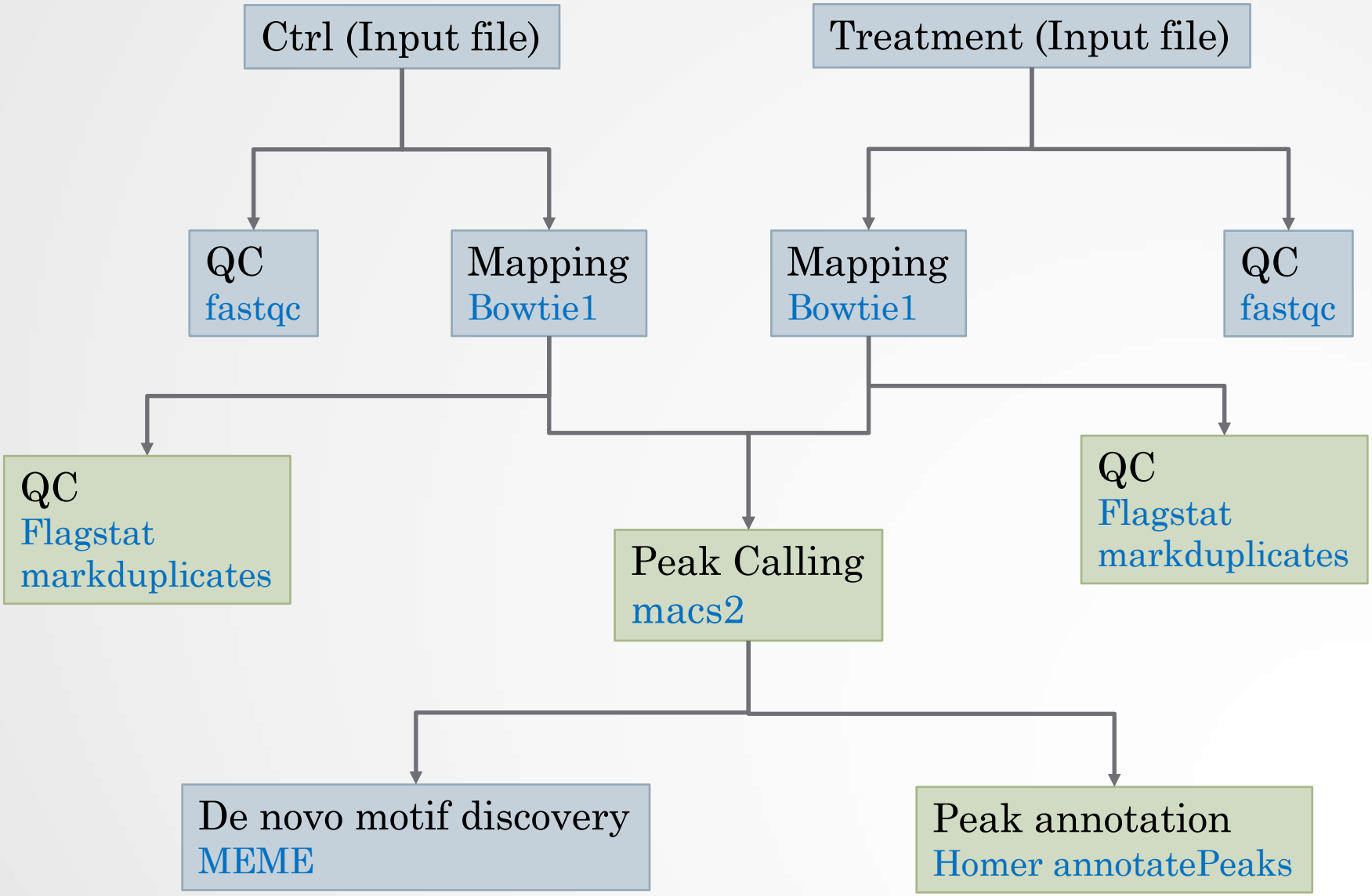
Stéphanie Le Gras
(slegras@igbmc.fr)

Exercise: your workflows for NGS data analysis

1.



Exercise: your workflows for NGS data analysis



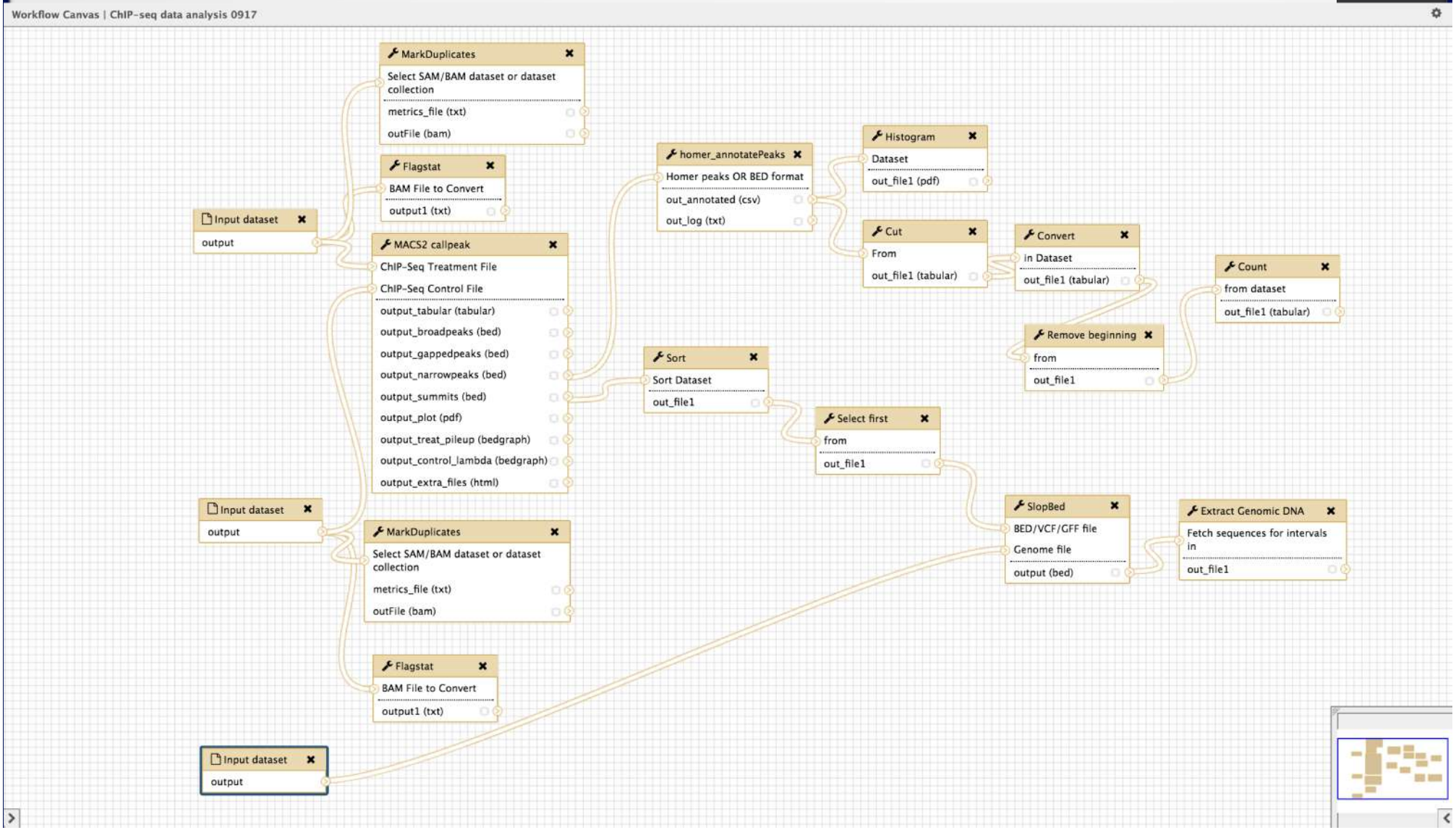
Exercise: your workflows for NGS data analysis

- 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

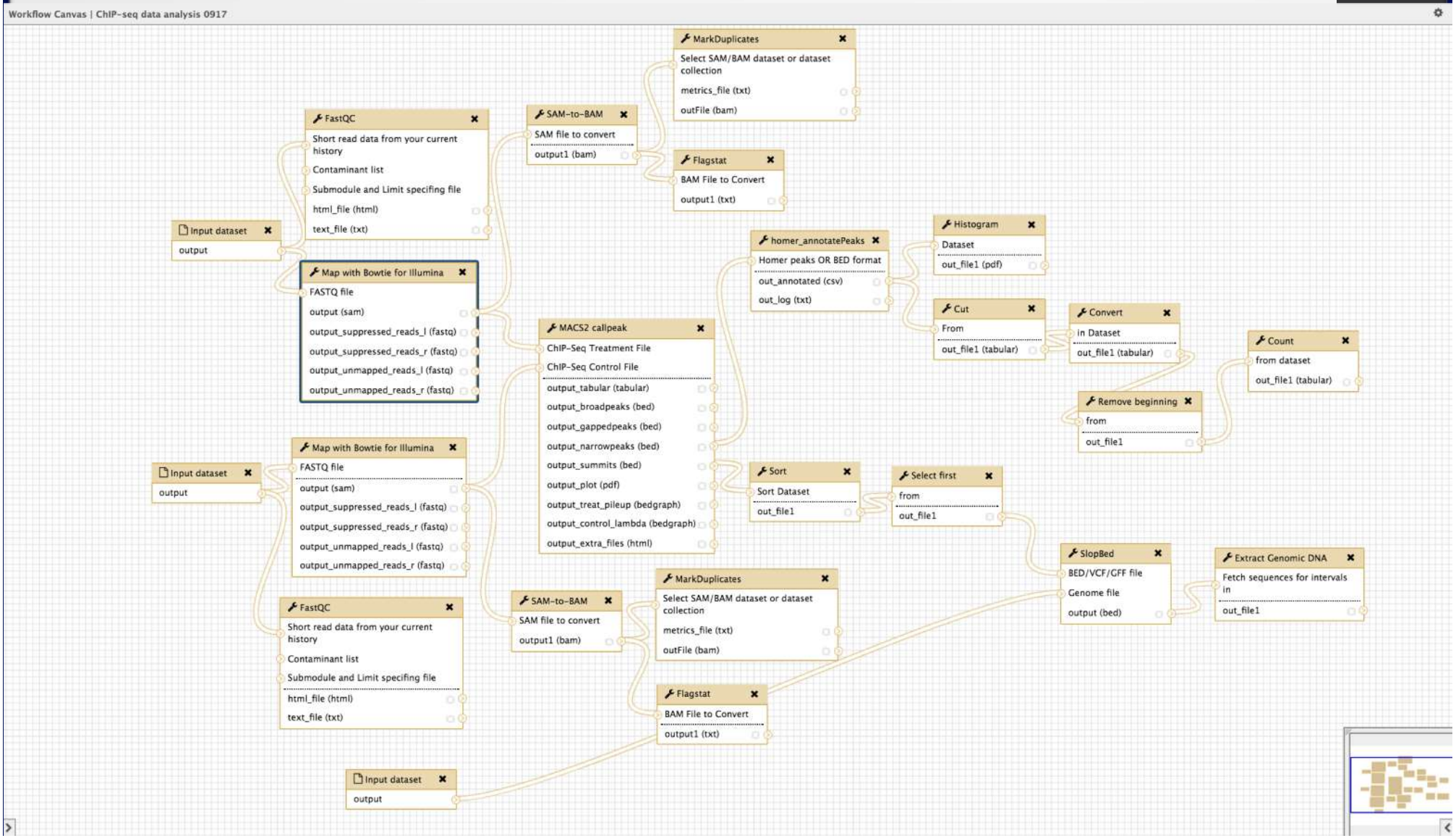
  Hint: You can give a name to input datasets to know what kind of file/data is expected to run the workflow. Name the input datasets:
    - “Treatment” for IP input sample
    - “Control” for control sample
    - “Chrom length” for chromosome length dataset
  - Click on 

  - Then to edit the workflow:
    - Click on Workflow (top menu)
    - Click on ChIP-seq data analysis > Edit

Exercise (before editing)



Exercise (after editing)



Exercise: your workflows for NGS data analysis

- Bowtie 1 parameters:
  - Select a reference genome : **set at runtime**
  - 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**

  Hint: Do it for the two alignment steps

- SAM-to-BAM
  - Reference Genome: set at run time
- MACS2
  - Build Model: **Build the shifting model**

Exercise: your workflows for NGS data analysis

- Homer annotatePeaks
  - Genome version: set at run time
- 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

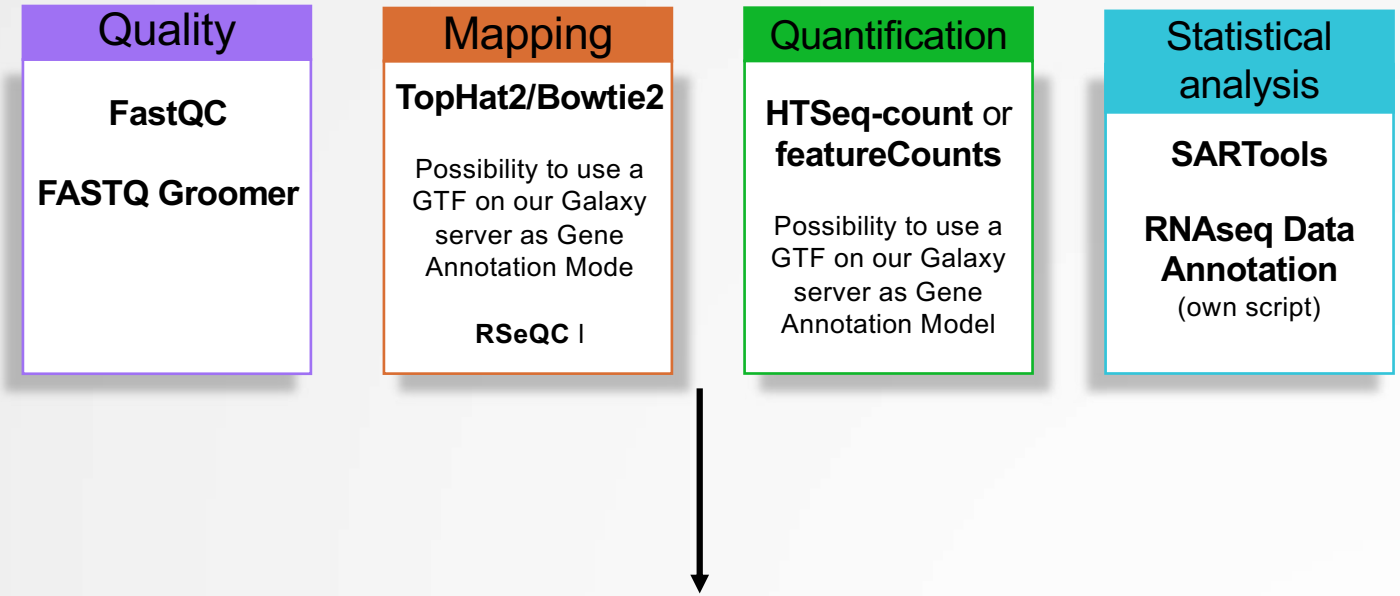
- 3.
  - Click on Analyze Data (top menu)
  - Go to Shared data > Data Libraries > NGS data analysis training > ChIPseq > workflow and add the two datasets to your history.
    - Import all data
    - Click on the button To history
    - Add the datasets to the new history “ChIP-seq test workflow”
  - 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
    - Chrom length: 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
    - Step 13: Homer annotatePeaks
      - Genome version: hg38
    - Click on Run workflow

Exercise: your workflows for NGS data analysis

- 3.
  - Click on the workflow “ChIP-seq data analysis” and select Run.
    - Treatment: chr10_mitf_2.fastq
    - Control: chr10_ctrl2_1.fastq
    - Chrom length: 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
    - Step 6: Sam-to-BAM
      - Using reference genome: hg38
    - Step 7: Sam-to-BAM
      - Using reference genome: hg38
    - Step 13: Homer annotatePeaks
      - Genome version: hg38
    - Click on Run workflow

Exercise: your workflows for NGS data analysis

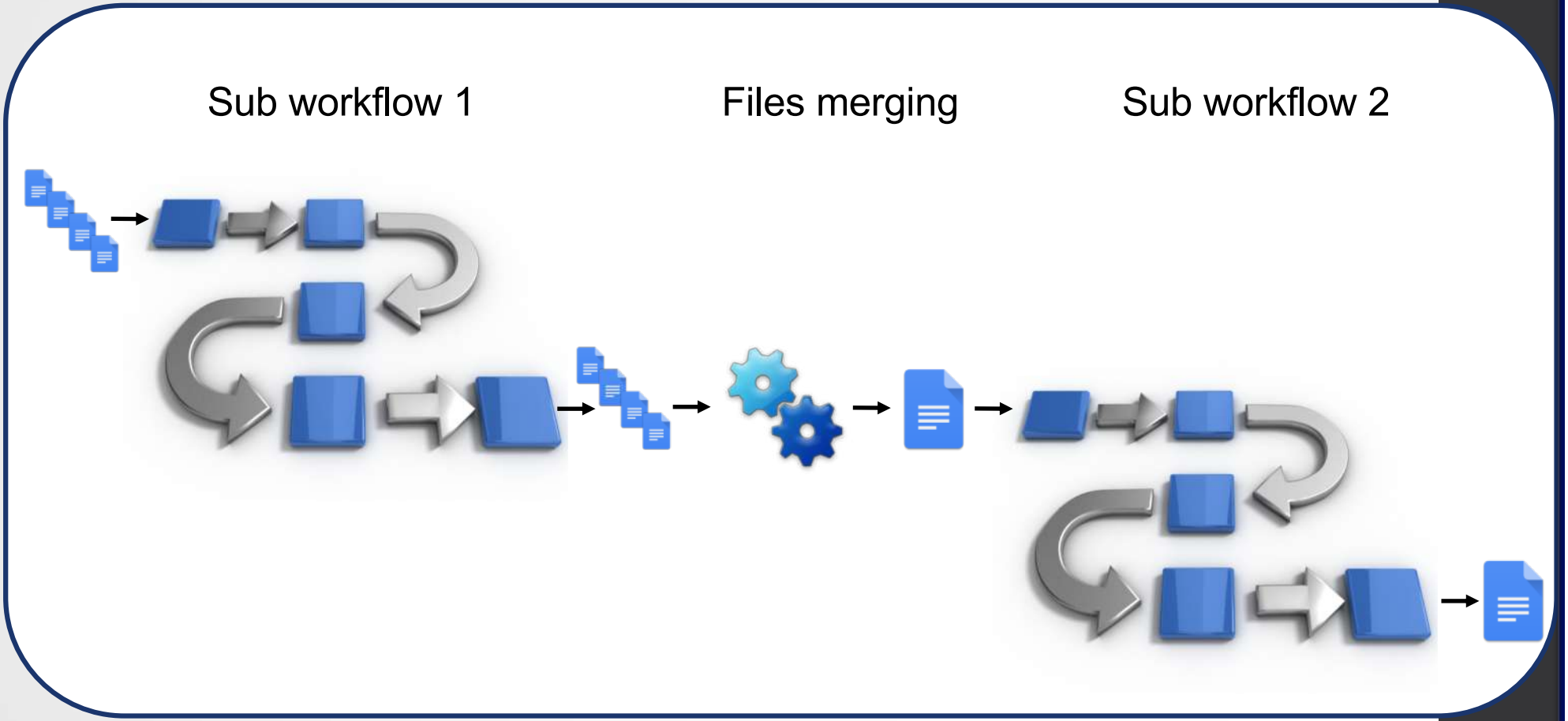
- 4.



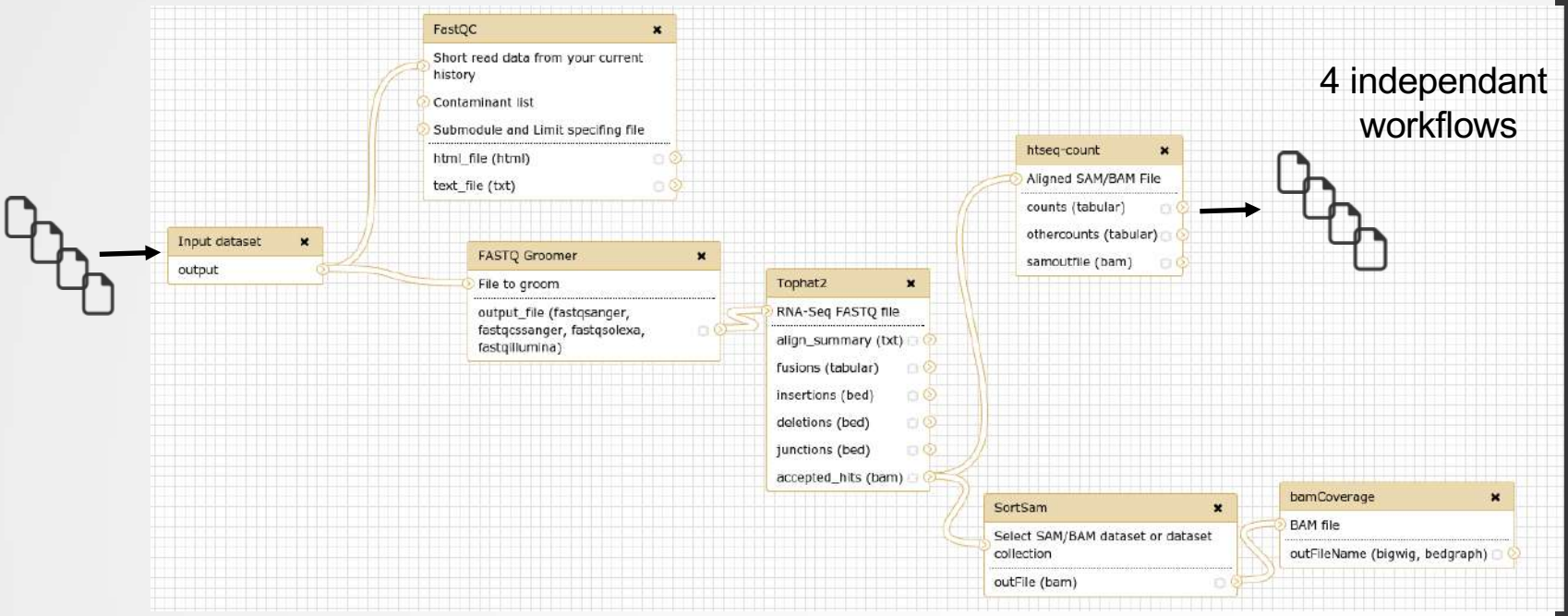
Problem : all steps can't be in a same workflow

RNAseq workflow : limits

Main workflow



RNAseq workflow : limits



HTSeq-count outputs

→ Merge → SARTools