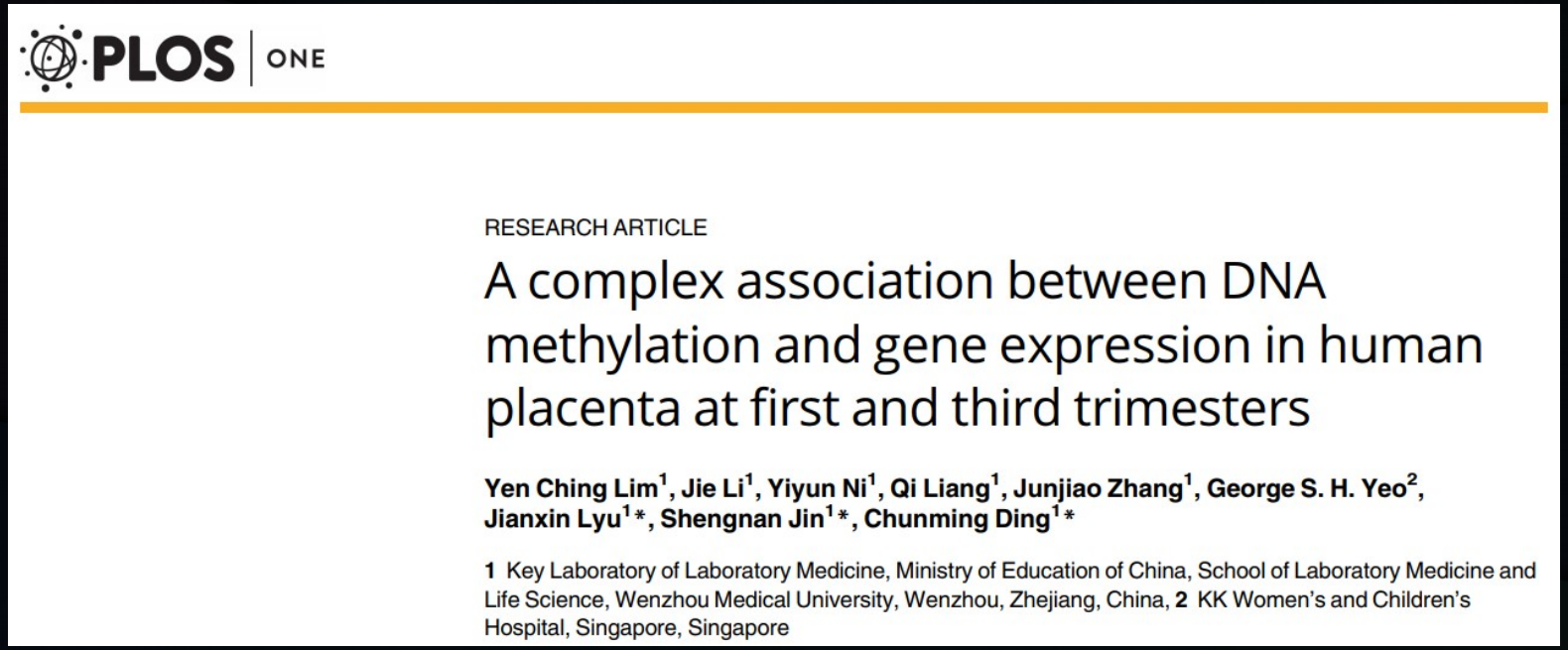
RNA-Seq analysis using Galaxy

Konstantin Riege, Leibniz Institute on Aging (FLI)

Correlation with gene expression



PLOS ONE

RESEARCH ARTICLE

A complex association between DNA methylation and gene expression in human placenta at first and third trimesters

Yen Ching Lim[1], Jie Li[1], Yiyun Ni[1], Qi Liang[1], Junjiao Zhang[1], George S. H. Yeo[2], Jianxin Lyu[1]*, Shengnan Jin[1]*, Chunming Ding[1]*

1 Key Laboratory of Laboratory Medicine, Ministry of Education of China, School of Laboratory Medicine and Life Science, Wenzhou Medical University, Wenzhou, Zhejiang, China, **2** KK Women's and Children's Hospital, Singapore, Singapore

Correlation with gene expression

- 1st trimester → 3rd trimester
  - ~200 diff. meth. promoter regions
  - ~2200 diff. meth. genes
    → Hypermethylation for > 90%
  - 2400 diff. expressed genes
    → Downregulation for ~80%
    → 25 with diff. Meth. promoter regions

Data download

https://share.leibniz-fli.de/index.php/s/mrWaNafRMnLc7be

25 DEG with DME

- Data preperation for 22 protein coding genes as listed in GENCODE – Ensembl v92
  - Minimalistic HG38 derived genome representation
    - Genes & 21 x 4kb intergenic regions: Fasta format
  - Minimalistic annotation
    - Genes: GTF format
    - Promoter - 1kb gene upstream regions: BED format

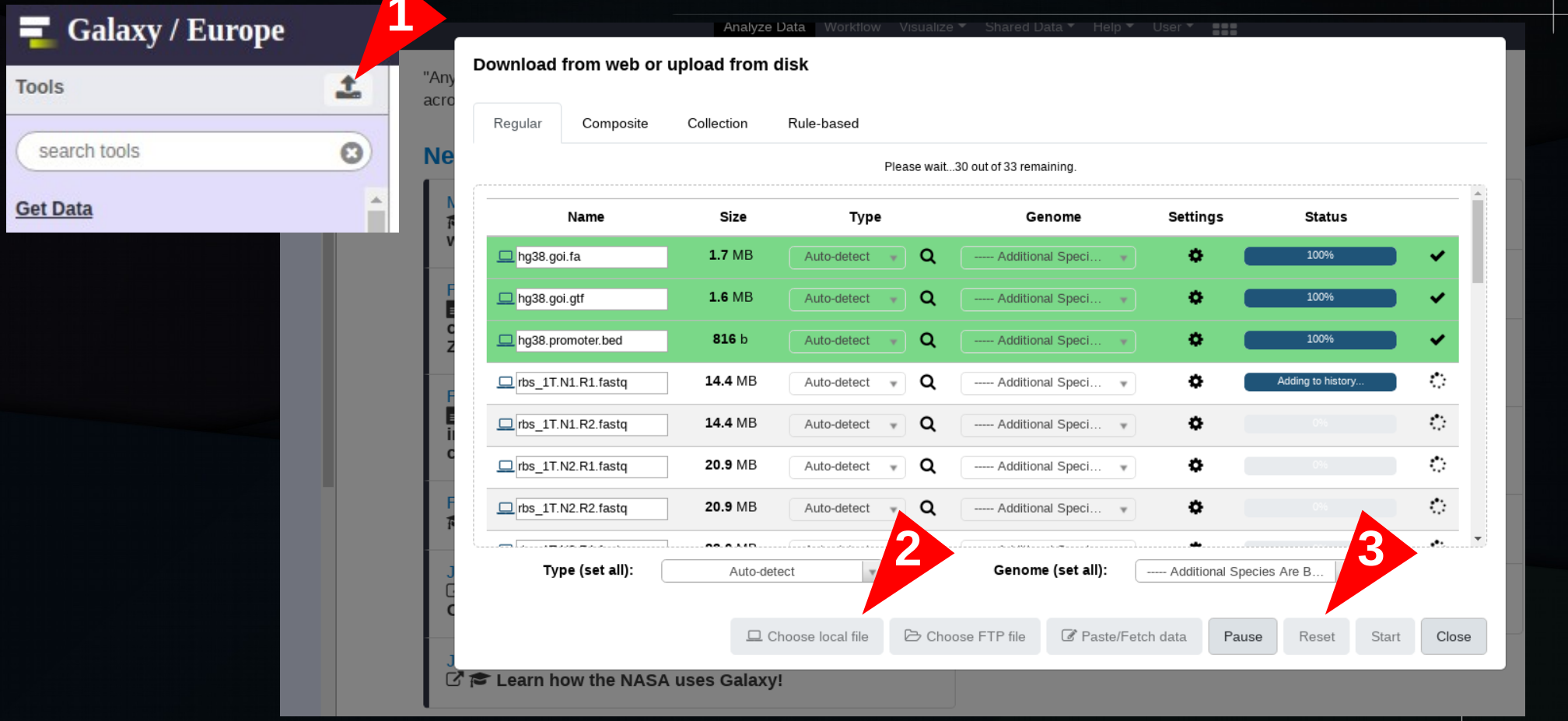
25 DEG with DME

- Data preperation for RNA-Seq
  - Preprocessing & mapping of
    - RNA-Seq data
      - → Subsampling to 5%
      - → extraction from RAW FASTQ data

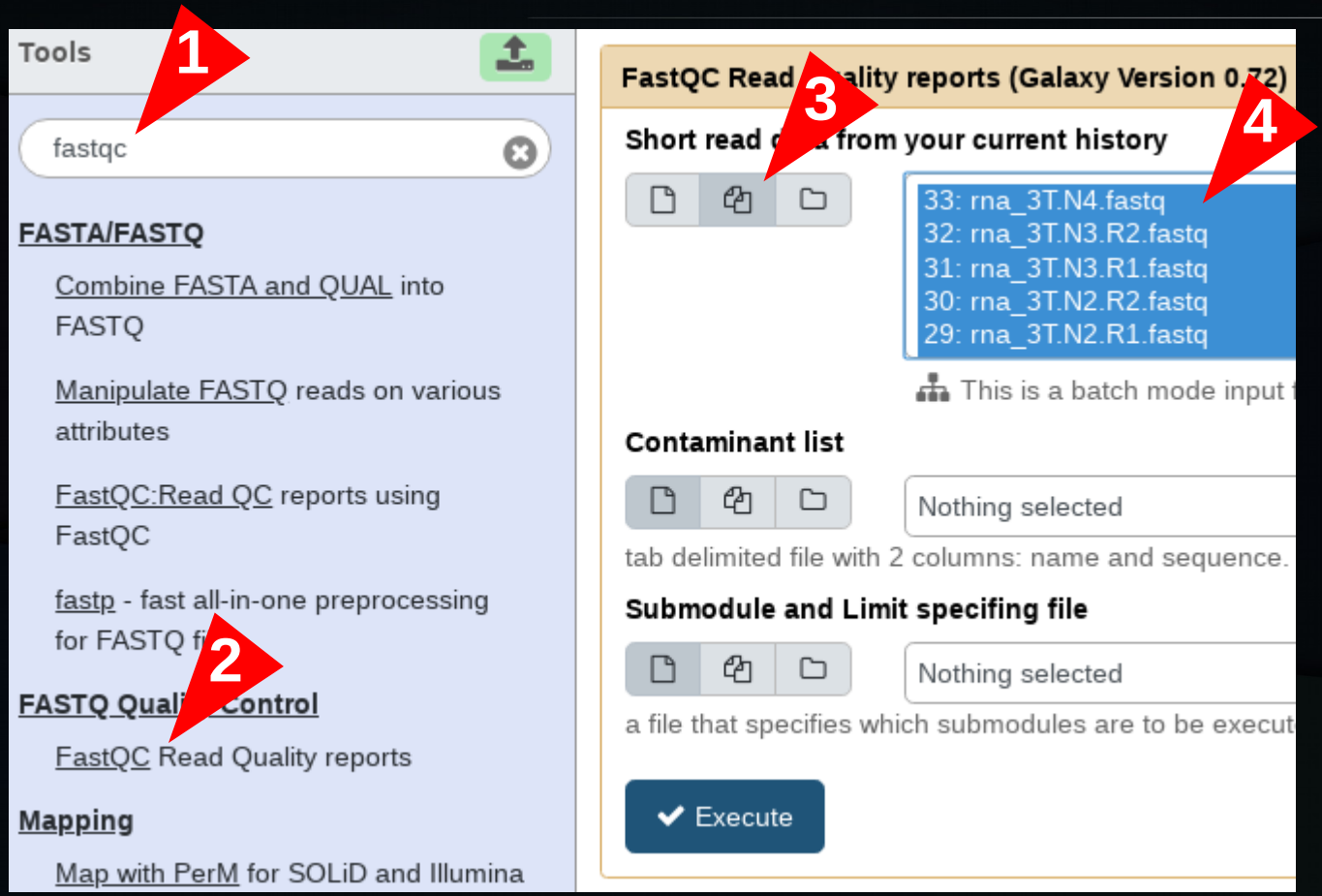
25 DEG with DME

- RNA-Seq
  - 1T & 3T @ N=4
    - 1T paired-end (PE)
    - **3T N1,N4 single (SE)**
    - **3T N2,N3 paired-end (PE)**

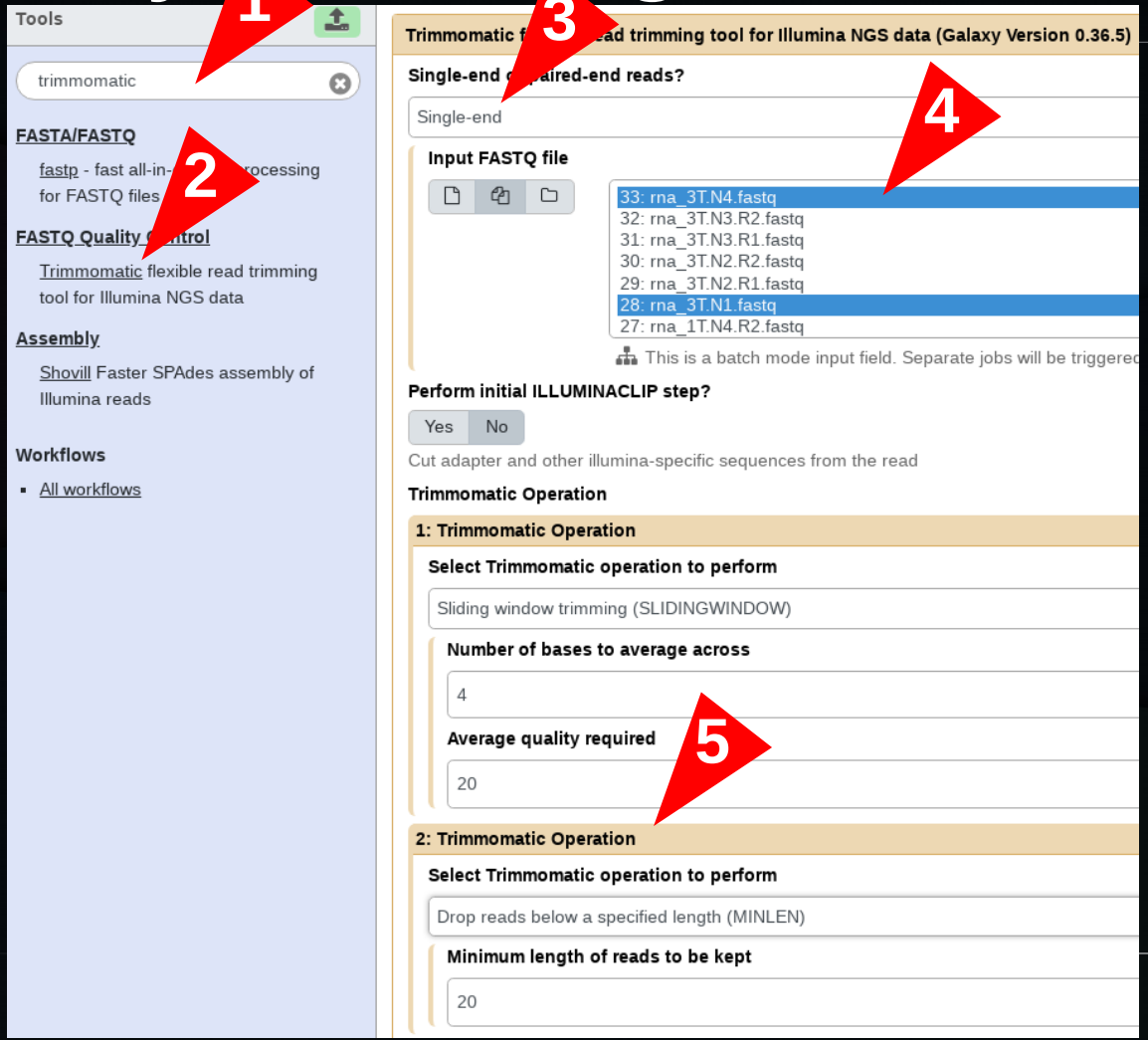
Galaxy – Upload



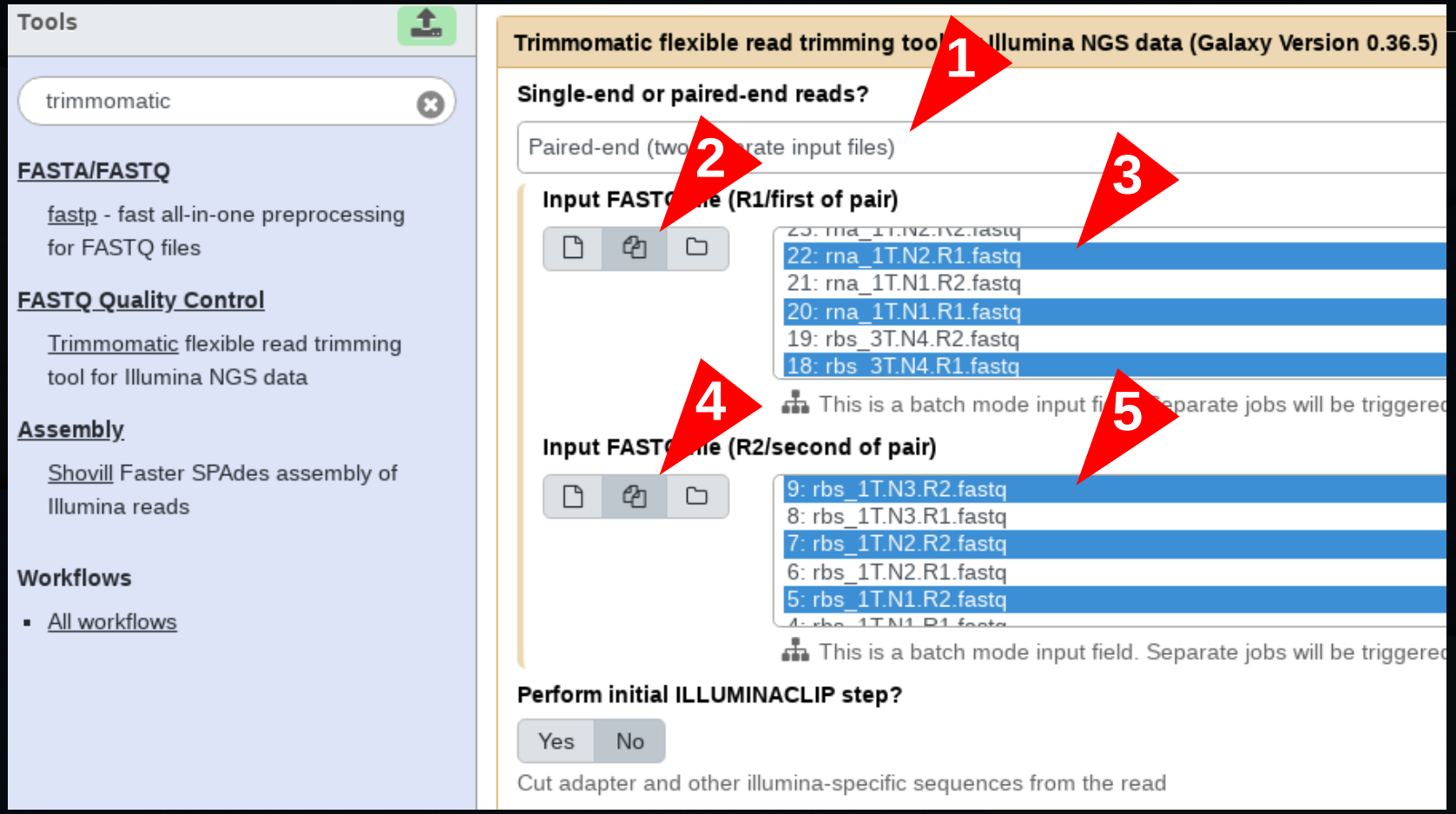
Quality analysis: FastQC



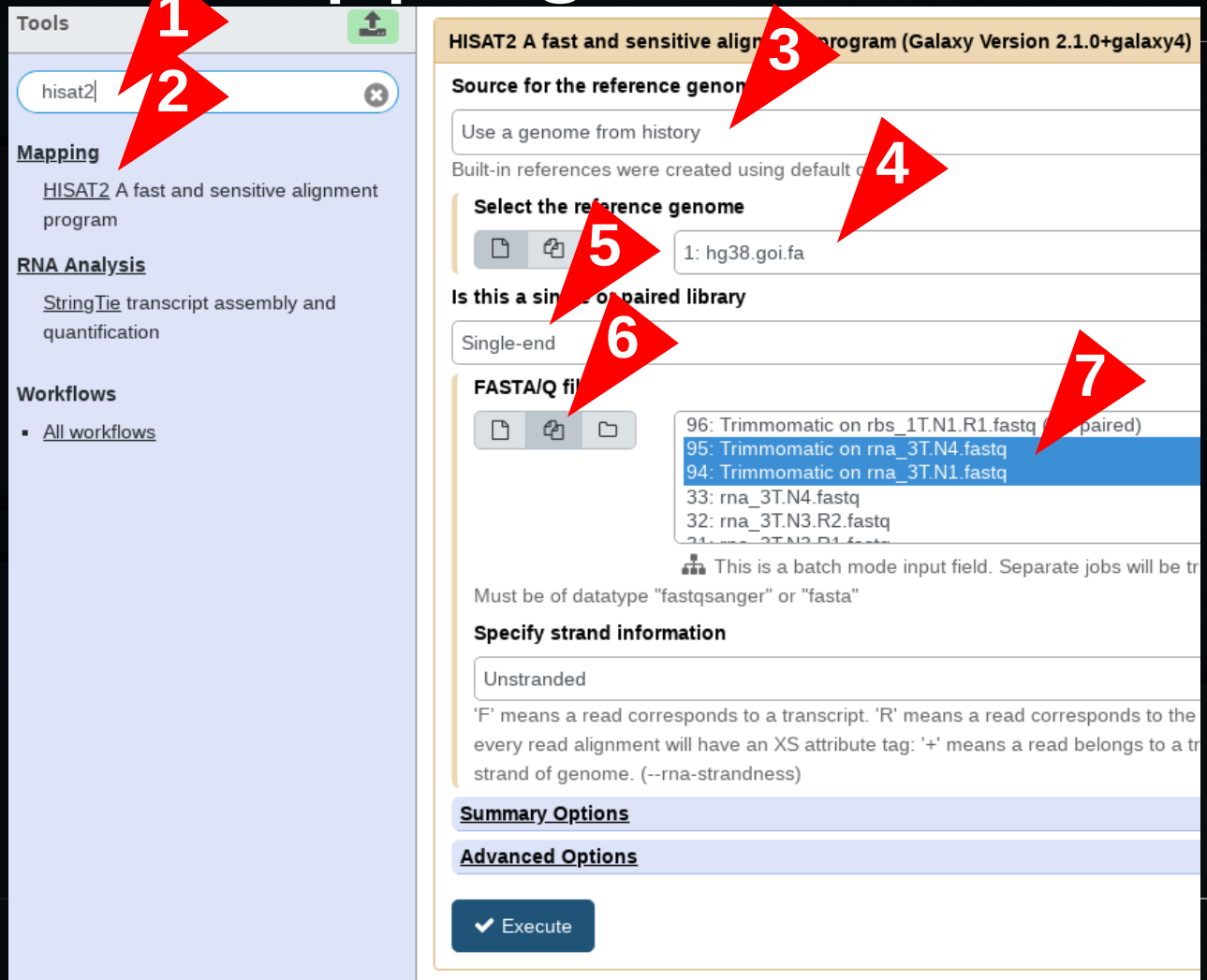
SE – Quality trimming: Trimmomatic



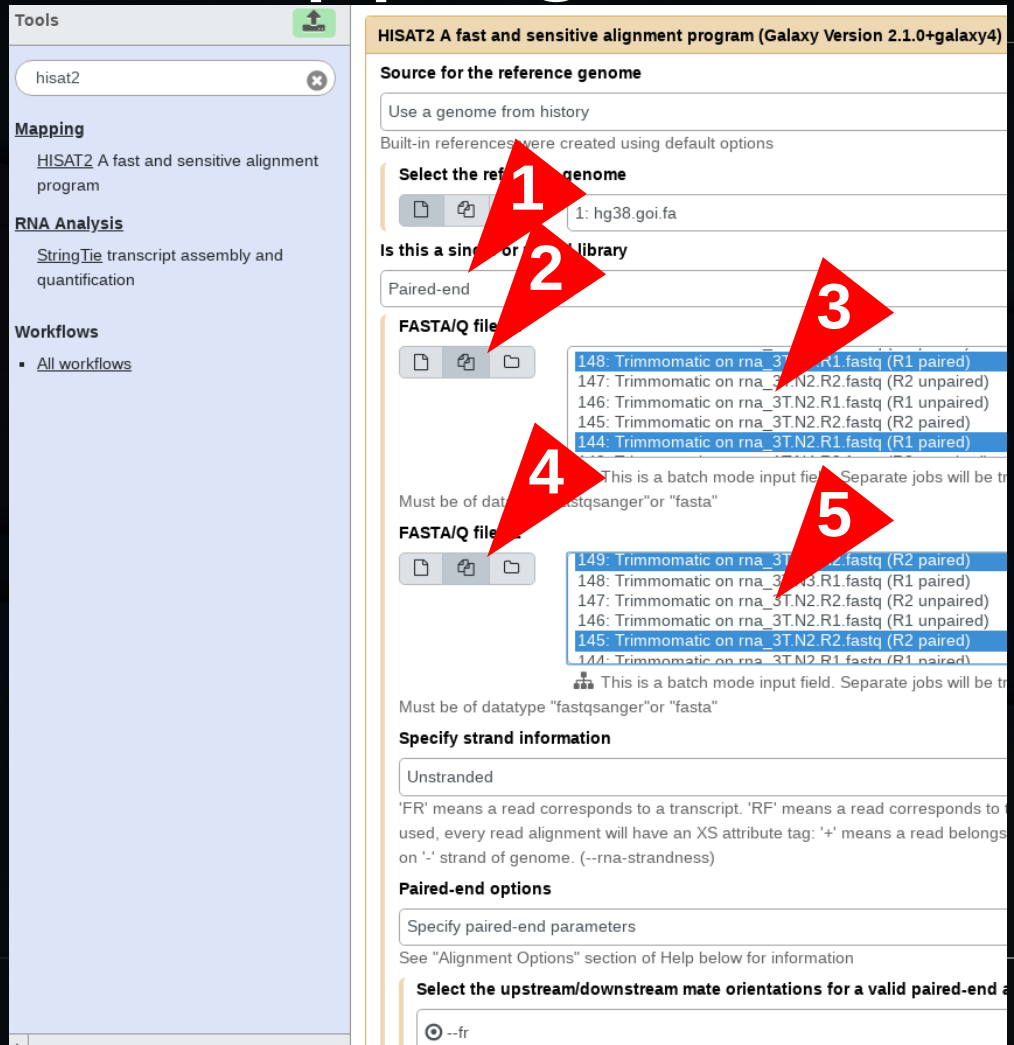
PE – Quality trimming: Trimmomatic



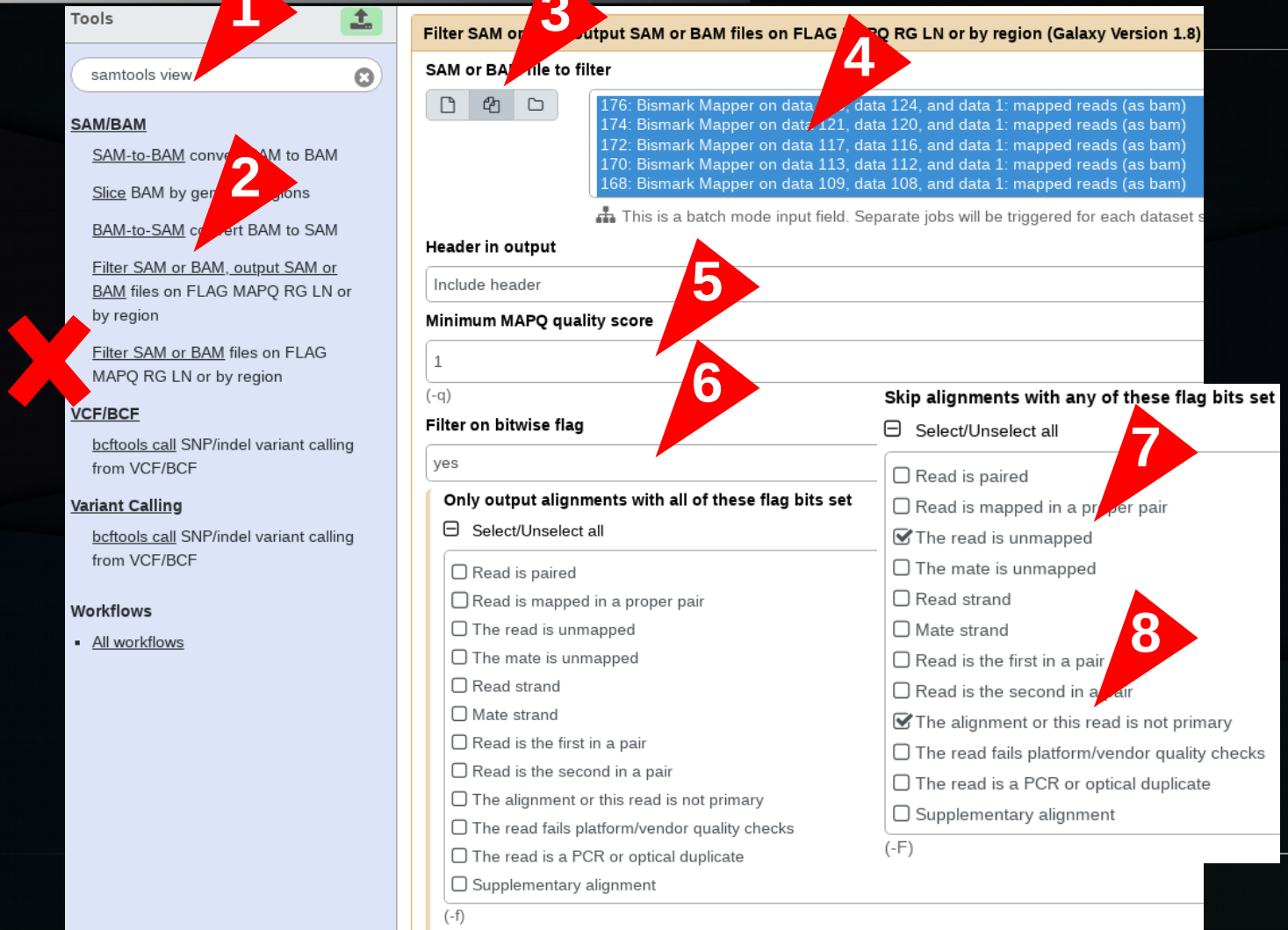
RNA – SE – Mapping: Hisat2



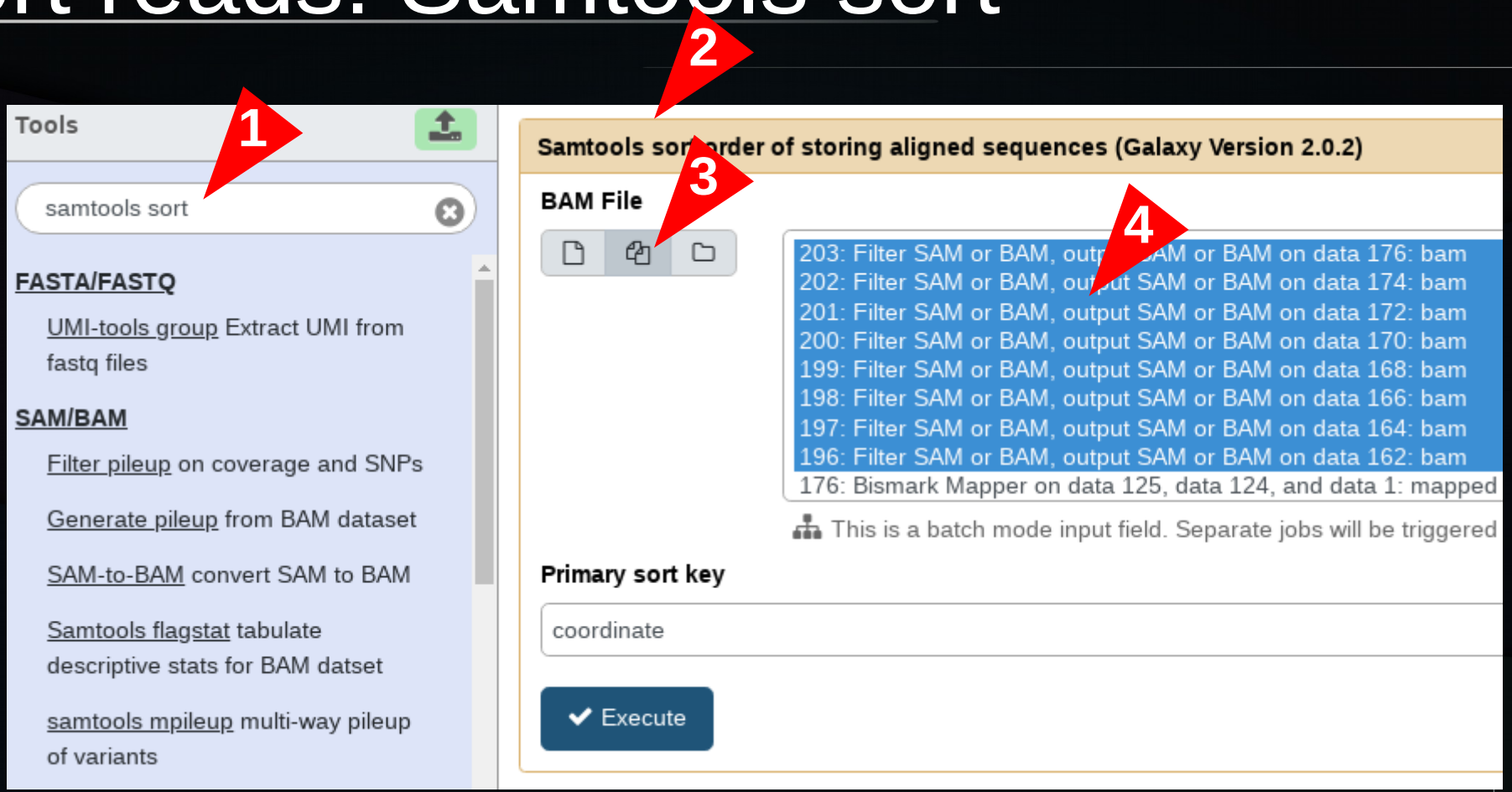
RNA – PE – Mapping: Hisat2



Read filter: Samtools view



Sort reads: Samtools sort

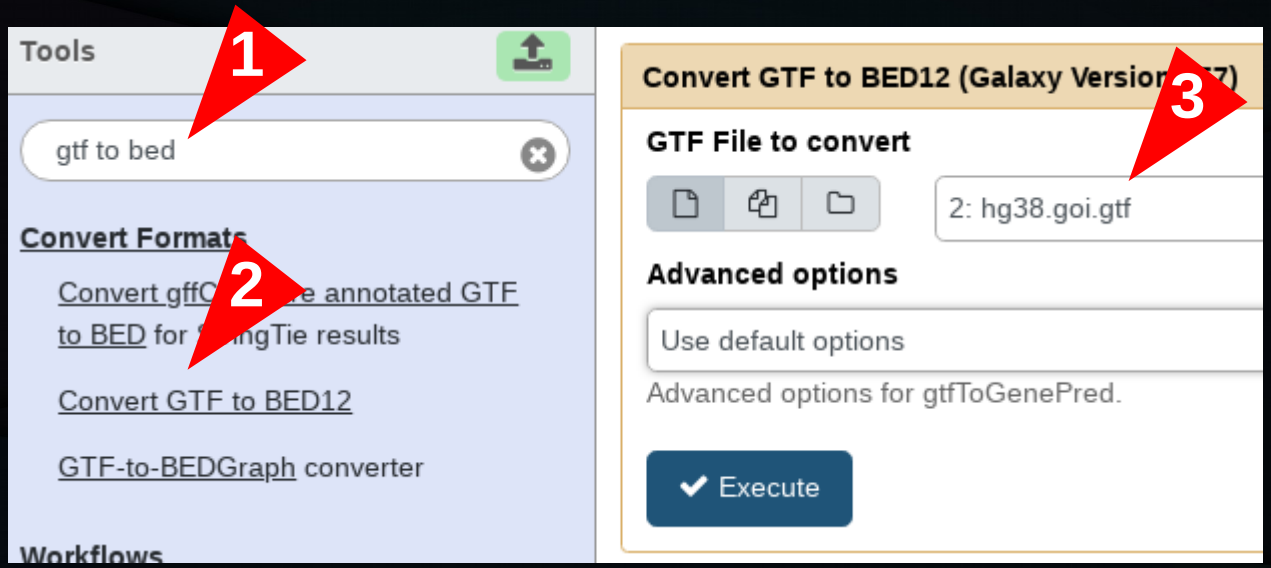


Differential Gene expression

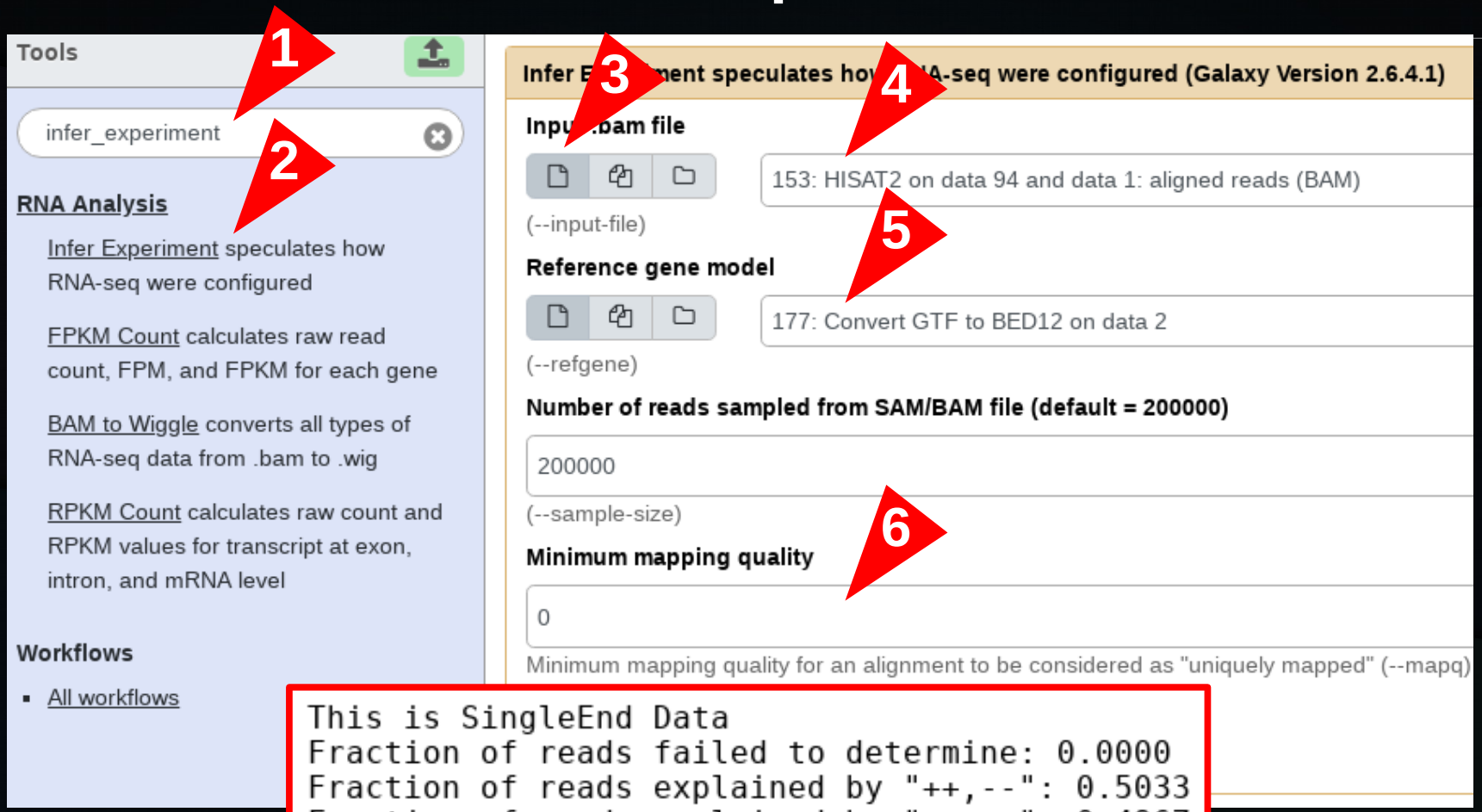
RNA – Quantification

- Requierements
  - Experiment strandness information
    - Prediction needs annotation in BED format

GTF to BED

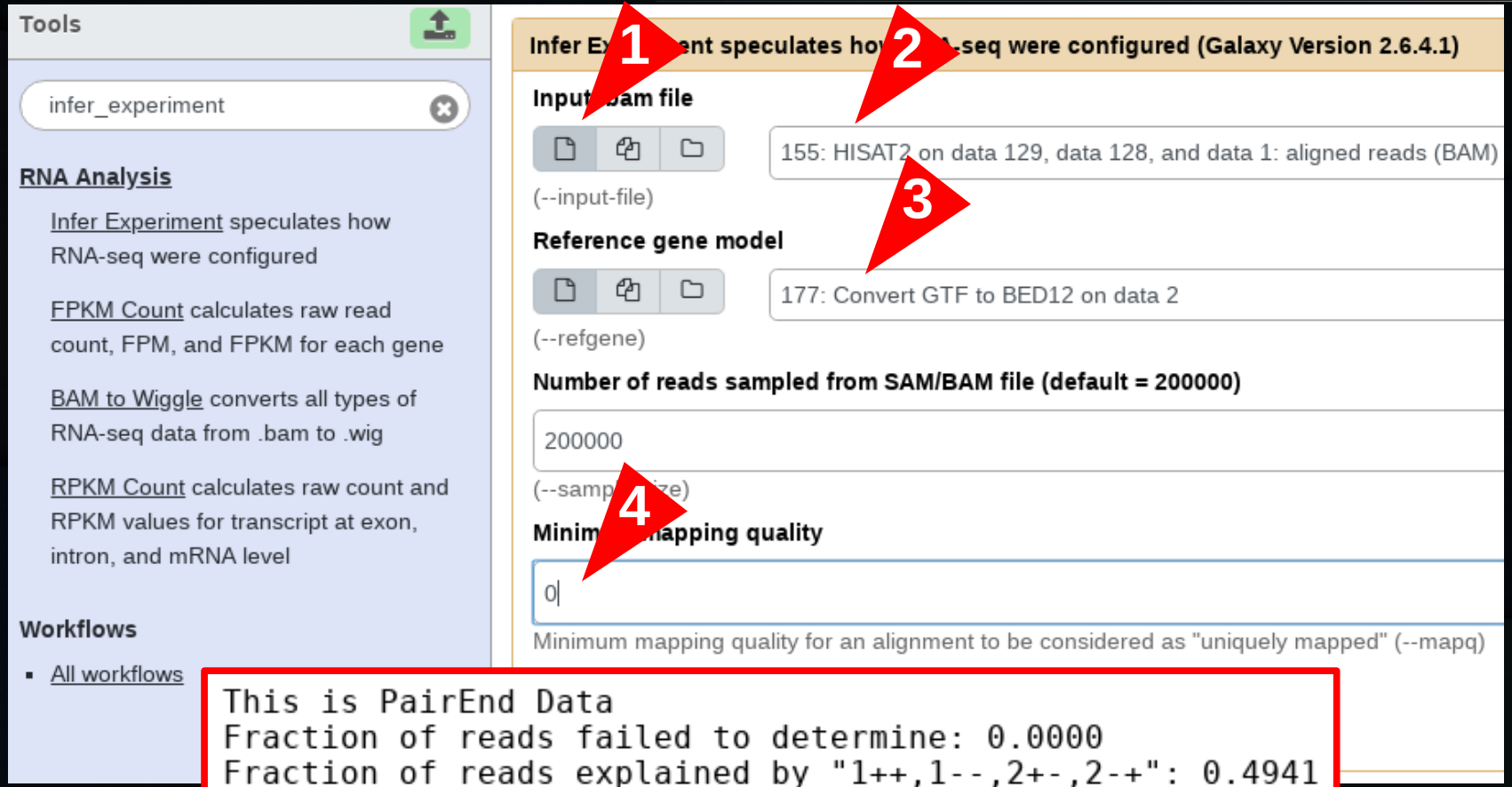

Tools
1
gtf to bed
Convert Formats
Convert gffC e annotated GTF to BED for ingTie results
2
Convert GTF to BED12
GTF-to-BEDGraph converter
Workflows
Convert GTF to BED12 (Galaxy Version 7)
GTF File to convert
2: hg38.goi.gtf
3
Advanced options
Use default options
Advanced options for gtfToGenePred.
Execute

RNA – SE – Infer experiment

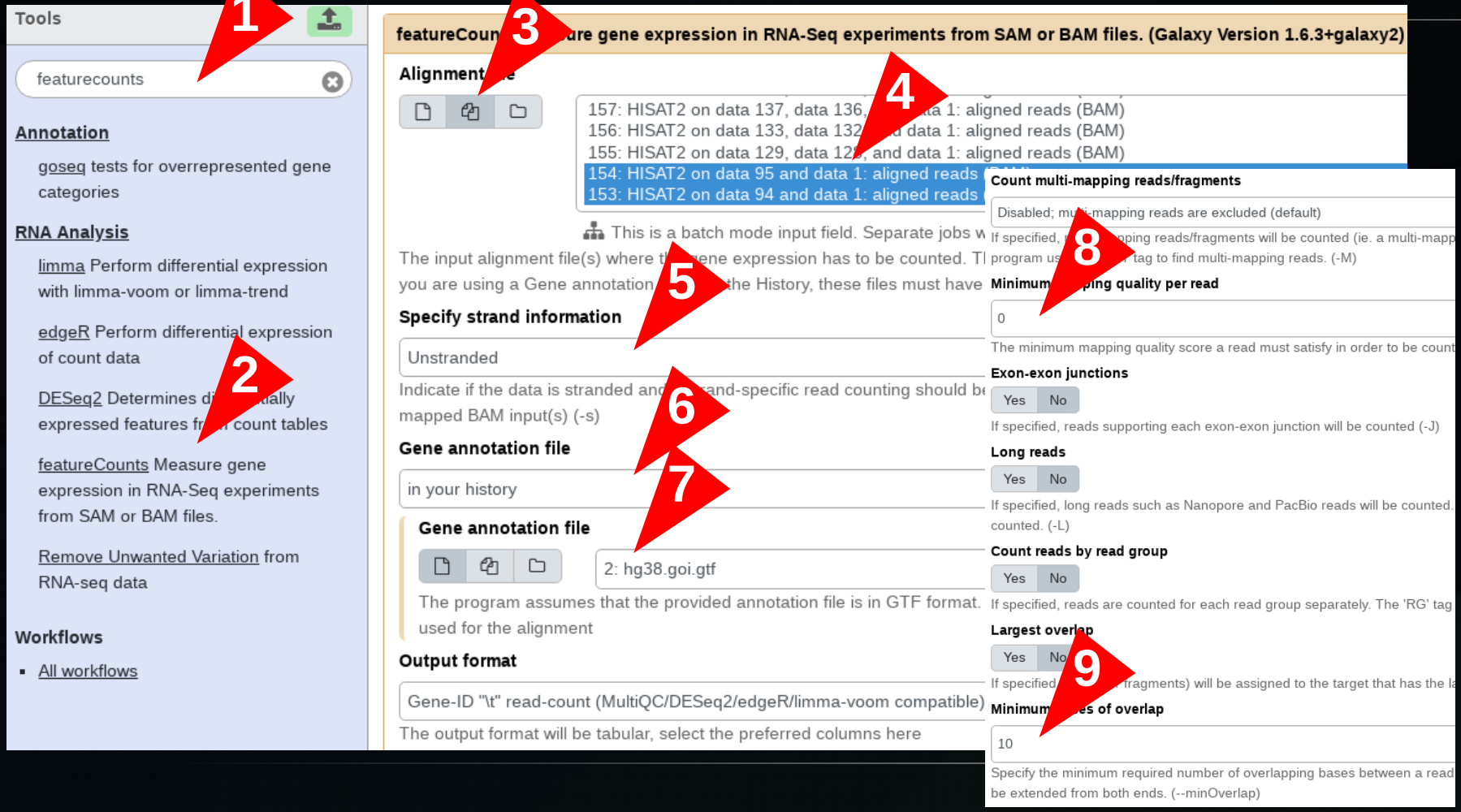



```
This is SingleEnd Data
Fraction of reads failed to determine: 0.0000
Fraction of reads explained by "++,--": 0.5033
Fraction of reads explained by "+-,-+": 0.4967
```

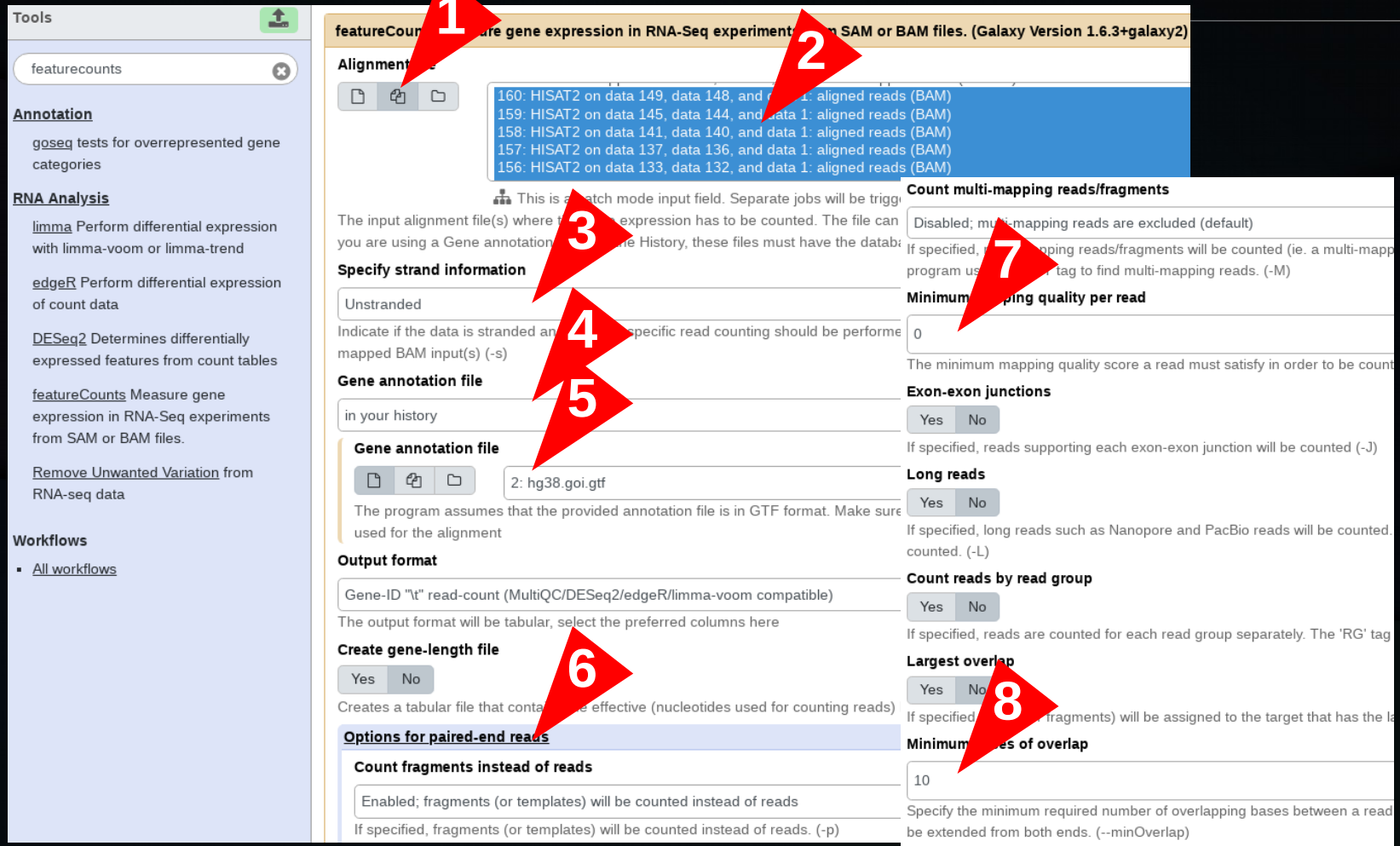
RNA – PE – Infer experiment



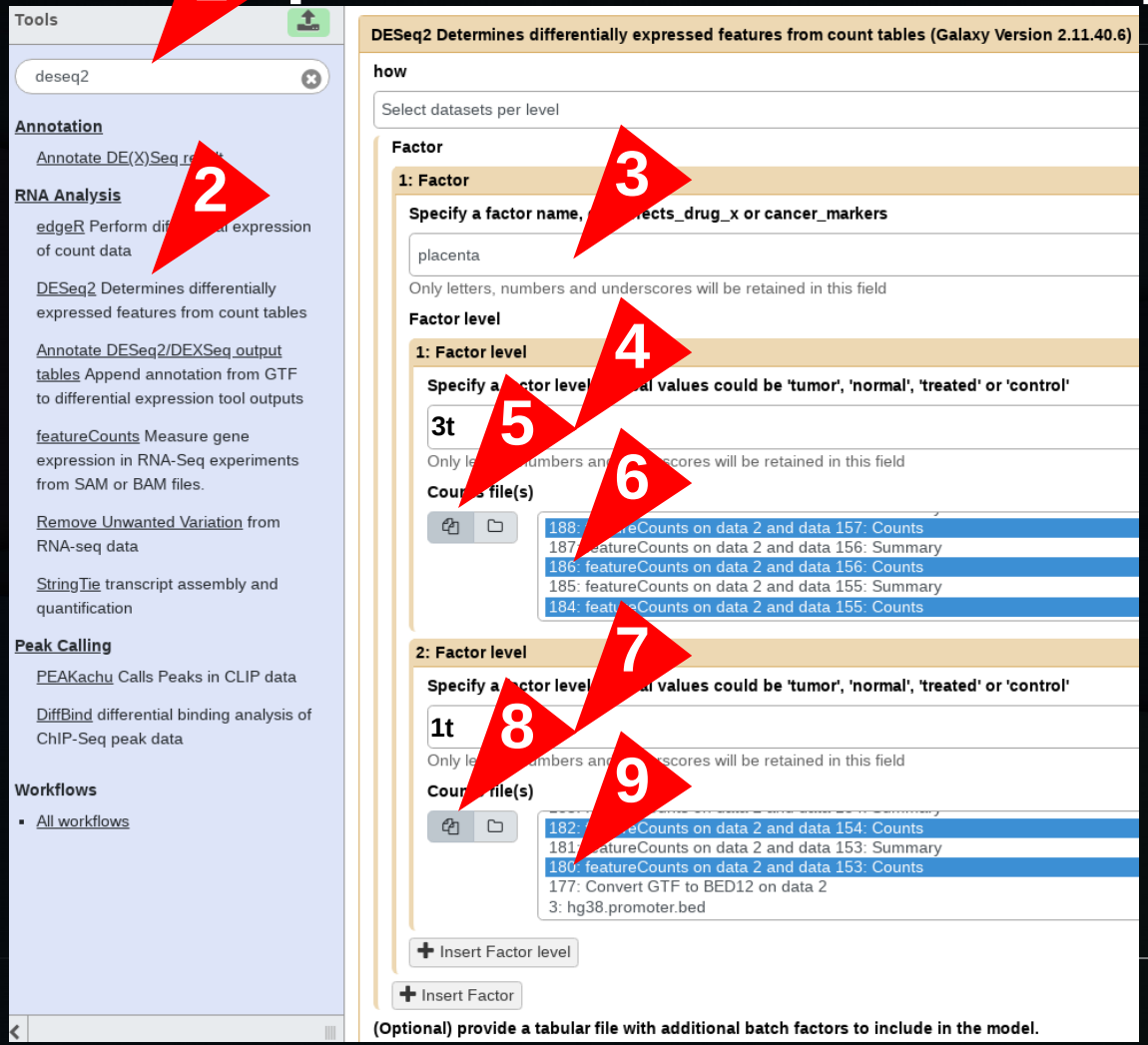
RNA – SE – Quantify: featurecounts



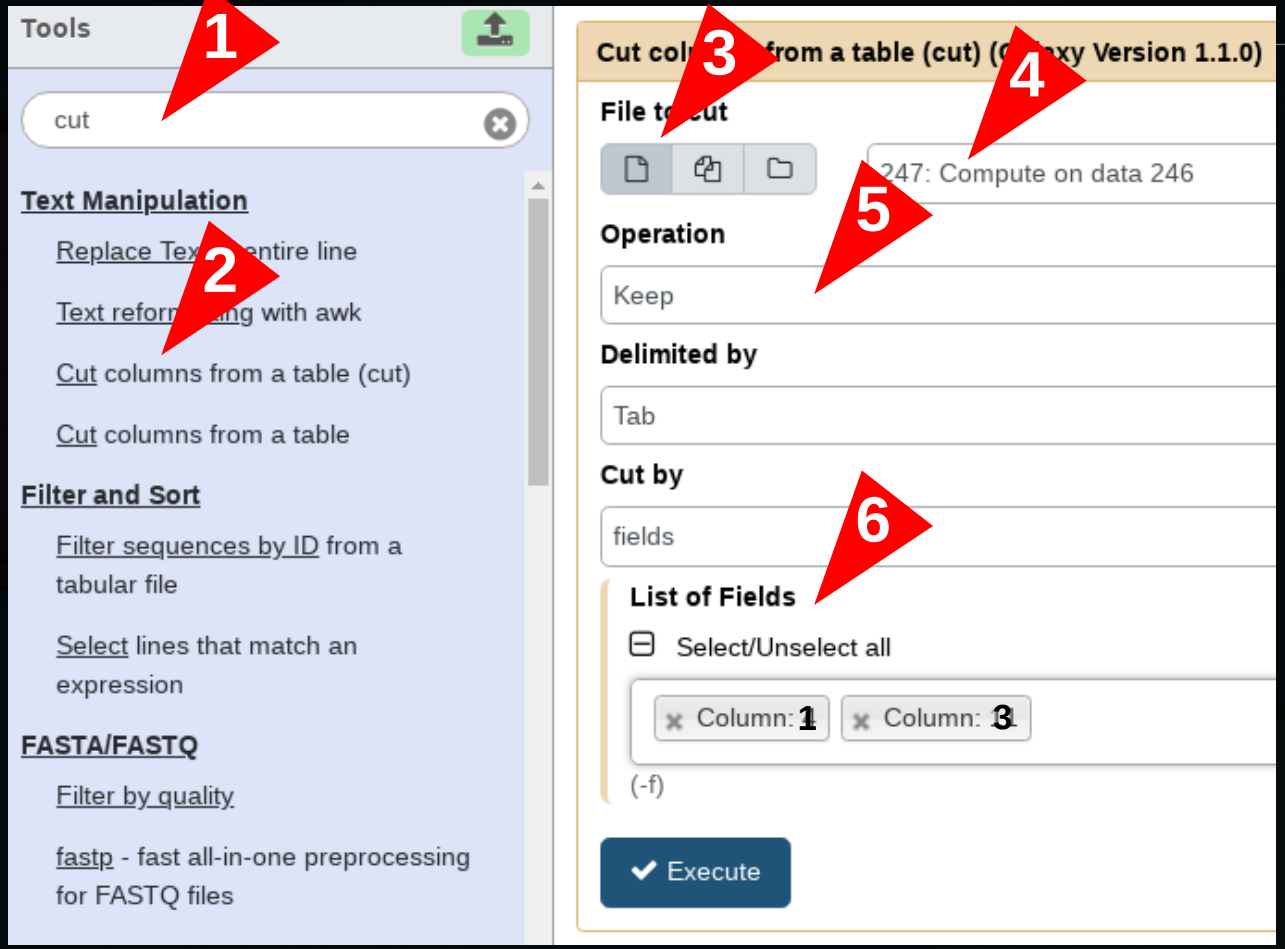
RNA – PE – Quantify: featurecounts



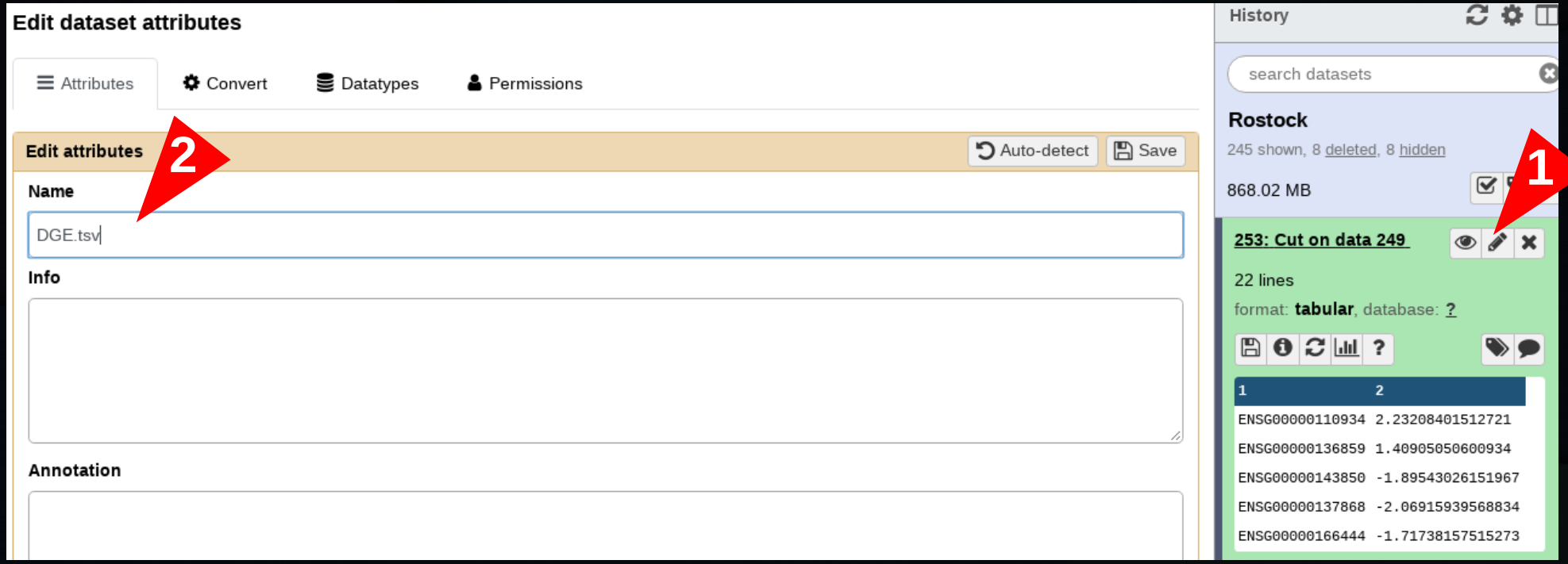
RNA – Diff. expression: DESeq2



Cut columns: cut



Tag results


Edit dataset attributes
Attributes
Convert
Datatypes
Permissions
Edit attributes
Auto-detect
Save
2
Name
DGE.tsv
Info
Annotation
History
search datasets
Rostock
245 shown, 8 deleted, 8 hidden
868.02 MB
1
253: Cut on data 249
22 lines
format: tabular, database: ?
1 2
ENSG00000110934 2.23208401512721
ENSG00000136859 1.40905050600934
ENSG00000143850 -1.89543026151967
ENSG00000137868 -2.06915939568834
ENSG00000166444 -1.71738157515273