nf-core/RNAseq nextflow workflow on genotoul

Nextflow is a bioinformatics workflow tool to run tasks across multiple compute infrastructures in a very portable manner. Nextflow is a workflow manager. The community develop and maintain workflows for several kind of high throughput data into nf-core repository (`https://github.com/nf-core`)

The rnaseq workflow is available to all genotoul cluster users.
"The workflow processes raw data from FastQ inputs (FastQC, Trim Galore!), aligns the reads (STAR or HiSAT2), generates gene counts (featureCounts, StringTie) and performs extensive quality-control on the results (RSeQC, dupRadar, Preseq, edgeR, MultiQC). See the output documentation for more details of the results."

The documentation summarize how to use this workflow on genotoul cluster. The description of all the used tools and the options are available at `https://github.com/nf-core/rnaseq/blob/master/docs/usage.md` and will not be explain in this document.

File How_to_use

`/usr/local/bioinfo/src/NextflowWorkflows/How_to_use_SLURM_NextflowWorkflows`

Load module:

```
module load bioinfo/nfcore
```

Nextflow help

```
nextflow run -help
```

Workflow help

```
nextflow run nf-core/rnaseq --help
```

Example of sbatch file

```
#!/bin/bash
#SBATCH -p workq
#SBATCH -t 1:00  #time in minutes

module load bioinfo/nfcore-rnaseq-1.1

nextflow run nf-core/rnaseq \
 --reads
'/usr/local/bioinfo/src/NextflowWorkflows/example_on_cluster/data/*_{1,2}_Ch6.f
astq.gz'\
  --fasta
/usr/local/bioinfo/src/NextflowWorkflows/example_on_cluster/data/ITAG2.3_genomi
c_Ch6.fasta\
  --gtf
/usr/local/bioinfo/src/NextflowWorkflows/example_on_cluster/data/ITAG_pre2.3_ge
ne_models_Ch6.gtf
```

Default options

```
params {
  container = 'nfcore/rnaseq:1.1' // Container slug. Stable releases should
specify release tag!

  // Pipeline Options
  aligner = 'star'
  genome = false
  forward_stranded = false
  reverse_stranded = false
  unstranded = false
  splicesites = false
  outdir = './results'
  saveReference = false
  saveTrimmed = false
  saveAlignedIntermediates = false
  singleEnd = false
  reads = "data/*{1,2}.fastq.gz"
  outdir = './results'

  // Custom trimming options
  clip_r1 = 0
  clip_r2 = 0
  three_prime_clip_r1 = 0
  three_prime_clip_r2 = 0

  // AWS Batch
  awsqueue = false
  awsregion = 'eu-west-1'

  // Defaults
  sampleLevel = false
  clusterOptions = false
  hisatBuildMemory = 200 // Required amount of memory in GB to build HISAT2
index with splice sites
  subsampFilesizeThreshold = 10000000000 // Don't subsample BAMs for RSeQC
gene_body_coverage if less than this
  maxMultiqcEmailFileSize = 25.MB
  readPaths = null
  tracedir = "${params.outdir}/pipeline_info"
}
```

Results

The pipeline will create the following files in your working directory:

```
work            # Directory containing the nextflow working files
results         # Finished results (configurable, see below)
.nextflow_log   # Log file from Nextflow
# Other nextflow hidden files, eg. history of pipeline runs and old logs.
```

Relaunch an aborted workflow

Warning : if you relaunch the nextflow command line in a previous working directory, the entire workflow will be relaunch if you don’t set option -resume.

```
nextflow run nf-core/rnaseq -resume\
 --reads
'/usr/local/bioinfo/src/NextflowWorkflows/example_on_cluster/data/*_{1,2}_Ch6.f
astq.gz'\
  --fasta
/usr/local/bioinfo/src/NextflowWorkflows/example_on_cluster/data/ITAG2.3_genomi
c_Ch6.fasta\
  --gtf
//usr/local/bioinfo/src/NextflowWorkflows/example_on_cluster/data/ITAG_pre2.3_g
ene_models_Ch6.gtf
```