NGS+BI+GM HW4

RNA-seq using NTU Galaxy

https://galaxy2.cc.ntu.edu.tw/

01

Data preprocessing

FastQC+Trimmomatic

02

Method-1

Hisat2+StringTie+DESeq2

03

Method-2

Kallisto+DESeq2

04

Analysis

Drawing Venn diagram

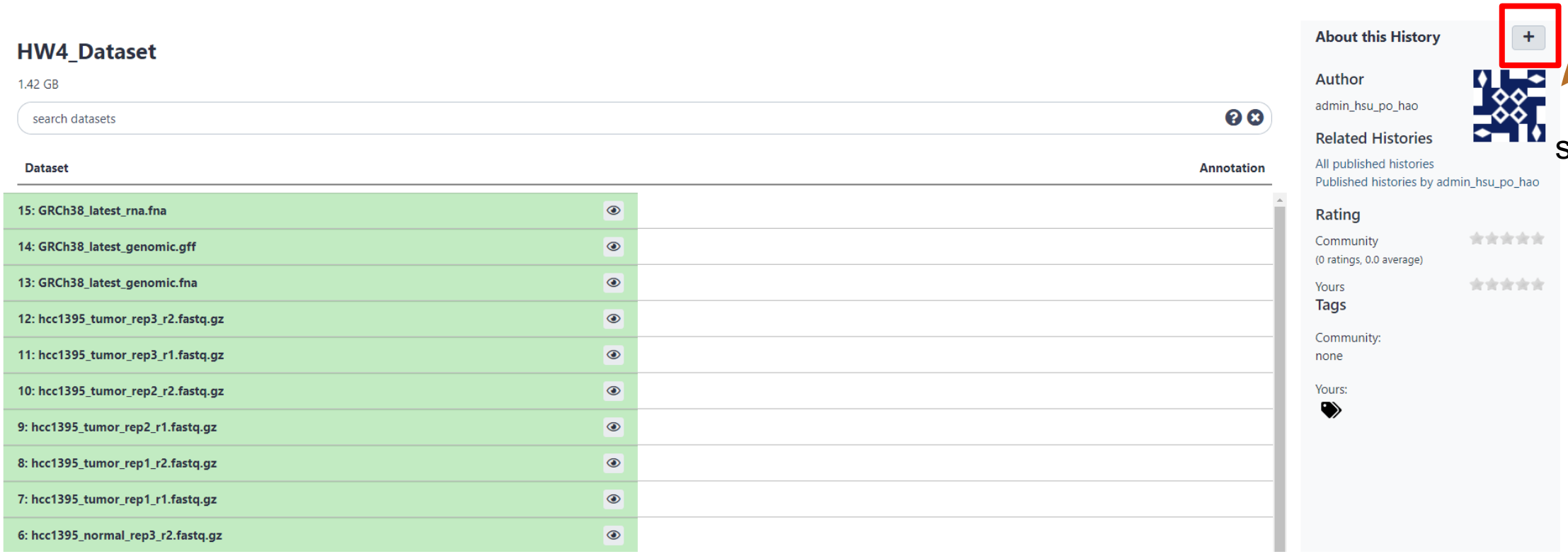
Data preprocessing

Import data + FastQC + Trimmomatic

Import data

Loading data from following URL : (paste URL to your website and login NTU galaxy

https://galaxy2.cc.ntu.edu.tw/u/galaxy_test/h/hw4dataset

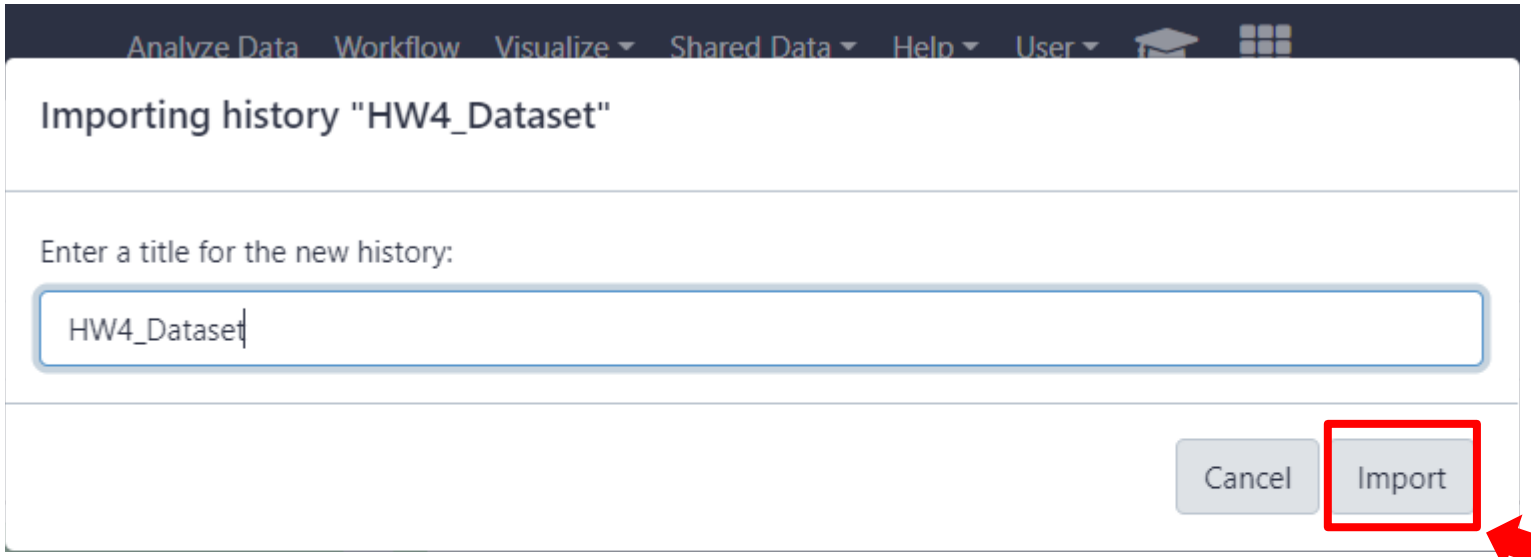


step1

Datasets were downloaded from https://github.com/griffithlab/rnaseq_tutorial/wiki/RNAseq-Data

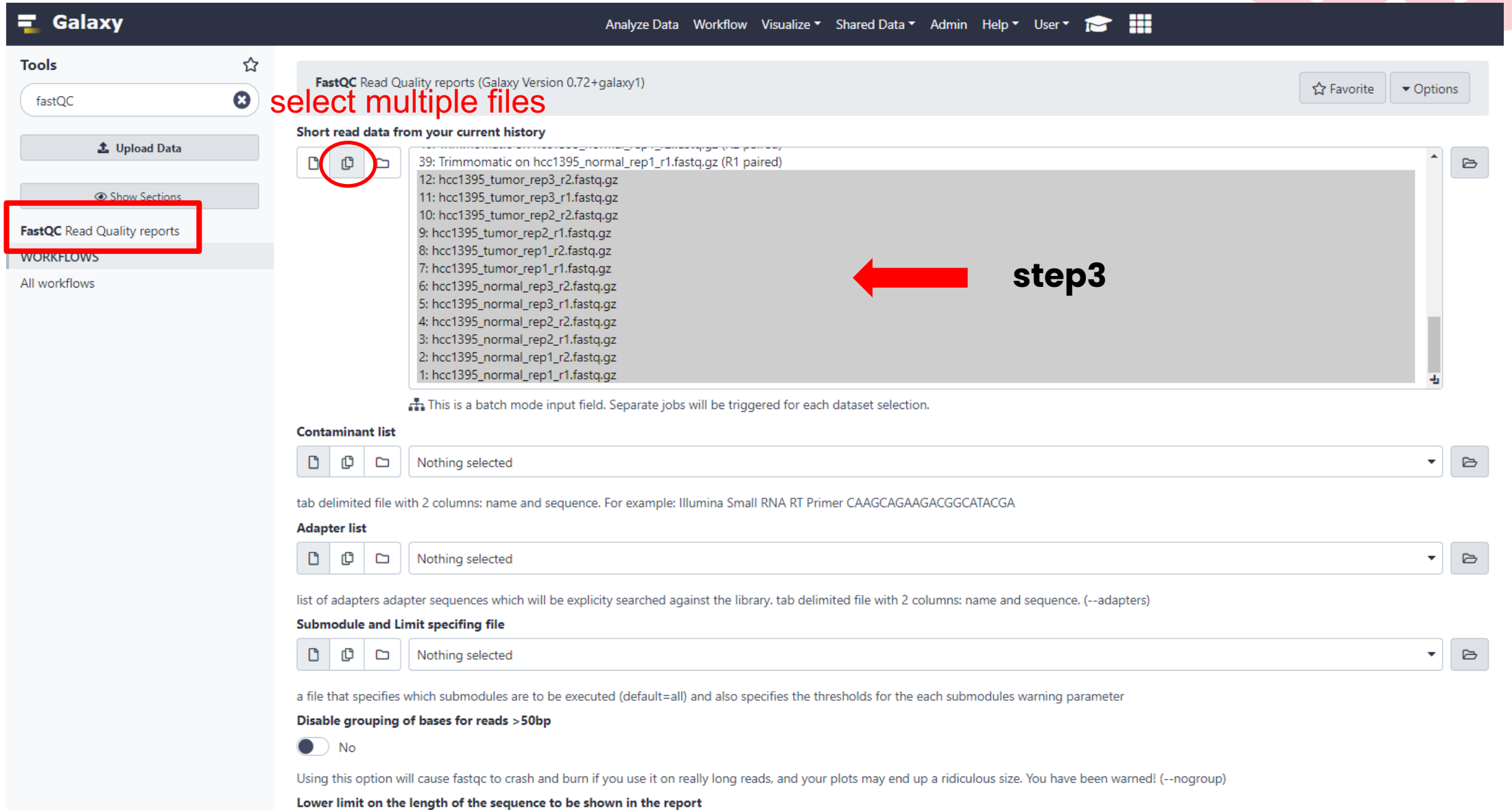
Import data

Import to your own history



step2

FastQC


Galaxy
Analyze Data
Workflow
Visualize
Shared Data
Admin
Help
User
Tools
fastQC
Upload Data
Show Sections
FastQC Read Quality reports
WORKFLOWS
All workflows
FastQC Read Quality reports (Galaxy Version 0.72+galaxy1)
Favorite
Options
select multiple files
Short read data from your current history
39: Trimmomatic on hcc1395_normal_rep1_r1.fastq.gz (R1 paired)
12: hcc1395_tumor_rep3_r2.fastq.gz
11: hcc1395_tumor_rep3_r1.fastq.gz
10: hcc1395_tumor_rep2_r2.fastq.gz
9: hcc1395_tumor_rep2_r1.fastq.gz
8: hcc1395_tumor_rep1_r2.fastq.gz
7: hcc1395_tumor_rep1_r1.fastq.gz
6: hcc1395_normal_rep3_r2.fastq.gz
5: hcc1395_normal_rep3_r1.fastq.gz
4: hcc1395_normal_rep2_r2.fastq.gz
3: hcc1395_normal_rep2_r1.fastq.gz
2: hcc1395_normal_rep1_r2.fastq.gz
1: hcc1395_normal_rep1_r1.fastq.gz
step3
This is a batch mode input field. Separate jobs will be triggered for each dataset selection.
Contaminant list
Nothing selected
tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA
Adapter list
Nothing selected
list of adapters adapter sequences which will be explicity searched against the library. tab delimited file with 2 columns: name and sequence. (--adapters)
Submodule and Limit specifing file
Nothing selected
a file that specifies which submodules are to be executed (default=all) and also specifies the thresholds for the each submodules warning parameter
Disable grouping of bases for reads >50bp
No
Using this option will cause fastqc to crash and burn if you use it on really long reads, and your plots may end up a ridiculous size. You have been warned! (--nogroup)
Lower limit on the length of the sequence to be shown in the report

FastQC

Executed **FastQC** and successfully added 12 jobs to the queue.

The tool uses 12 inputs:

- **1: hcc1395_normal_rep1_r1.fastq.gz**
- **2: hcc1395_normal_rep1_r2.fastq.gz**
- **3: hcc1395_normal_rep2_r1.fastq.gz**
- **4: hcc1395_normal_rep2_r2.fastq.gz**
- **5: hcc1395_normal_rep3_r1.fastq.gz**
- **6: hcc1395_normal_rep3_r2.fastq.gz**
- **7: hcc1395_tumor_rep1_r1.fastq.gz**
- **8: hcc1395_tumor_rep1_r2.fastq.gz**
- **9: hcc1395_tumor_rep2_r1.fastq.gz**
- **10: hcc1395_tumor_rep2_r2.fastq.gz**
- **11: hcc1395_tumor_rep3_r1.fastq.gz**
- **12: hcc1395_tumor_rep3_r2.fastq.gz**

It produces 24 outputs:

- **15: FastQC on data 1: Webpage**
- **16: FastQC on data 1: RawData**
- **17: FastQC on data 2: Webpage**
- **18: FastQC on data 2: RawData**
- **19: FastQC on data 3: Webpage**
- **20: FastQC on data 3: RawData**
- **21: FastQC on data 4: Webpage**
- **22: FastQC on data 4: RawData**
- **23: FastQC on data 5: Webpage**
- **24: FastQC on data 5: RawData**
- **25: FastQC on data 6: Webpage**
- **26: FastQC on data 6: RawData**

QC report

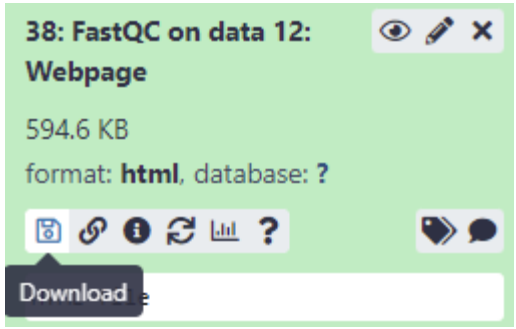
FastQC Report

Summary

- Basic Statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content

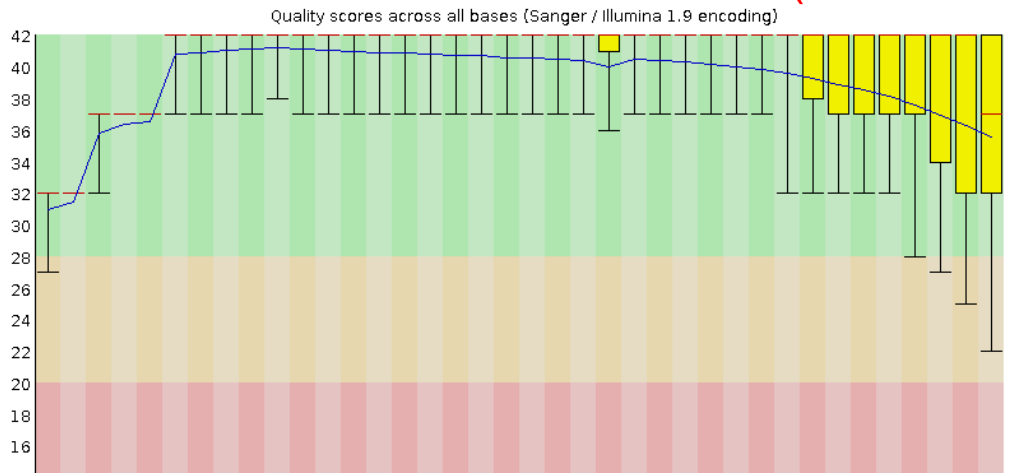
Basic Statistics

Measure	Value
Filename	hcc1395_normal_rep1_r1_fastq_gz.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	331958
Sequences flagged as poor quality	0
Sequence length	151
%GC	54



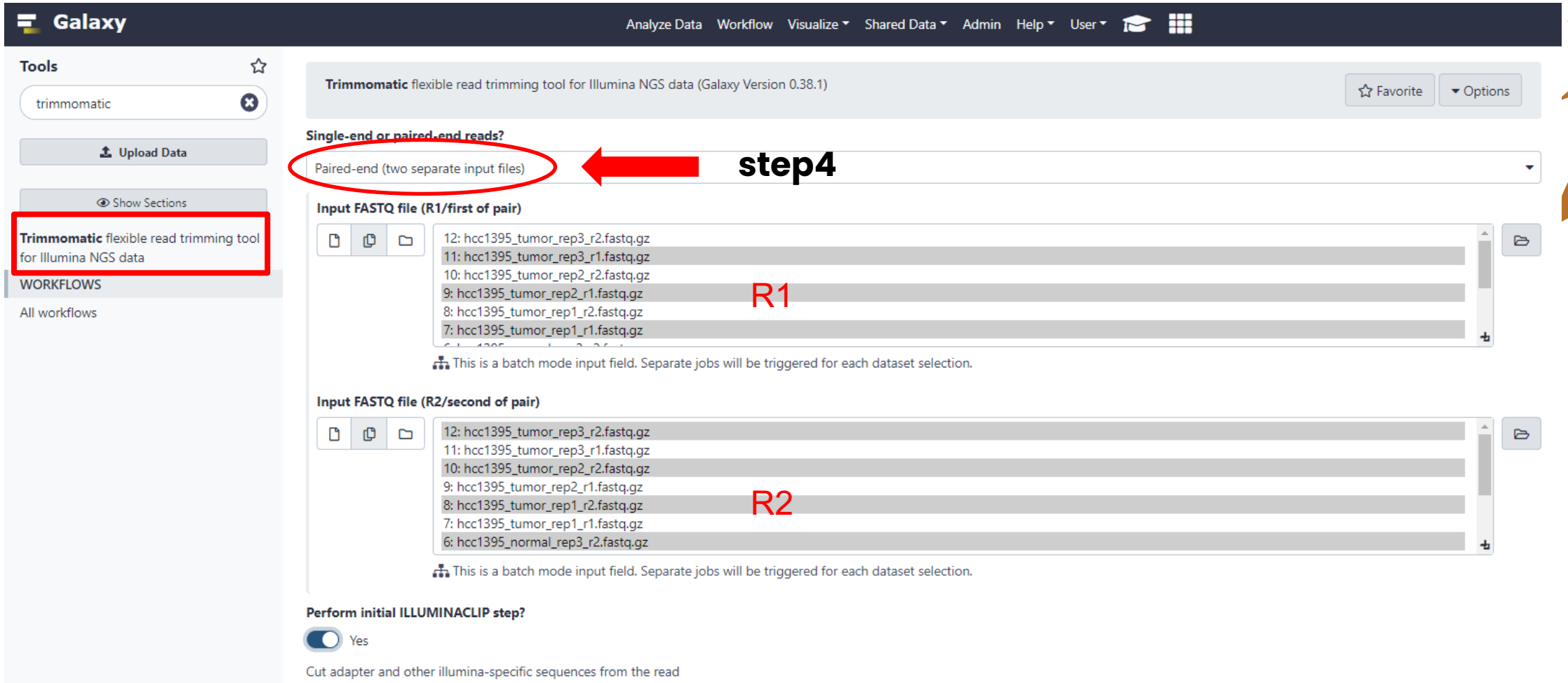
Downloading the html results to your computer
to see the complete result
(there are some features can be found in RawData)

Per base sequence quality



Trimmomatic

select multiple files



Trimmomatic

Perform initial ILLUMINACLIP step?

Yes

step5

Cut adapter and other illumina-specific sequences from the read

Select standard adapter sequences or provide custom?

Standard

Adapter sequences to use

TruSeq3 (paired-ended, for MiSeq and HiSeq)

Maximum mismatch count which will still allow a full match to be performed

2

How accurate the match between the two 'adapter ligated' reads must be for PE palindrome read alignment

30

How accurate the match between any adapter etc. sequence must be against a read

10

Minimum length of adapter that needs to be detected (PE specific/palindrome mode)

2

Trimmomatic

Always keep both reads (PE specific/palindrome mode)?

Yes

See help below

Trimmomatic Operation

1: Trimmomatic Operation

Select Trimmomatic operation to perform

Cut bases off the start of a read, if below a threshold quality (LEADING)

Minimum quality required to keep a base

3

Bases at the start of the read with quality below the threshold will be removed

2: Trimmomatic Operation

Select Trimmomatic operation to perform

Cut bases off the end of a read, if below a threshold quality (TRAILING)

Minimum quality required to keep a base

3

Bases at the end of the read with quality below the threshold will be removed

3: Trimmomatic Operation

Select Trimmomatic operation to perform

Drop reads with average quality lower than a specified level (AVGQUAL)

Minimum average quality required to keep a read

36

Trimmomatic

Executed **Trimmomatic** and successfully added 6 jobs to the queue.

The tool uses 12 inputs:

- **1: hcc1395_normal_rep1_r1.fastq.gz**
- **3: hcc1395_normal_rep2_r1.fastq.gz**
- **5: hcc1395_normal_rep3_r1.fastq.gz**
- **7: hcc1395_tumor_rep1_r1.fastq.gz**
- **9: hcc1395_tumor_rep2_r1.fastq.gz**
- **11: hcc1395_tumor_rep3_r1.fastq.gz**
- **2: hcc1395_normal_rep1_r2.fastq.gz**
- **4: hcc1395_normal_rep2_r2.fastq.gz**
- **6: hcc1395_normal_rep3_r2.fastq.gz**
- **8: hcc1395_tumor_rep1_r2.fastq.gz**
- **10: hcc1395_tumor_rep2_r2.fastq.gz**
- **12: hcc1395_tumor_rep3_r2.fastq.gz**

It produces 24 outputs:

- **39: Trimmomatic on hcc1395_normal_rep1_r1.fastq.gz (R1 paired)**
- **40: Trimmomatic on hcc1395_normal_rep1_r2.fastq.gz (R2 paired)**
- **41: Trimmomatic on hcc1395_normal_rep1_r1.fastq.gz (R1 unpaired)**
- **42: Trimmomatic on hcc1395_normal_rep1_r2.fastq.gz (R2 unpaired)**
- **43: Trimmomatic on hcc1395_normal_rep2_r1.fastq.gz (R1 paired)**
- **44: Trimmomatic on hcc1395_normal_rep2_r2.fastq.gz (R2 paired)**
- **45: Trimmomatic on hcc1395_normal_rep2_r1.fastq.gz (R1 unpaired)**
- **46: Trimmomatic on hcc1395_normal_rep2_r2.fastq.gz (R2 unpaired)**
- **47: Trimmomatic on hcc1395_normal_rep3_r1.fastq.gz (R1 paired)**
- **48: Trimmomatic on hcc1395_normal_rep3_r2.fastq.gz (R2 paired)**
- **49: Trimmomatic on hcc1395_normal_rep3_r1.fastq.gz (R1 unpaired)**
- **50: Trimmomatic on hcc1395_normal_rep3_r2.fastq.gz (R2 unpaired)**
- **51: Trimmomatic on hcc1395_tumor_rep1_r1.fastq.gz (R1 paired)**
- **52: Trimmomatic on hcc1395_tumor_rep1_r2.fastq.gz (R2 paired)**
- **53: Trimmomatic on hcc1395_tumor_rep1_r1.fastq.gz (R1 unpaired)**

Note :
The Trimmomatic result would be used in following chapter: method1 and method2 directly.

Re-run FastQC

select multiple files

Tools

fastQC

Upload Data

Show Sections

FastQC Read Quality reports

WORKFLOWS

All workflows

step9

FastQC Read Quality reports (Galaxy Version 0.72+galaxy1)

Favorite | Options

Short read data from your current history

62: Trimmomatic on hcc1395_tumor_rep3_r2.fastq.gz (R2 unpaired)
61: Trimmomatic on hcc1395_tumor_rep3_r1.fastq.gz (R1 unpaired)
60: Trimmomatic on hcc1395_tumor_rep3_r2.fastq.gz (R2 paired)
59: Trimmomatic on hcc1395_tumor_rep3_r1.fastq.gz (R1 paired)
58: Trimmomatic on hcc1395_tumor_rep2_r2.fastq.gz (R2 unpaired)
57: Trimmomatic on hcc1395_tumor_rep2_r1.fastq.gz (R1 unpaired)
56: Trimmomatic on hcc1395_tumor_rep2_r2.fastq.gz (R2 paired)
55: Trimmomatic on hcc1395_tumor_rep2_r1.fastq.gz (R1 paired)
54: Trimmomatic on hcc1395_tumor_rep1_r2.fastq.gz (R2 unpaired)
53: Trimmomatic on hcc1395_tumor_rep1_r1.fastq.gz (R1 unpaired)
52: Trimmomatic on hcc1395_tumor_rep1_r2.fastq.gz (R2 paired)
51: Trimmomatic on hcc1395_tumor_rep1_r1.fastq.gz (R1 paired)
50: Trimmomatic on hcc1395_normal_rep3_r2.fastq.gz (R2 unpaired)
49: Trimmomatic on hcc1395_normal_rep3_r1.fastq.gz (R1 unpaired)

This is a batch mode input field. Separate jobs will be triggered for each dataset selection.

Contaminant list

Nothing selected

tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA

Adapter list

Nothing selected

list of adapters adapter sequences which will be explicity searched against the library. tab delimited file with 2 columns: name and sequence. (--adapters)

Submodule and Limit specifing file

Nothing selected

Check QC report after Trimmomatic

FastQC Report

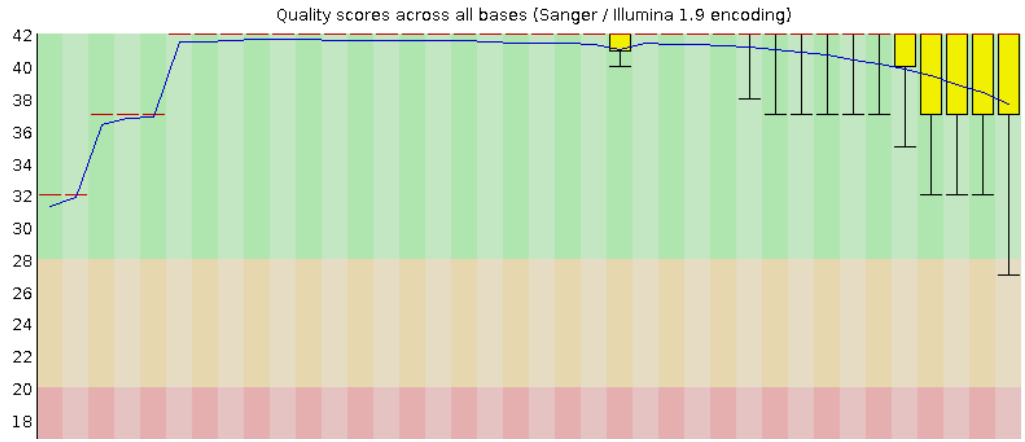
Summary

- Basic Statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content

Basic Statistics

Measure	Value
Filename	Trimmomatic on hcc1395_normal_rep1_r1_fastq_gz _R1 paired_.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	222746
Sequences flagged as poor quality	0
Sequence length	116-151
%GC	54

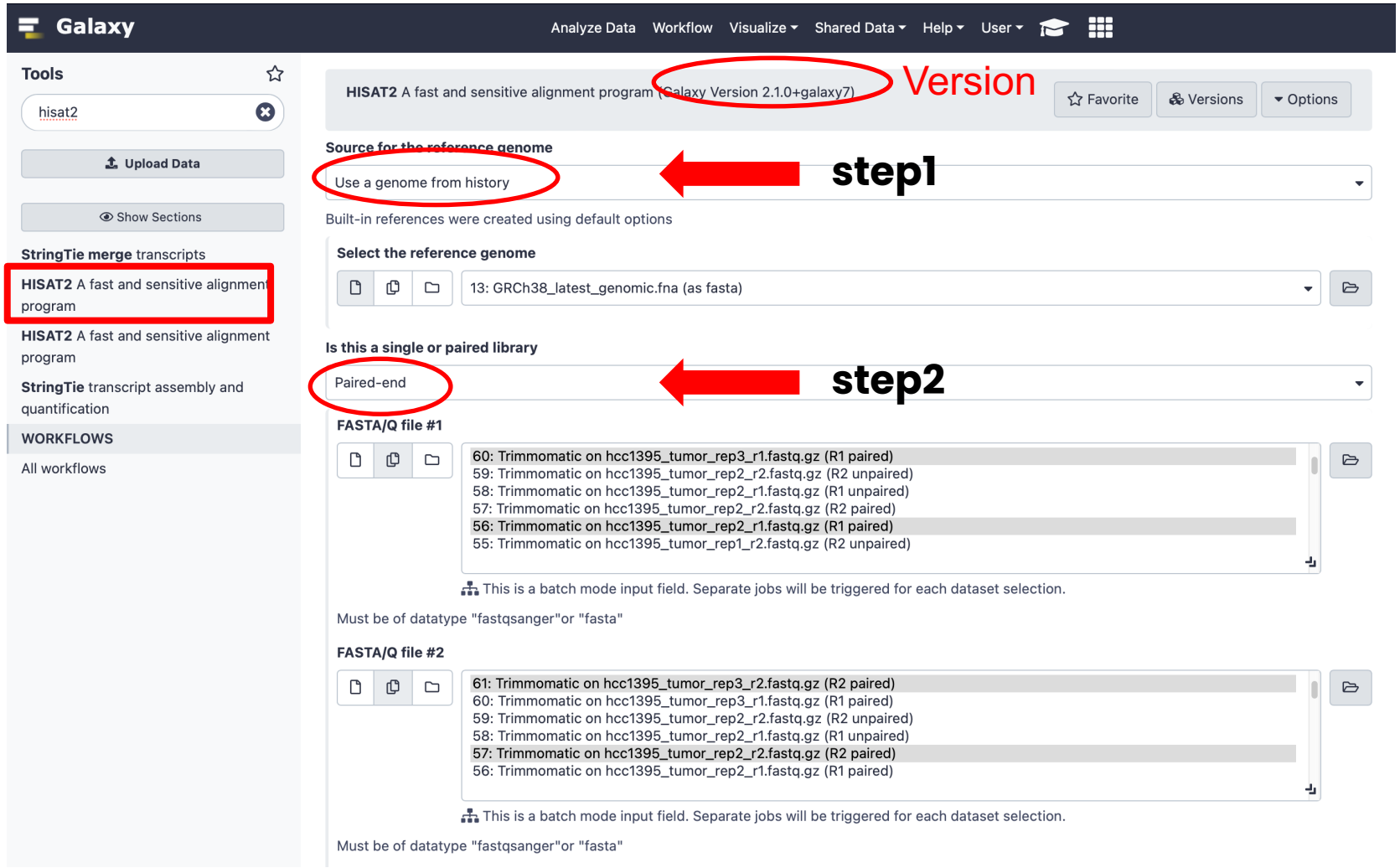
Per base sequence quality



Method-1

Hisat2 + StringTie + DESeq2

Hisat2



Strand Settings: https://rnabio.org/module-09-appendix/0009/12/01/StrandSettings/

Hisat2

The tasks could take longer than 20 minutes.

Executed **HISAT2** and successfully added 6 jobs to the queue.

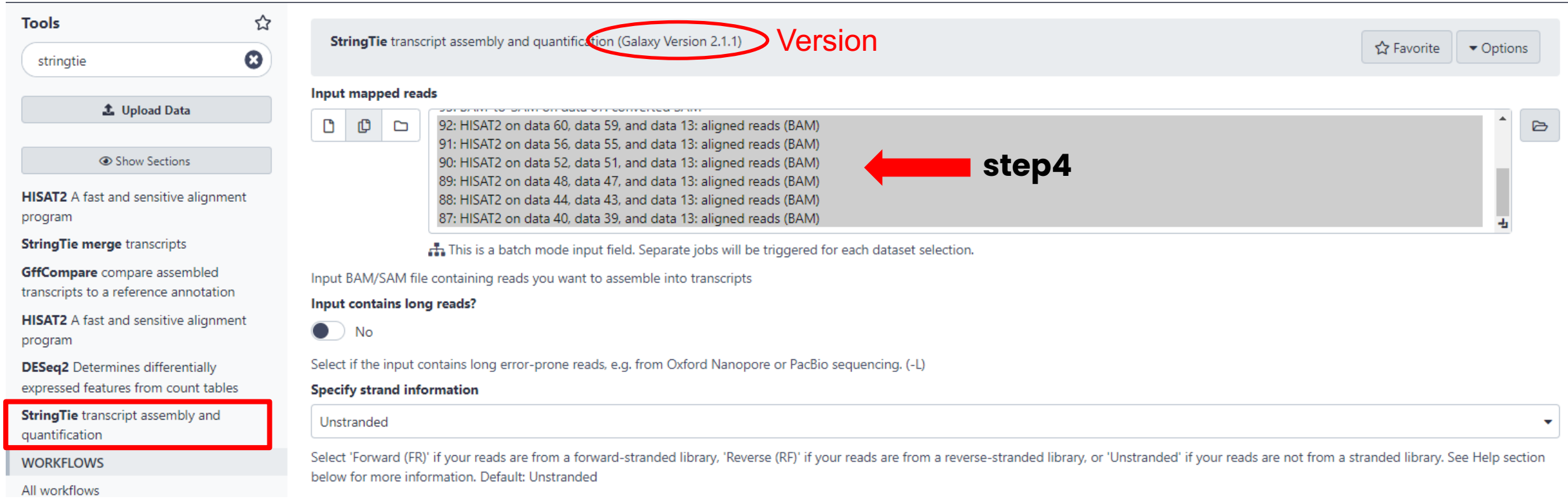
The tool uses 13 inputs:

- **13: hg38_refseq (as fasta)**
- **39: Trimmomatic on hcc1395_normal_rep1_r1.fastq.gz (R1 paired)**
- **43: Trimmomatic on hcc1395_normal_rep2_r1.fastq.gz (R1 paired)**
- **47: Trimmomatic on hcc1395_normal_rep3_r1.fastq.gz (R1 paired)**
- **51: Trimmomatic on hcc1395_tumor_rep1_r1.fastq.gz (R1 paired)**
- **55: Trimmomatic on hcc1395_tumor_rep2_r1.fastq.gz (R1 paired)**
- **59: Trimmomatic on hcc1395_tumor_rep3_r1.fastq.gz (R1 paired)**
- **40: Trimmomatic on hcc1395_normal_rep1_r2.fastq.gz (R2 paired)**
- **44: Trimmomatic on hcc1395_normal_rep2_r2.fastq.gz (R2 paired)**
- **48: Trimmomatic on hcc1395_normal_rep3_r2.fastq.gz (R2 paired)**
- **52: Trimmomatic on hcc1395_tumor_rep1_r2.fastq.gz (R2 paired)**
- **56: Trimmomatic on hcc1395_tumor_rep2_r2.fastq.gz (R2 paired)**
- **60: Trimmomatic on hcc1395_tumor_rep3_r2.fastq.gz (R2 paired)**

It produces 6 outputs:

- **87: HISAT2 on data 40, data 39, and data 13: aligned reads (BAM)**
- **88: HISAT2 on data 44, data 43, and data 13: aligned reads (BAM)**
- **89: HISAT2 on data 48, data 47, and data 13: aligned reads (BAM)**
- **90: HISAT2 on data 52, data 51, and data 13: aligned reads (BAM)**
- **91: HISAT2 on data 56, data 55, and data 13: aligned reads (BAM)**
- **92: HISAT2 on data 60, data 59, and data 13: aligned reads (BAM)**

StringTie



https://rnabio.org/module-09-appendix/0009/12/01/StrandSettings/

StringTie

Specify strand information

Unstranded

Select 'Forward (FR)' if your reads are from a forward-stranded library, 'Reverse (RF)' if your reads are from a reverse-stranded library, or 'Unstranded' if your reads are not from a stranded library. See Help section below for more information. Default: Unstranded

Use a reference file to guide assembly?

Use reference GTF/GFF3 ← step5

Use the reference annotation file (in GTF or GFF3 format) to guide the assembly process. The output will include expressed reference transcripts as well as any novel transcripts that are assembled. This option is required by option -e (Use Reference transcripts only), see below. (-G)

Reference file

Use a file from history ← step6

GTF/GFF3 dataset to guide assembly

14: GRCh38_latest_genomic.gff (as gff3) ← step7

Use Reference transcripts only?

No

Limit the processing of read alignments to only estimate and output the assembled transcripts matching the reference transcripts given with the -G option. With this option, read bundles with no reference transcripts (novel transcripts) will be entirely skipped, which may provide a considerable speed boost when the given set of reference transcripts is limited to a set of target genes, for example. Default: No (-e)

Output files for differential expression?

No additional output

Select to output additional files that can be used with Ballgown or DESeq2/edgeR. See Help section below for more information

Output coverage file?

No

StringTie

Executed **StringTie** and successfully added 6 jobs to the queue.

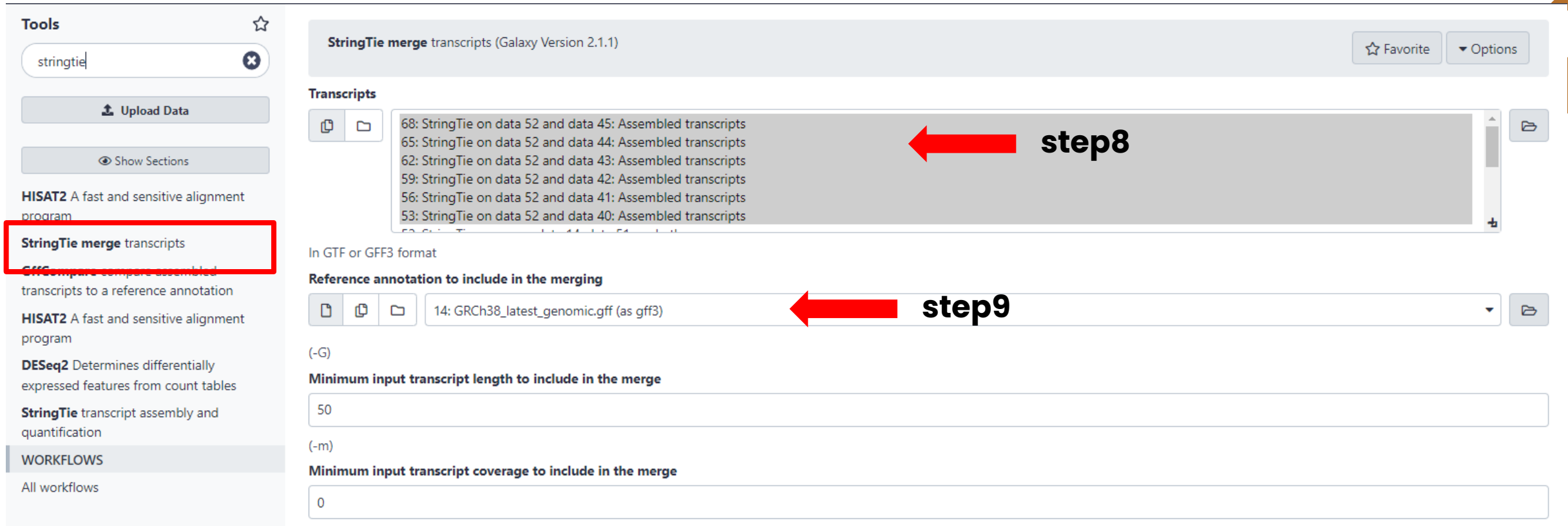
The tool uses 7 inputs:

- **87: HISAT2 on data 40, data 39, and data 13: aligned reads (BAM)**
- **88: HISAT2 on data 44, data 43, and data 13: aligned reads (BAM)**
- **89: HISAT2 on data 48, data 47, and data 13: aligned reads (BAM)**
- **90: HISAT2 on data 52, data 51, and data 13: aligned reads (BAM)**
- **91: HISAT2 on data 56, data 55, and data 13: aligned reads (BAM)**
- **92: HISAT2 on data 60, data 59, and data 13: aligned reads (BAM)**
- **14: hg38_refseq_gff (as gff3)**

It produces 6 outputs:

- **99: StringTie on data 14 and data 87: Assembled transcripts**
- **100: StringTie on data 14 and data 88: Assembled transcripts**
- **101: StringTie on data 14 and data 89: Assembled transcripts**
- **102: StringTie on data 14 and data 90: Assembled transcripts**
- **103: StringTie on data 14 and data 91: Assembled transcripts**
- **104: StringTie on data 14 and data 92: Assembled transcripts**

StringTie merge



StringTie merge

Executed **StringTie merge** and successfully added 1 job to the queue.

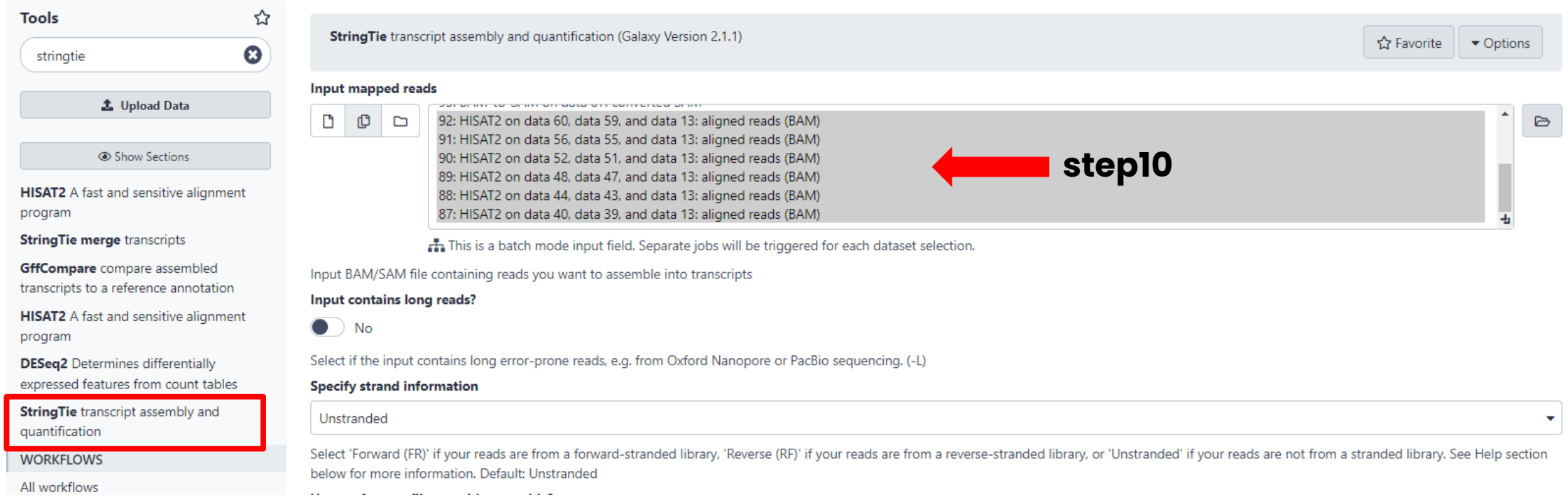
The tool uses 7 inputs:

- **99: StringTie on data 14 and data 87: Assembled transcripts**
- **100: StringTie on data 14 and data 88: Assembled transcripts**
- **101: StringTie on data 14 and data 89: Assembled transcripts**
- **102: StringTie on data 14 and data 90: Assembled transcripts**
- **103: StringTie on data 14 and data 91: Assembled transcripts**
- **104: StringTie on data 14 and data 92: Assembled transcripts**
- **14: hg38_refseq_gff (as gff3)**

It produces this output:

- **105: StringTie merge on data 14, data 104, and others**

Re-run StringTie



https://rnabio.org/module-09-appendix/0009/12/01/StrandSettings/

Re-run StringTie

Input contains long reads?

No

Select if the input contains long error-prone reads, e.g. from Oxford Nanopore or PacBio sequencing. (-L)

Specify strand information

Unstranded

Select 'Forward (FR)' if your reads are from a forward-stranded library, 'Reverse (RF)' if your reads are from a reverse-stranded library, or 'Unstranded' if your reads are not from a stranded library. See Help section below for more information. Default: Unstranded

Use a reference file to guide assembly?

Use reference GTF/GFF3 ← **step11**

Use the reference annotation file (in GTF or GFF3 format) to guide the assembly process. The output will include expressed reference transcripts as well as any novel transcripts that are assembled. This option is required by option -e (Use Reference transcripts only), see below. (-G)

Reference file

Use a file from history ← **step12**

GTF/GFF3 dataset to guide assembly

105: StringTie merge on data 14, data 104, and others

Use Reference transcripts only?

Yes

Limit the processing of read alignments to only estimate and output the assembled transcripts matching the reference transcripts given with the -G option. With this option, read bundles with no reference transcripts (novel transcripts) will be entirely skipped, which may provide a considerable speed boost when the given set of reference transcripts is limited to a set of target genes, for example. Default: No (-e)

Output files for differential expression?

DESeq2/edgeR/limma-voom ← **step13**

Select to output additional files that can be used with Ballgown or DESeq2/edgeR. See Help section below for more information

Re-run StringTie

Executed **StringTie** and successfully added 6 jobs to the queue.

The tool uses 7 inputs:

- **87: HISAT2 on data 40, data 39, and data 13: aligned reads (BAM)**
- **88: HISAT2 on data 44, data 43, and data 13: aligned reads (BAM)**
- **89: HISAT2 on data 48, data 47, and data 13: aligned reads (BAM)**
- **90: HISAT2 on data 52, data 51, and data 13: aligned reads (BAM)**
- **91: HISAT2 on data 56, data 55, and data 13: aligned reads (BAM)**
- **92: HISAT2 on data 60, data 59, and data 13: aligned reads (BAM)**
- **105: StringTie merge on data 14, data 104, and others**

It produces 18 outputs:

- **112: StringTie on data 105 and data 87: Assembled transcripts**
- **113: StringTie on data 105 and data 87: Gene counts**
- **114: StringTie on data 105 and data 87: Transcript counts**
- **115: StringTie on data 105 and data 88: Assembled transcripts**
- **116: StringTie on data 105 and data 88: Gene counts**
- **117: StringTie on data 105 and data 88: Transcript counts**
- **118: StringTie on data 105 and data 89: Assembled transcripts**
- **119: StringTie on data 105 and data 89: Gene counts**
- **120: StringTie on data 105 and data 89: Transcript counts**
- **121: StringTie on data 105 and data 90: Assembled transcripts**
- **122: StringTie on data 105 and data 90: Gene counts**
- **123: StringTie on data 105 and data 90: Transcript counts**
- **124: StringTie on data 105 and data 91: Assembled transcripts**
- **125: StringTie on data 105 and data 91: Gene counts**
- **126: StringTie on data 105 and data 91: Transcript counts**
- **127: StringTie on data 105 and data 92: Assembled transcripts**
- **128: StringTie on data 105 and data 92: Gene counts**
- **129: StringTie on data 105 and data 92: Transcript counts**

DESeq2


Tools
deseq2
Upload Data
Show Sections
StringTie merge transcripts
DESeq2 Determines differentially expressed features from count tables
StringTie transcript assembly and quantification
WORKFLOWS
All workflows
DESeq2 Determines differentially expressed features from count tables (Galaxy Version 2.11.40.6+galaxy1)
Favorite
Options
how
Select datasets per level
step14
Factor
1: Factor
Specify a factor name, e.g. effects_drug_x or cancer_markers
Human
Only letters, numbers and underscores will be retained in this field
Factor level
1: Factor level
Specify a factor level, typical values could be 'tumor', 'normal', 'treated' or 'control'
Normal
Only letters, numbers and underscores will be retained in this field
Counts file(s)
123: StringTie on data 105 and data 90: Transcript counts
122: StringTie on data 105 and data 90: Gene counts
121: StringTie on data 105 and data 90: Assembled transcripts
120: StringTie on data 105 and data 89: Transcript counts
119: StringTie on data 105 and data 89: Gene counts
118: StringTie on data 105 and data 89: Assembled transcripts
117: StringTie on data 105 and data 88: Transcript counts

DESeq2

Factor level

1: Factor level

Specify a factor level, typical values could be 'tumor', 'normal', 'treated' or 'control'

Normal

step15

Only letters, numbers and underscores will be retained in this field

Counts file(s)

120: StringTie on data 105 and data 89: Transcript counts
119: StringTie on data 105 and data 89: Gene counts
118: StringTie on data 105 and data 89: Assembled transcripts
117: StringTie on data 105 and data 88: Transcript counts
116: StringTie on data 105 and data 88: Gene counts
115: StringTie on data 105 and data 88: Assembled transcripts
114: StringTie on data 105 and data 87: Transcript counts

2: Factor level

Specify a factor level, typical values could be 'tumor', 'normal', 'treated' or 'control'

Tumor

step16

Only letters, numbers and underscores will be retained in this field

Counts file(s)

129: StringTie on data 105 and data 92: Transcript counts
128: StringTie on data 105 and data 92: Gene counts
127: StringTie on data 105 and data 92: Assembled transcripts
126: StringTie on data 105 and data 91: Transcript counts
125: StringTie on data 105 and data 91: Gene counts
124: StringTie on data 105 and data 91: Assembled transcripts
123: StringTie on data 105 and data 90: Transcript counts

+ Insert Factor level

DESeq2

Executed **DESeq2** and successfully added 1 job to the queue.

The tool uses 6 inputs:

- **114: StringTie on data 105 and data 87: Transcript counts**
- **117: StringTie on data 105 and data 88: Transcript counts**
- **120: StringTie on data 105 and data 89: Transcript counts**
- **123: StringTie on data 105 and data 90: Transcript counts**
- **126: StringTie on data 105 and data 91: Transcript counts**
- **129: StringTie on data 105 and data 92: Transcript counts**

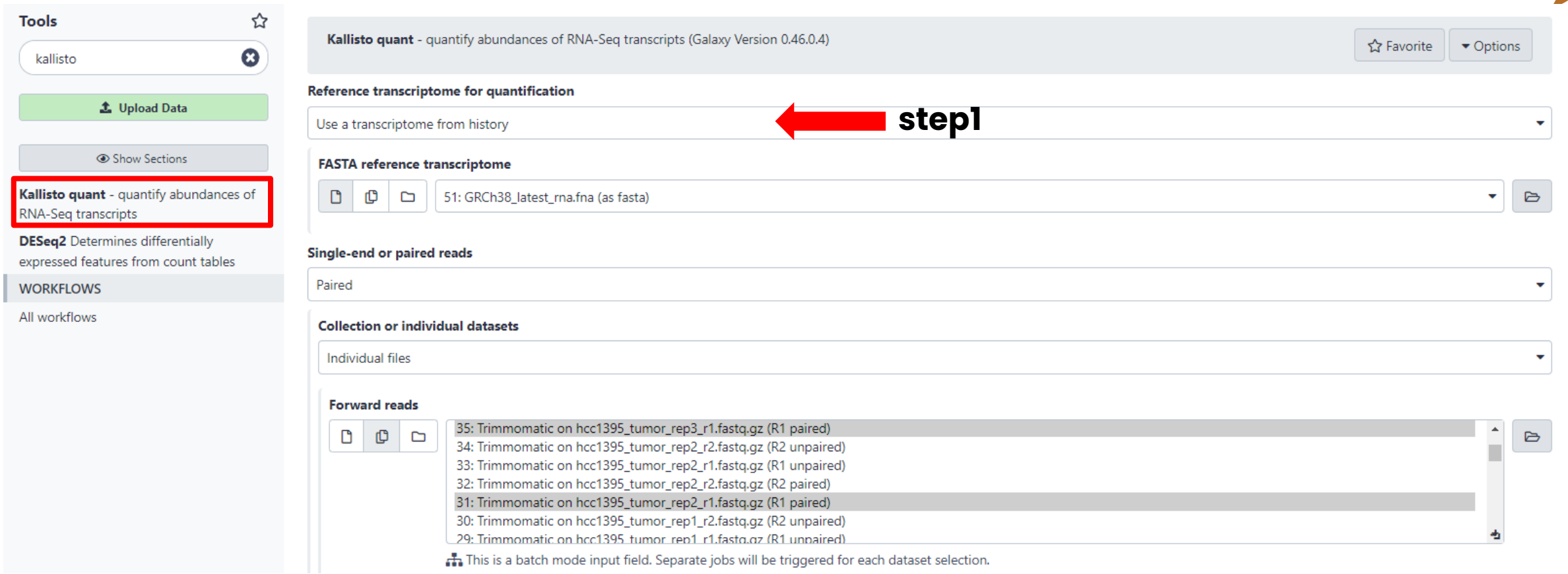
It produces 2 outputs:

- **130: DESeq2 result file on data 129, data 126, and others**
- **131: DESeq2 plots on data 129, data 126, and others**

Method-2

Kallisto+DESeq2

Kallisto


Tools
kallisto
Upload Data
Show Sections
Kallisto quant - quantify abundances of RNA-Seq transcripts
DESeq2 Determines differentially expressed features from count tables
WORKFLOWS
All workflows
Kallisto quant - quantify abundances of RNA-Seq transcripts (Galaxy Version 0.46.0.4)
Favorite
Options
Reference transcriptome for quantification
Use a transcriptome from history
step1
FASTA reference transcriptome
51: GRCh38_latest_rna.fna (as fasta)
Single-end or paired reads
Paired
Collection or individual datasets
Individual files
Forward reads
35: Trimmomatic on hcc1395_tumor_rep3_r1.fastq.gz (R1 paired)
34: Trimmomatic on hcc1395_tumor_rep2_r2.fastq.gz (R2 unpaired)
33: Trimmomatic on hcc1395_tumor_rep2_r1.fastq.gz (R1 unpaired)
32: Trimmomatic on hcc1395_tumor_rep2_r2.fastq.gz (R2 paired)
31: Trimmomatic on hcc1395_tumor_rep2_r1.fastq.gz (R1 paired)
30: Trimmomatic on hcc1395_tumor_rep1_r2.fastq.gz (R2 unpaired)
29: Trimmomatic on hcc1395_tumor_rep1_r1.fastq.gz (R1 unpaired)
This is a batch mode input field. Separate jobs will be triggered for each dataset selection.

Kallisto

Single-end or paired reads

Paired

step2

Collection or individual datasets

Individual files

Forward reads

35: Trimmomatic on hcc1395_tumor_rep3_r1.fastq.gz (R1 paired)
34: Trimmomatic on hcc1395_tumor_rep2_r2.fastq.gz (R2 unpaired)
33: Trimmomatic on hcc1395_tumor_rep2_r1.fastq.gz (R1 unpaired)
32: Trimmomatic on hcc1395_tumor_rep2_r2.fastq.gz (R2 paired)
31: Trimmomatic on hcc1395_tumor_rep2_r1.fastq.gz (R1 paired)
30: Trimmomatic on hcc1395_tumor_rep1_r2.fastq.gz (R2 unpaired)
29: Trimmomatic on hcc1395_tumor_rep1_r1.fastq.gz (R1 unpaired)

This is a batch mode input field. Separate jobs will be triggered for each dataset selection.

Reverse reads

37: Trimmomatic on hcc1395_tumor_rep3_r1.fastq.gz (R1 unpaired)
36: Trimmomatic on hcc1395_tumor_rep3_r2.fastq.gz (R2 paired)
35: Trimmomatic on hcc1395_tumor_rep3_r1.fastq.gz (R1 paired)
34: Trimmomatic on hcc1395_tumor_rep2_r2.fastq.gz (R2 unpaired)
33: Trimmomatic on hcc1395_tumor_rep2_r1.fastq.gz (R1 unpaired)
32: Trimmomatic on hcc1395_tumor_rep2_r2.fastq.gz (R2 paired)
31: Trimmomatic on hcc1395_tumor_rep2_r1.fastq.gz (R1 paired)

This is a batch mode input field. Separate jobs will be triggered for each dataset selection.

https://rnabio.org/module-09-appendix/0009/12/01/StrandSettings/

Kallisto

Executed **Kallisto quant** and successfully added 6 jobs to the queue.

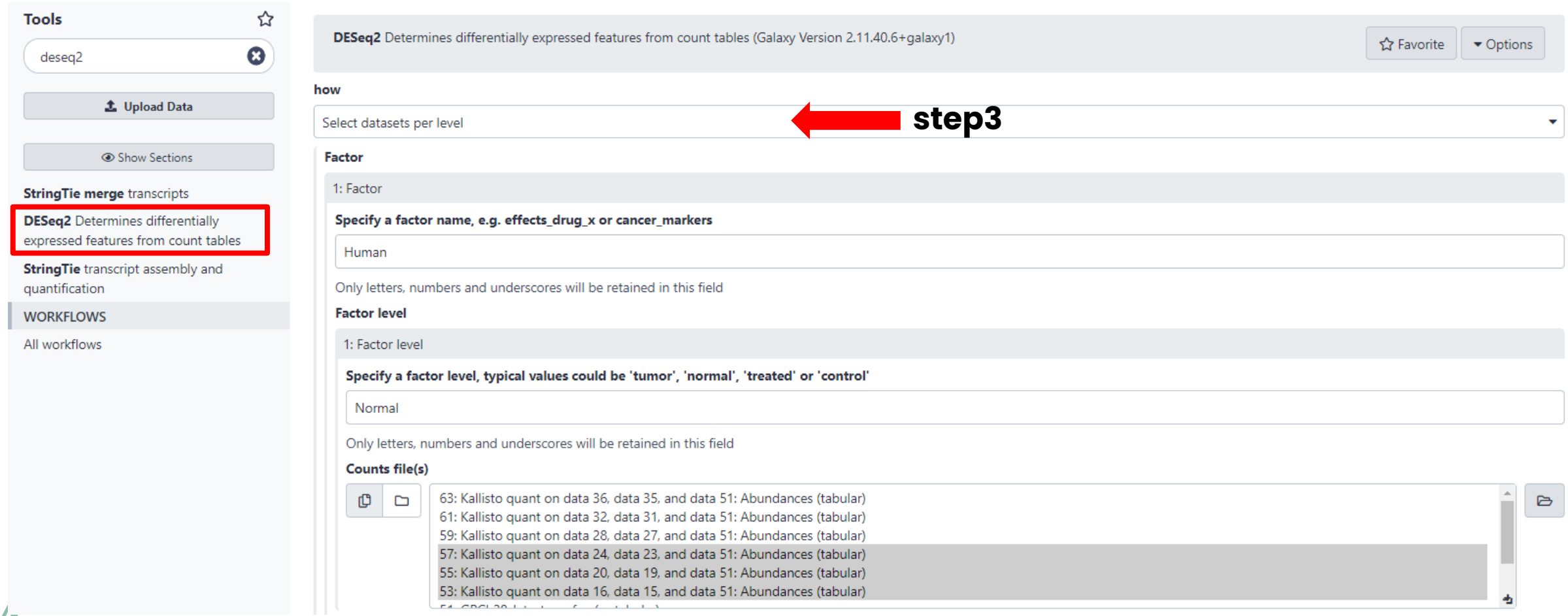
The tool uses 13 inputs:

- **51: GRCh38_latest_rna.fna (as fasta)**
- **15: Trimmomatic on hcc1395_normal_rep1_r1.fastq.gz (R1 paired)**
- **19: Trimmomatic on hcc1395_normal_rep2_r1.fastq.gz (R1 paired)**
- **23: Trimmomatic on hcc1395_normal_rep3_r1.fastq.gz (R1 paired)**
- **27: Trimmomatic on hcc1395_tumor_rep1_r1.fastq.gz (R1 paired)**
- **31: Trimmomatic on hcc1395_tumor_rep2_r1.fastq.gz (R1 paired)**
- **35: Trimmomatic on hcc1395_tumor_rep3_r1.fastq.gz (R1 paired)**
- **16: Trimmomatic on hcc1395_normal_rep1_r2.fastq.gz (R2 paired)**
- **20: Trimmomatic on hcc1395_normal_rep2_r2.fastq.gz (R2 paired)**
- **24: Trimmomatic on hcc1395_normal_rep3_r2.fastq.gz (R2 paired)**
- **28: Trimmomatic on hcc1395_tumor_rep1_r2.fastq.gz (R2 paired)**
- **32: Trimmomatic on hcc1395_tumor_rep2_r2.fastq.gz (R2 paired)**
- **36: Trimmomatic on hcc1395_tumor_rep3_r2.fastq.gz (R2 paired)**

It produces 12 outputs:

- **52: Kallisto quant on data 16, data 15, and data 51: Abundances (HDF5)**
- **53: Kallisto quant on data 16, data 15, and data 51: Abundances (tabular)**
- **54: Kallisto quant on data 20, data 19, and data 51: Abundances (HDF5)**
- **55: Kallisto quant on data 20, data 19, and data 51: Abundances (tabular)**
- **56: Kallisto quant on data 24, data 23, and data 51: Abundances (HDF5)**
- **57: Kallisto quant on data 24, data 23, and data 51: Abundances (tabular)**
- **58: Kallisto quant on data 28, data 27, and data 51: Abundances (HDF5)**
- **59: Kallisto quant on data 28, data 27, and data 51: Abundances (tabular)**
- **60: Kallisto quant on data 32, data 31, and data 51: Abundances (HDF5)**
- **61: Kallisto quant on data 32, data 31, and data 51: Abundances (tabular)**
- **62: Kallisto quant on data 36, data 35, and data 51: Abundances (HDF5)**
- **63: Kallisto quant on data 36, data 35, and data 51: Abundances (tabular)**

DESeq2


Tools
deseq2
Upload Data
Show Sections
StringTie merge transcripts
DESeq2 Determines differentially expressed features from count tables
StringTie transcript assembly and quantification
WORKFLOWS
All workflows
DESeq2 Determines differentially expressed features from count tables (Galaxy Version 2.11.40.6+galaxy1)
Favorite
Options
how
Select datasets per level
step3
Factor
1: Factor
Specify a factor name, e.g. effects_drug_x or cancer_markers
Human
Only letters, numbers and underscores will be retained in this field
Factor level
1: Factor level
Specify a factor level, typical values could be 'tumor', 'normal', 'treated' or 'control'
Normal
Only letters, numbers and underscores will be retained in this field
Counts file(s)
63: Kallisto quant on data 36, data 35, and data 51: Abundances (tabular)
61: Kallisto quant on data 32, data 31, and data 51: Abundances (tabular)
59: Kallisto quant on data 28, data 27, and data 51: Abundances (tabular)
57: Kallisto quant on data 24, data 23, and data 51: Abundances (tabular)
55: Kallisto quant on data 20, data 19, and data 51: Abundances (tabular)
53: Kallisto quant on data 16, data 15, and data 51: Abundances (tabular)

DESeq2

Factor level

1: Factor level

Specify a factor level, typical values could be 'tumor', 'normal', 'treated' or 'control'

Normal

step4

Only letters, numbers and underscores will be retained in this field

Counts file(s)

63: Kallisto quant on data 36, data 35, and data 51: Abundances (tabular)
61: Kallisto quant on data 32, data 31, and data 51: Abundances (tabular)
59: Kallisto quant on data 28, data 27, and data 51: Abundances (tabular)
57: Kallisto quant on data 24, data 23, and data 51: Abundances (tabular)
55: Kallisto quant on data 20, data 19, and data 51: Abundances (tabular)
53: Kallisto quant on data 16, data 15, and data 51: Abundances (tabular)

2: Factor level

Specify a factor level, typical values could be 'tumor', 'normal', 'treated' or 'control'

Tumor

step5

Only letters, numbers and underscores will be retained in this field

Counts file(s)

63: Kallisto quant on data 36, data 35, and data 51: Abundances (tabular)
61: Kallisto quant on data 32, data 31, and data 51: Abundances (tabular)
59: Kallisto quant on data 28, data 27, and data 51: Abundances (tabular)
57: Kallisto quant on data 24, data 23, and data 51: Abundances (tabular)
55: Kallisto quant on data 20, data 19, and data 51: Abundances (tabular)
53: Kallisto quant on data 16, data 15, and data 51: Abundances (tabular)

DESeq2

(Optional) provide a tabular file with additional batch factors to include in the model.

Nothing selected

You can produce this file using RUVSeq or svaseq.

Files have header?

Yes

If this option is set to Yes, the tool will assume that the count files have column headers in the first row. Default: Yes

Choice of Input data

TPM values (e.g. from kallisto, sailfish or salmon)

step6

Program used to generate TPMs

kallisto

Gene mapping format

GTF/GFF3

GTF/GFF3 annotation file

14: GRCh38_latest_genomic.gff (as gff3)

DESeq2

Executed **DESeq2** and successfully added 1 job to the queue.

The tool uses 7 inputs:

- **123: Kallisto quant on data 41, data 40, and data 15: Abundances (tabular)**
- **125: Kallisto quant on data 45, data 44, and data 15: Abundances (tabular)**
- **127: Kallisto quant on data 49, data 48, and data 15: Abundances (tabular)**
- **129: Kallisto quant on data 53, data 52, and data 15: Abundances (tabular)**
- **131: Kallisto quant on data 57, data 56, and data 15: Abundances (tabular)**
- **133: Kallisto quant on data 61, data 60, and data 15: Abundances (tabular)**
- **14: GRCh38_latest_genomic.gff (as gff3)**

It produces 2 outputs:

- **136: DESeq2 result file on data 14, data 133, and others**
- **137: DESeq2 plots on data 14, data 133, and others**

You can check the status of queued jobs and view the resulting data by refreshing the History panel. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

Analysis

Drawing Venn diagram

Subject :

1. **Filter the gene/transcript in P-adj value < 0.01**
2. **Translate transcript ID (result in Hisat2+StringTie+DESeq2) to GeneID**
3. **Compare the gene names from the two methods (using Venn diagram)**

Result in kallisto + DESeq2

GeneID	Base mean	log2(FC)	StdErr	Wald-Stats	P-value	P-adj
APOBEC3C	1238.53686914908	4.2537109038166	0.0810897590769195	52.4568200995843	0	0
ATF4	5516.89270803537	-1.88657896148456	0.0271759715088854	-69.4208470474636	0	0
EMID1	1130.78468458702	3.98655082720641	0.0788707094417703	50.545390746734	0	0
GRK3	1585.17617784959	2.22321663035666	0.0511766622447742	43.4420013505996	0	0

Result in hisat2 + StringTie + DESeq2

GeneID	Base mean	log2(FC)	StdErr	Wald-Stats	P-value	P-adj
rna-NM_000362.5	7300.93519814528	-9.60933647500719	0.870220615898659	-11.042414187216	2.38536614773783e-28	6.08983977517467e-25
rna-NM_002872.5	2950.55873141655	7.13956949515504	0.788612924278076	9.05332549766522	1.38679151138018e-19	1.7702393642768e-16
id-IGLC1	2083.44730064633	11.0887014007596	1.29504162902587	8.56242853683435	1.10514542867322e-17	9.40478759800911e-15
MSTRG.789.3	2358.52402884027	13.4970013263314	1.69684454692317	7.95417668094876	1.80326635470545e-15	1.15093475089076e-12

Operations

Method-1
Hisat2+SringTie+DESeq2

Filtered by P-adj value

Filtered by transcript ID, get gene list

Venn diagram

Method-2
Kallisto+DESeq2

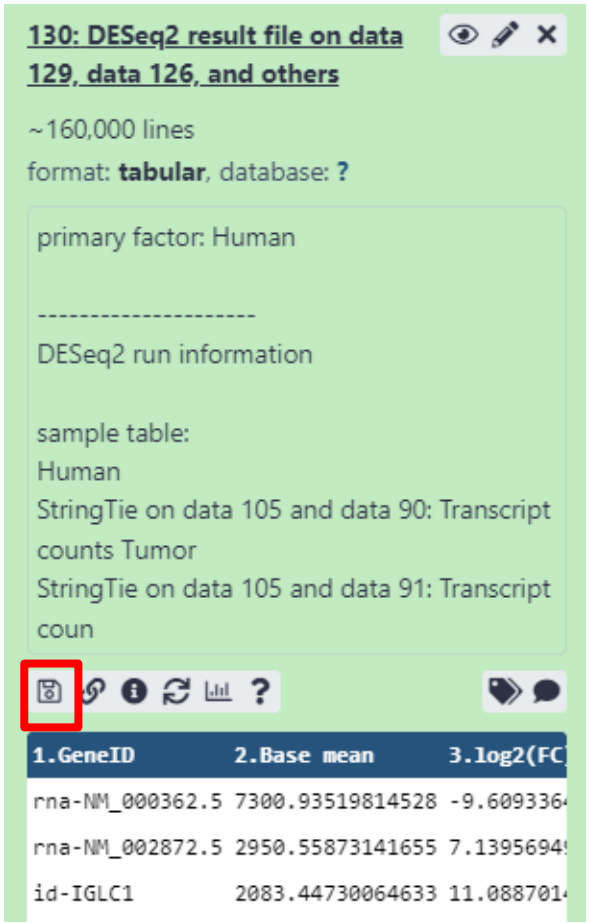
Filtered by P-adj value

You may use programming language to analyze the data, or follow the following guideline using excel

Method-1 for example

GeneID column

P-adj column



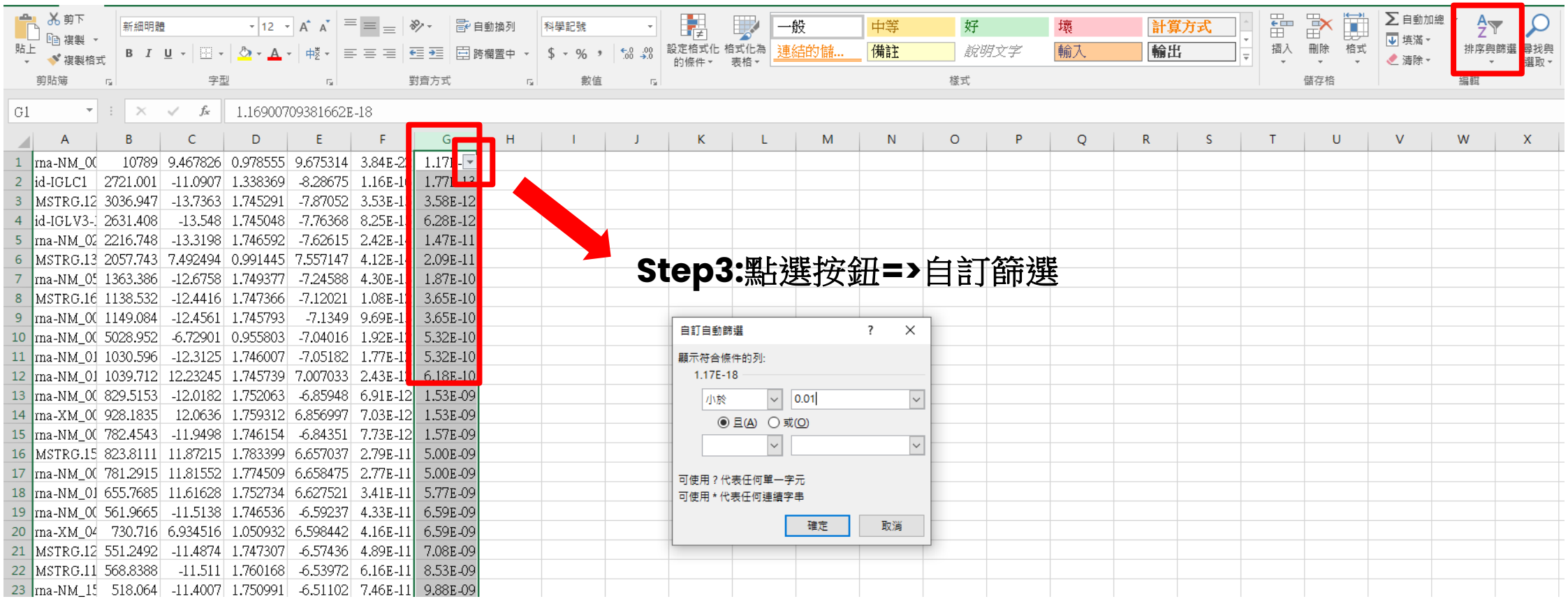
Open with Excel

	A	B	C	D	E	F	G	H	I
1	rna-NM_0(	10789	9.467826	0.978555	9.675314	3.84E-22	1.17E-18		
2	id-IGLC1	2721.001	-11.0907	1.338369	-8.28675	1.16E-16	1.77E-13		
3	MSTRG.12	3036.947	-13.7363	1.745291	-7.87052	3.53E-15	3.58E-12		
4	id-IGLV3-1	2631.408	-13.548	1.745048	-7.76368	8.25E-15	6.28E-12		
5	rna-NM_02	2216.748	-13.3198	1.746592	-7.62615	2.42E-14	1.47E-11		
6	MSTRG.13	2057.743	7.492494	0.991445	7.557147	4.12E-14	2.09E-11		
7	rna-NM_05	1363.386	-12.6758	1.749377	-7.24588	4.30E-13	1.87E-10		
8	MSTRG.16	1138.532	-12.4416	1.747366	-7.12021	1.08E-12	3.65E-10		
9	rna-NM_0(	1149.084	-12.4561	1.745793	-7.1349	9.69E-13	3.65E-10		
10	rna-NM_0(	5028.952	-6.72901	0.955803	-7.04016	1.92E-12	5.32E-10		
11	rna-NM_01	1030.596	-12.3125	1.746007	-7.05182	1.77E-12	5.32E-10		
12	rna-NM_01	1039.712	12.23245	1.745739	7.007033	2.43E-12	6.18E-10		
13	rna-NM_0(	829.5153	-12.0182	1.752063	-6.85948	6.91E-12	1.53E-09		
14	rna-XM_0(	928.1835	12.0636	1.759312	6.856997	7.03E-12	1.53E-09		
15	rna-NM_0(	782.4543	-11.9498	1.746154	-6.84351	7.73E-12	1.57E-09		
16	MSTRG.15	823.8111	11.87215	1.783399	6.657037	2.79E-11	5.00E-09		
17	rna-NM_0(	781.2915	11.81552	1.774509	6.658475	2.77E-11	5.00E-09		
18	rna-NM_01	655.7685	11.61628	1.752734	6.627521	3.41E-11	5.77E-09		

If you can’t open .tabular file, you can simply change the file extension to .tsv

Step1: Select column

Step2:點選篩選



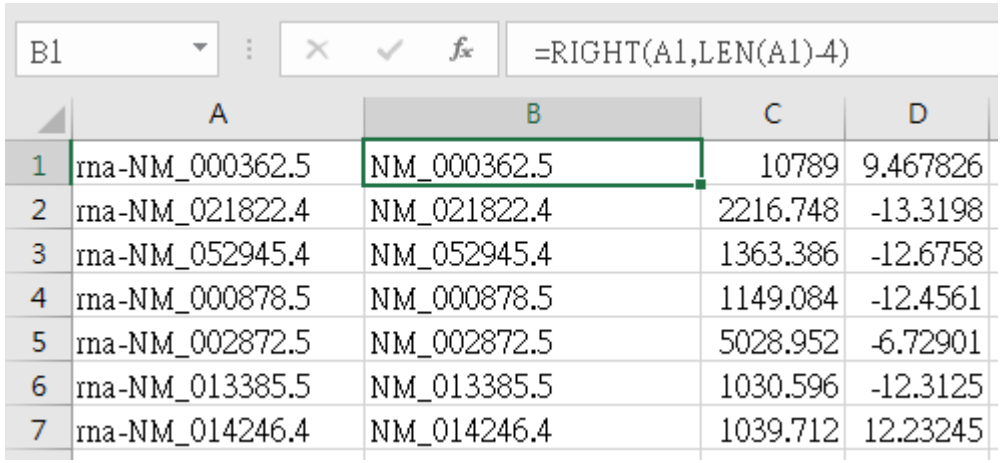
	A	B	C	D	E	F	G
1	rna-NM_00	10789	9.467826	0.978555	9.675314	3.84E-2	1.17E-
2	id-IGLC1	2721.001	-11.0907	1.338369	-8.28675	1.16E-1	1.77E-12
3	MSTRG.12	3036.947	-13.7363	1.745291	-7.87052	3.53E-1	3.58E-12
4	id-IGLV3-	2631.408	-13.548	1.745048	-7.76368	8.25E-1	6.28E-12
5	rna-NM_02	2216.748	-13.3198	1.746592	-7.62615	2.42E-1	1.47E-11
6	MSTRG.13	2057.743	7.492494	0.991445	7.557147	4.12E-1	2.09E-11
7	rna-NM_05	1363.386	-12.6758	1.749377	-7.24588	4.30E-1	1.87E-10
8	MSTRG.16	1138.532	-12.4416	1.747366	-7.12021	1.08E-1	3.65E-10
9	rna-NM_00	1149.084	-12.4561	1.745793	-7.1349	9.69E-1	3.65E-10
10	rna-NM_00	5028.952	-6.72901	0.955803	-7.04016	1.92E-1	5.32E-10
11	rna-NM_01	1030.596	-12.3125	1.746007	-7.05182	1.77E-1	5.32E-10
12	rna-NM_01	1039.712	12.23245	1.745739	7.007033	2.43E-1	6.18E-10
13	rna-NM_00	829.5153	-12.0182	1.752063	-6.85948	6.91E-12	1.53E-09
14	rna-XM_00	928.1835	12.0636	1.759312	6.856997	7.03E-12	1.53E-09
15	rna-NM_00	782.4543	-11.9498	1.746154	-6.84351	7.73E-12	1.57E-09
16	MSTRG.15	823.8111	11.87215	1.783399	6.657037	2.79E-11	5.00E-09
17	rna-NM_00	781.2915	11.81552	1.774509	6.658475	2.77E-11	5.00E-09
18	rna-NM_01	655.7685	11.61628	1.752734	6.627521	3.41E-11	5.77E-09
19	rna-NM_00	561.9665	-11.5138	1.746536	-6.59237	4.33E-11	6.59E-09
20	rna-XM_04	730.716	6.934516	1.050932	6.598442	4.16E-11	6.59E-09
21	MSTRG.12	551.2492	-11.4874	1.747307	-6.57436	4.89E-11	7.08E-09
22	MSTRG.11	568.8388	-11.511	1.760168	-6.53972	6.16E-11	8.53E-09
23	rna-NM_15	518.064	-11.4007	1.750991	-6.51102	7.46E-11	9.88E-09

Step3:點選按鈕=>自訂篩選

Step4: 分別對P-adj column 小於0.01以及包含”rna-”的geneID column篩選

Step5: Remove “rna-” to generate ID list

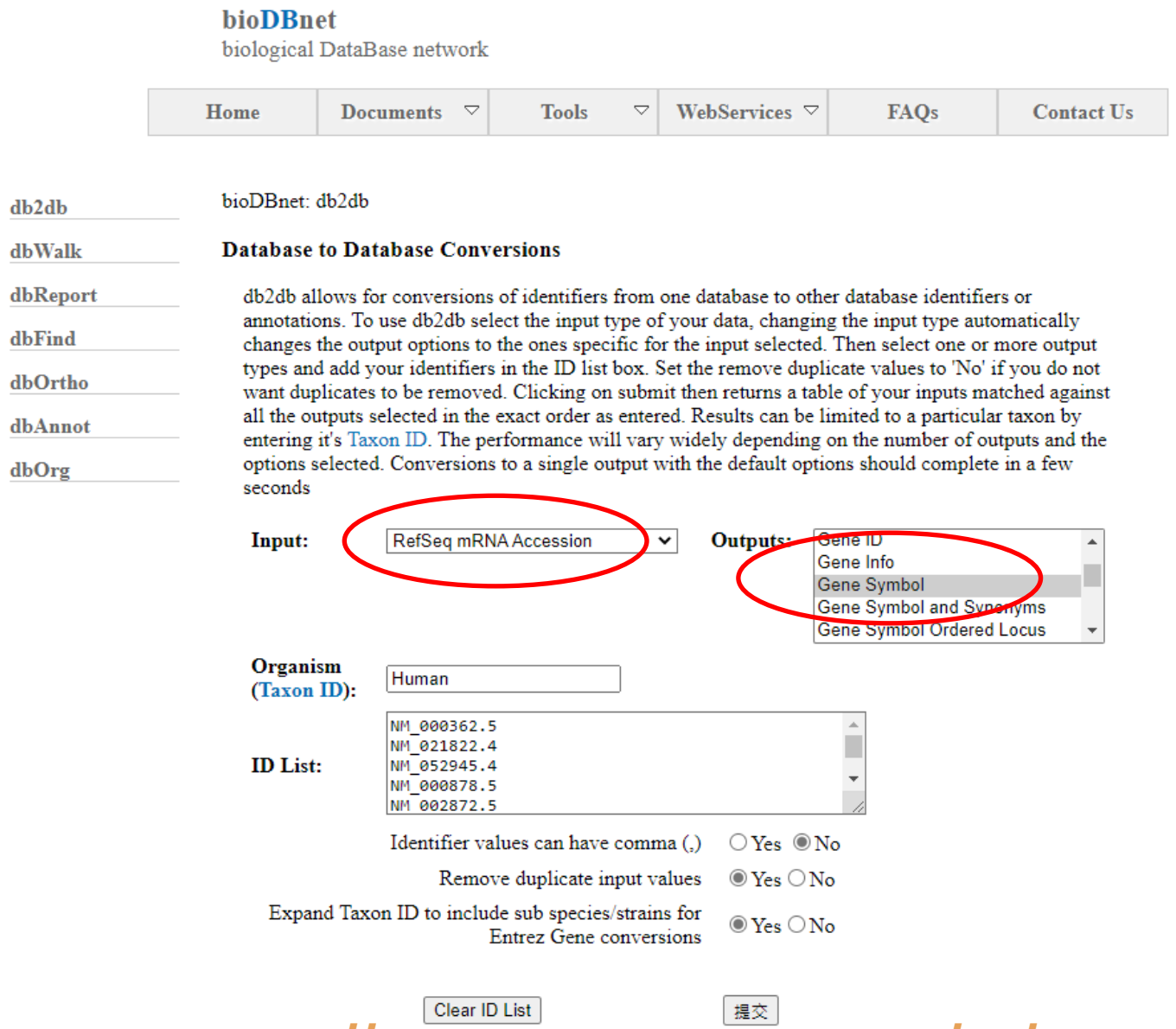
取”A1”欄位的右邊”LEN(A1)-4”個字母



B1 | =RIGHT(A1,LEN(A1)-4)

	A	B	C	D
1	rna-NM_000362.5	NM_000362.5	10789	9.467826
2	rna-NM_021822.4	NM_021822.4	2216.748	-13.3198
3	rna-NM_052945.4	NM_052945.4	1363.386	-12.6758
4	rna-NM_000878.5	NM_000878.5	1149.084	-12.4561
5	rna-NM_002872.5	NM_002872.5	5028.952	-6.72901
6	rna-NM_013385.5	NM_013385.5	1030.596	-12.3125
7	rna-NM_014246.4	NM_014246.4	1039.712	12.23245

Step6: Translate ID list to gene list



Tool: https://biodbnet-abcc.ncifcrf.gov/db/db2db.php

Step7: Draw Venn diagram

Tool: https://bioinformatics.psb.ugent.be/webtools/Venn/

THANKS!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.