

Reproducible, scalable bioinformatics workflows with nextflow and nf-core

Shirley Li

Bioinformatician

TTS Research Technology

Yucheng Zhang

Bioinformatics Engineer

TTS Research Technology

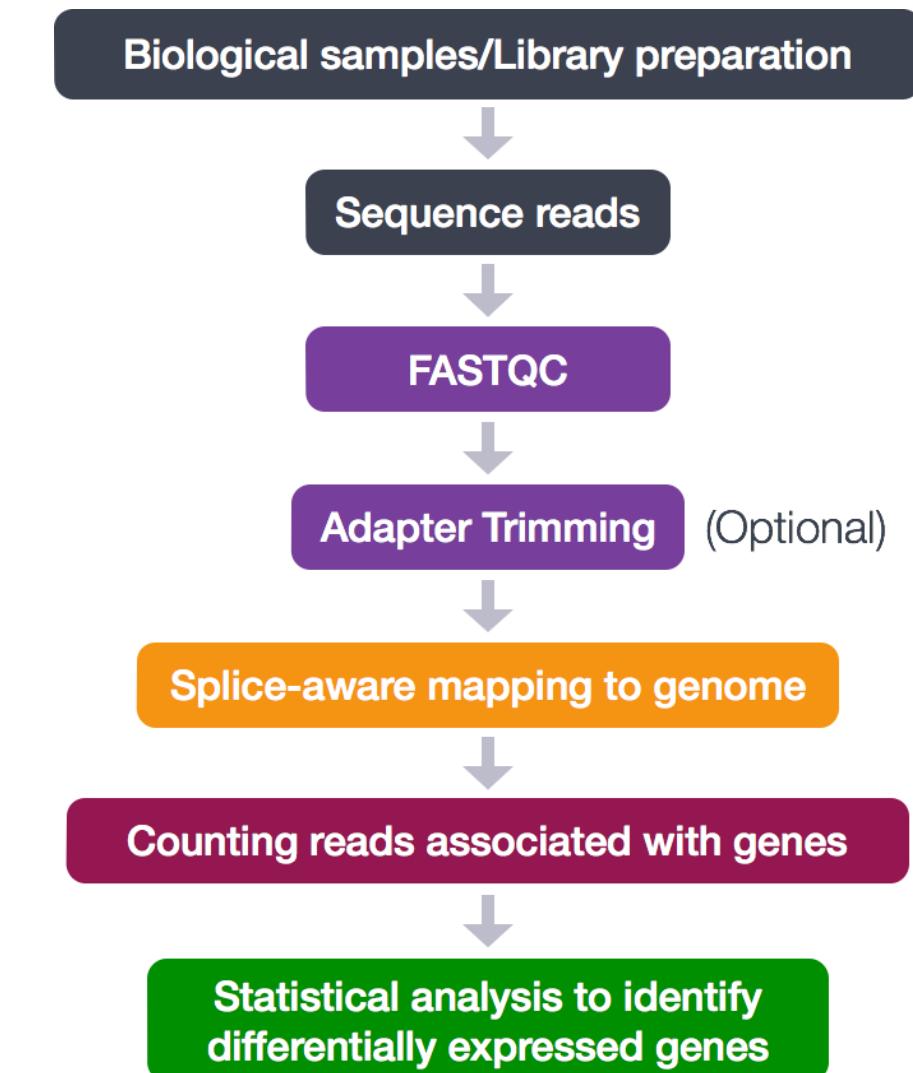
Overview

1. Intro to nextflow and nf-core
2. Clean cache data
3. Nextflow configuration files
4. How to run nf-core pipelines on Tufts HPC
5. Troubleshooting
6. Hands-on demo

Nextflow

Workflow

- ❖ A pipeline is a collection of several analysis steps
- ❖ Steps are linked by input/output files
- ❖ One often needs to run the same workflow for several samples



Bad workflows

```
## fastp
fastp -i SRR1553607_1.fastq -o SRR1553607_1.fastq.trimmed.fq --max_len1 20
fastp -i SRR1553607_2.fastq -o SRR1553607_2.fastq.trimmed.fq --max_len1 20
fastp -i SRR1972917_1.fastq -o SRR1972917_1.fastq.trimmed.fq --max_len1 20
fastp -i SRR1972917_2.fastq -o SRR1972917_2.fastq.trimmed.fq --max_len1 20
## fastqc
fastqc SRR1553607_1.fastq.trimmed.fq
fastqc SRR1553607_2.fastq.trimmed.fq
fastqc SRR1972917_1.fastq.trimmed.fq
fastqc SRR1972917_1.fastq.trimmed.fq
```

Bad workflows: for loop

```
## fastp
for name in *.fastq; do
    fastp -i $name -o ${name%.*}.trimmed.fq --max_len1 20
done
```

```
## fastqc
for name in *.trimmed.fq; do
    fastqc -i $name
done
```

- For loop runs only one command at a time.
- Our computers have many cores so that we could run multiple commands at the same time.
- We could add & operator to the end of the command to run it in the background.
- But then it runs all commands simultaneously, which we don't want either.
- **We want to run as many commands as we have compute cores, but no more.**

What is a good workflow?

- **Automated:** Runs automatically without manual effort.
- **Scalable:** Can process large datasets and many samples efficiently.
- **Reproducible:** Allows others to easily repeat and get the same results.
- **Error Handling:** Includes checks to catch and manage errors.
- **Modular:** Steps can be reused or adapted for different analyses.

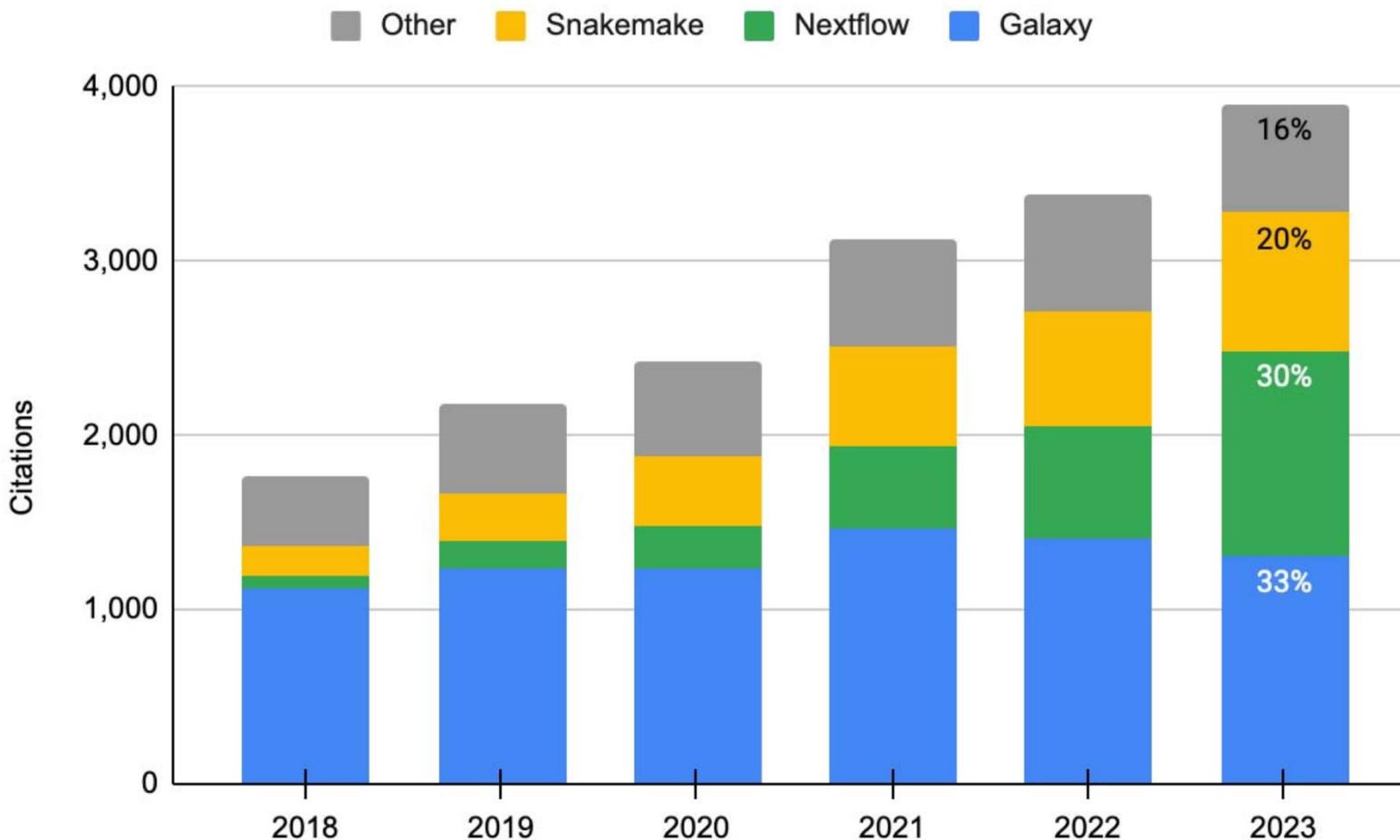


FIGURE 1: Google Scholar citation counts for bioinformatics workflow management systems. Sum of citations of the major publications of Galaxy, Nextflow, and Snakemake between 2018 and 2023 (Data in Supplementary Table 1).

Langer, Bjorn E., et al. "Empowering bioinformatics communities with Nextflow and nf-core." *bioRxiv* (2024): 2024-05.

nextflow pipeline

Write code
in any language



Orchestrate tasks with
dataflow programming



Define software
dependencies
via containers



Built-in version
control with Git



nextflow runtime

Task orchestration
and execution

Supported Platforms



nextflow run

```
#!/usr/bin/env nextflow

params.greeting = 'Hello world!'
greeting_ch = Channel.of(params.greeting)

process SPLITLETTERS {
    input:
    val x

    output:
    path 'chunk_*'

    script:
    """
    printf '$x' | split -b 6 - chunk_
    """
}

process CONVERTTOUPPER {
    input:
    path y

    output:
    stdout

    script:
    """
    cat $y | tr '[a-z]' '[A-Z]'
    """
}

workflow {
    letters_ch = SPLITLETTERS(greeting_ch)
    results_ch = CONVERTTOUPPER(letters_ch.flatten())
    results_ch.view{ it }
}
```

```
[yzhang85@c1cmp063 nf-training]$ nextflow run hello.nf
Nextflow 23.10.1 is available - Please consider updating your version to it
N E X T F L O W ~ version 23.10.0
Launching `hello.nf` [furious_newton] DSL2 - revision: 3c3d5e1897
executor > local (3)
[8f/3b8107] process > SPLITLETTERS (1) [100%] 1 of 1 ✓
[d3/4546d4] process > CONVERTTOUPPER (1) [100%] 2 of 2 ✓
WORLD!
HELLO
```

More information can be found on their website

Documentation: <https://www.nextflow.io/docs/latest/index.html>

Training: <https://training.nextflow.io/>

Examples: <https://www.nextflow.io/example1.html>

Running a Nextflow Pipeline from GitHub on HPC

```
1 module load nextflow/24.04.1  
2 module load singularity  
3  
4 nextflow run nf-core/rnaseq ...
```

- Load required modules
- Run the pipeline using nextflow

<https://github.com/nf-core/rnaseq/tree/3.16.1>

The screenshot shows the GitHub repository page for `nf-core / rnaseq`. The repository is public and has 49 issues and 11 pull requests. The current branch is `3.16.1`, which has 23 branches and 33 tags. A merge commit by `maxulysse` has been pushed to the master branch. The repository contains several directories: `.devcontainer`, `.github`, `assets`, and `bin`.

Code Issues 49 Pull requests 11 Actions

rnaseq Public

3.16.1 23 Branches 33 Tags

maxulysse Merge branch 'dev' into master ✓

.devcontainer Template update for n

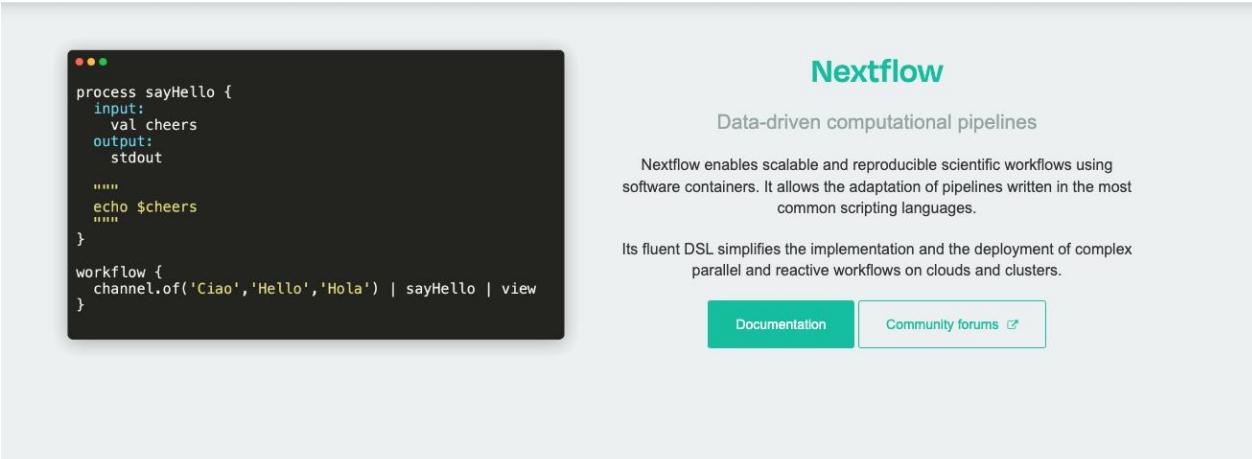
.github Properly disable cond

assets Merge branch 'dev' in

bin Remove restatement c

<https://www.nextflow.io/>

More information can be found on their website



The screenshot shows the official Nextflow website. At the top, there's a navigation bar with links for Documentation, Examples, Training, Resources, Forums, and a search icon. Below the navigation is a large image of a computer terminal window displaying Nextflow code. To the right of the terminal, the word "Nextflow" is written in a large, bold, teal font, followed by the subtitle "Data-driven computational pipelines". A brief description explains that Nextflow enables scalable and reproducible scientific workflows using software containers, allowing adaptation of pipelines written in various scripting languages. It highlights the fluent DSL for implementing complex parallel and reactive workflows. At the bottom of this section are two buttons: "Documentation" and "Community forums".



nf-core: Curated Analysis Pipelines



A community effort to collect a curated set of analysis pipelines built using Nextflow.

<https://nf-co.re/pipelines>

Pipelines

Browse the 113 pipelines that are currently available as part of nf-core.

Search

Released 68

Under development 32

Archived 13

Stars

88

≡

rnaseq ✓ ★ 885

RNA sequencing analysis pipeline using STAR, RSEM, HISAT2 or Salmon with gene/isoform counts and extensive quality control.

rna rna-seq

3.16.1 released 2 days ago

sarek ✓ ★ 399

Analysis pipeline to detect germline or somatic variants (pre-processing, variant calling and annotation) from WGS / targeted sequencing

annotation cancer gatk4 genomics
germline pre-processing somatic
target-panels variant-calling
whole-exome-sequencing
whole-genome-sequencing

3.4.4 released about 2 months ago

mag ✓ ★ 211

Assembly and binning of metagenomes

annotation assembly binning
long-read-sequencing metagenomes
metagenomics nanopore nanopore-sequencing

3.1.0 released 14 days ago

scrnaseq ✓ ★ 210

A single-cell RNAseq pipeline for 10X genomics data

10x-genomics 10xgenomics alevin bustools
cellranger kallisto rna-seq single-cell
star-solo

2.7.1 released 2 months ago

chipseq ✓ ★ 190

ChIP-seq peak-calling, QC and differential analysis pipeline.

chip chip-seq chromatin-immunoprecipitation

ampliseq ✓ ★ 184

Amplicon sequencing analysis workflow using DADA2 and QIIME2

16s 18s amplicon-sequencing edna illumina iontorrent its metabarcoding metagenomics microbiome pablo qiime2

atacseq ✓ ★ 184

ATAC-seq peak-calling and QC analysis pipeline

nanoseq ✓ ★ 177

Nanopore demultiplexing, QC and alignment pipeline

Local nf-core pipelines

HPC system administrators have downloaded popular nf-core pipelines and stored them in the following directory:

/cluster/tufts/biocontainers/nf-core/pipelines/

```
[yzhang85@login-prod-03 ~]$ ls /cluster/tufts/biocontainers/nf-core/pipelines/
nf-core-ampliseq/          nf-core-mag/           nf-core-rnasplice/
nf-core-atacseq/            nf-core-metatdenovo/   nf-core-sarek/
nf-core-chipseq/            nf-core-methylseq/    nf-core-scrnaseq/
nf-core-differentialabundance/ nf-core-nanoseq/    nf-core-smrnaseq/
nf-core-eager/              nf-core-nanostring/   nf-core-taxprofiler/
nf-core-fetchngs/           nf-core-pangenome/   nf-core-viralrecon/
nf-core-funcscan/           nf-core-rnafusion/  
nf-core-hic/                nf-core-rnaseq/
```

No downloads each time, faster runs, more efficient!

Run local nf-core pipelines

```
1 module load nf-core-rnaseq/3.16.0  
2  
3 rnaseq --help
```

- **Recommended!**
- No download each time

OR

```
1 module load nextflow/24.04.1  
2 module load singularity  
3  
4 nextflow run nf-core/rnaseq ...
```

- Download the pipeline from GitHub Repo each time, less efficient

Usage instructions and documentation

Each pipeline has its own webpage at https://nf-co.re/<pipeline_name>

The screenshot shows the nf-core Pipelines landing page. At the top, there is a green header with the title "Pipelines" and a subtitle "Browse the 113 pipelines that are currently available as part of nf-core.". Below the header is a search bar and navigation buttons for "Released" (68), "Under development" (32), and "Archived" (13). There are also buttons for "Stars" (with a dropdown arrow), a grid icon, and a menu icon. The main content area displays a table of pipelines, each with a name, description, release status, star count, and last release date.

Name	Description	Released	Stars	Last release
rnaseq	RNA sequencing analysis pipeline using STAR, RSEM, HISAT2 or Salmon with gene/isoform counts and extensive quality control.	✓	885	3.16.1
sarek	Analysis pipeline to detect germline or somatic variants (pre-processing, variant calling and annotation) from WGS / targeted sequencing	✓	399	3.4.4
mag	Assembly and binning of metagenomes	✓	211	3.1.0
scrnaseq	A single-cell RNAseq pipeline for 10X genomics data	✓	210	2.7.1
chipseq	ChIP-seq peak-calling, QC and differential analysis pipeline.	✓	190	2.1.0
ampliseq	Amplicon sequencing analysis workflow using DADA2 and QIIME2	✓	184	2.11.0
atacseq	ATAC-seq peak-calling and QC analysis pipeline	✓	184	2.1.2

The screenshot shows the nf-core/rnaseq documentation page. At the top, there's a green header with the title "nf-core/rnaseq" and a subtitle: "RNA sequencing analysis pipeline using STAR, RSEM, HISAT2 or Salmon with gene/isoform counts and extensive quality control." Below the header, there are two small buttons: "rna" and "rna-seq". On the right side of the header is an "Edit" button with a pencil icon. In the center of the page is a green button labeled "Launch version 3.16.1" with a key icon. Below it is a link to the GitHub repository: "https://github.com/nf-core/rnaseq". The main content area has a light gray background. At the top of this area is a navigation bar with tabs: "Introduction" (disabled), "Usage" (selected), "Parameters", "Output", "Results", "Releases", and a dropdown menu set to "3.16.1". To the right of the navigation bar is a sidebar titled "On this page" which lists various pipeline components: Pipeline parameters, Samplesheet input, FASTQ sampling, Adapter trimming options, Alignment options, Quantification options, Reference genome options, Contamination screening options, and Running the pipeline. The "Pipeline parameters" section contains text about providing pipeline parameters via the CLI or Nextflow `-params-file` option. The "Samplesheet input" section contains text about creating a samplesheet with sample information before running the pipeline.

RNA sequencing analysis pipeline using STAR, RSEM, HISAT2 or Salmon with gene/isoform counts and extensive quality control.

rna rna-seq

Launch version 3.16.1

<https://github.com/nf-core/rnaseq>

→ Introduction Usage Parameters Output Results ↲ Releases 3.16.1

Pipeline parameters

Please provide pipeline parameters via the CLI or Nextflow `-params-file` option. Custom config files including those provided by the `-c` Nextflow option can be used to provide any configuration except for parameters; see [docs](#).

Samplesheet input

You will need to create a samplesheet with information about the samples you would like to analyse before running the pipeline. Use this parameter to specify its location. It has to be a comma-separated file with 4 columns, and a header row as shown in the examples below.

On this page

- Pipeline parameters
- Samplesheet input
- FASTQ sampling
- Adapter trimming options
- Alignment options
- Quantification options
- Reference genome options
- Contamination screening options
- Running the pipeline

Check Instructions locally

```
1 nextflow run  
· /cluster/tufts/biocontainers/nf-core/pipelines/nf-core-rnaseq/3.14.0/3_14_0/ --help
```

```
[yzhang85@login-prod-03 ~]$ nextflow run /cluster/tufts/biocontainers/nf-core/pipelines/nf-core-rnaseq/3.14.0/3_14_0/ --help  
Nextflow 23.10.1 is available - Please consider updating your version to it  
N E X T F L O W ~ version 23.10.0  
Launching `/cluster/tufts/biocontainers/nf-core/pipelines/nf-core-rnaseq/3.14.0/3_14_0/main.nf` [lonely_wright] DSL2 - revision: 746820de9b
```



Typical pipeline command:

```
nextflow run nf-core/rnaseq --input samplesheet.csv --genome GRCh37 -profile docker
```

Input/output options

--input	[string]	Path to comma-separated file containing information about the samples in the experiment.
--outdir	[string]	The output directory where the results will be saved. You have to use absolute paths to storage
on Cloud		infrastructure.
--email	[string]	Email address for completion summary.
--multiqc_title	[string]	MultiQC report title. Printed as page header, used for filename if not otherwise specified.

Reference genome options

--genome	[string]	Name of iGenomes reference.
--fasta	[string]	Path to FASTA genome file.
--gtf	[string]	Path to GTF annotation file.
--gff	[string]	Path to GFF3 annotation file.

Singularity in nf-core Pipelines

In the context of nf-core pipelines, singularity is used to package and run all the software and dependencies required by the pipeline in a self-contained container. This ensures that the pipeline runs consistently, regardless of the system it's executed on—whether on an HPC cluster or a local machine.

Users can learn more about singularity usages from our previous container [training](#).



NXF_SINGULARITY_CACHEDIR in nf-core Pipelines

NXF_SINGULARITY_CACHEDIR: an environment variable used to specify where **singularity** images are stored on the cluster.

Storing these images locally can **speed up pipeline execution**, as they don't need to be downloaded every time.

Public & Personal NXF_SINGULARITY_CACHEDIR

If you want to run the nf-core pipelines managed by system admins, please define **NXF_SINGULARITY_CACHEDIR** like this:

```
1 | export  
· | NXF_SINGULARITY_CACHEDIR=/cluster/tufts/biocontainers/nf-core/singularity-images
```

However, if you need to run your own pipelines, you have to define **NXF_SINGULARITY_CACHEDIR** to your own directory.

Please do not use your \$HOME.

cache and resume

Cache and resume

The nextflow caching mechanism works by assigning a unique ID to each task which is used to create a separate execution directory where the tasks are executed and the results stored.

The task unique ID is generated as a 128-bit hash value composing the task input values, files and command string.



resume

Usage: **nextflow run <script> -resume**

-resume allows the continuation of a workflow execution from the last step that was completed successfully.

```
WORKFLOW=/cluster/tufts/biocontainers/nf-core/pipelines/nf-core-rnaseq/3.14.0/3_14_0
nextflow run $WORKFLOW \
    --input $input \
    --outdir $outdir
    --genome GRCh38 \
    --aligner star_rsem \
    -profile tufts \
    -resume
```

Clean up

After a pipeline is completed with success, it's better to clean up **work** directory to save space.

You can remove the work directory completely by:
rm -rf work



nextflow log & nextflow clean

- Check information on nextflow runs by running **nextflow log** inside your project folder
- **nextflow clean** together with the RUN NAME to clean cache.

```
nf-training -> nextflow log
TIMESTAMP          DURATION      RUN NAME    STATUS   REVISION ID    SESSION ID           COMMAND
2024-03-07 21:03:07  2.9s        clever_darwin  OK      3c3d5e1897  c6f83839-fb98-45af-9090-6807b02a1800  nextflow run hello.nf
2024-03-07 21:03:33  1.8s        chaotic_faggin OK      86d466d737  9a963a51-3351-4c1a-8d7d-ed7643c11c44  nextflow run script1.nf
nf-training -> nextflow clean clever_darwin -f
Removed /workspace/gitpod/nf-training/work/f2/14d3a75f9b4c683bcf5e361931bcc9
Removed /workspace/gitpod/nf-training/work/ea/0cf312c156b549204e8b8b438739ed
Removed /workspace/gitpod/nf-training/work/f8/91d79e889abde3cf52c41e1a078320
nf-training -> □
```

Configs

Config files

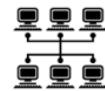
nf-core pipelines make use of nextflow's configuration files to specify how the pipelines runs, define custom parameters and what software management system to use e.g. docker, singularity or conda.



Default 'base' config (always loaded)



Core profiles (e.g. singularity, conda, test)



Institutional profiles (nf-core/configs)



Your local config files (-c flag)

Default base config

```
nextflow run nf-core/<pipeline>
```

-  Automatically loaded
-  Sensible default resource allocation
-  No software packaging specified
-  Runs locally, no job submission

Core profiles

```
nextflow run nf-core/<pipeline> -profile singularity
```

Specify software packaging



Docker



Singularity



Conda

Specify test profile



<https://github.com/zhan4429/ContainerWorkshop2024Spring-Tufts>

Institutional profiles

```
nextflow run nf-core/<pipeline> -profile mycluster
```

⇒ Specifies job submission

 Specify software packaging

Works for:

 For all pipelines

 For all users on your system

 Single point to update

Organisations

Some of the organisations running nf-core pipelines are listed below, along with a key person who you can contact for advice.

Note

Expand ▾

Zoom out



National Genomics Infrastructure



The NGI provides next-generation sequencing services for Swedish academic groups. Many of the nf-core pipelines started life as SciLifeLab / NGI workflows.

Quantitative Biology Center



The Quantitative Biology Center provides a one-stop-shop for access to high-throughput technologies in the life sciences and the required bioinformatics analysis. As a bioinformatics core facility we provide

<https://nf-co.re/contributors>

tufts profile



```
params {
    max_memory = 120.GB
    max_cpus = 72
    max_time = 168.h
    igenomes_base = '/cluster/tufts/biocontainers/datasets/igenomes/'
}

process {
    executor = 'slurm'
    clusterOptions = '-N 1 -n 1 -p batch'
}

executor {
    queueSize = 16
    pollInterval = '1 min'
    queueStatInterval = '5 min'
    submitRateLimit = '10 sec'
}

// Set $NXF_SINGULARITY_CACHEDIR in your ~/.bashrc
// to stop downloading the same image for every run
singularity {
    enabled = true
    autoMounts = true
}
```

<https://github.com/nf-core/configs/blob/master/conf/tufts.config>

tufts profile



SINGULARITYCE

```
// Perform work directory cleanup when the run has successfully completed  
trace {
```

```
    trace.overwrite = true
```

```
    enabled      = true
```

```
}
```

```
// On a successful completion of a Nextflow run, automatically delete all  
// intermediate files stored in the work/ directory
```

```
cleanup = true
```

```
// Allows to override the default cleanup = true behaviour for debugging
```

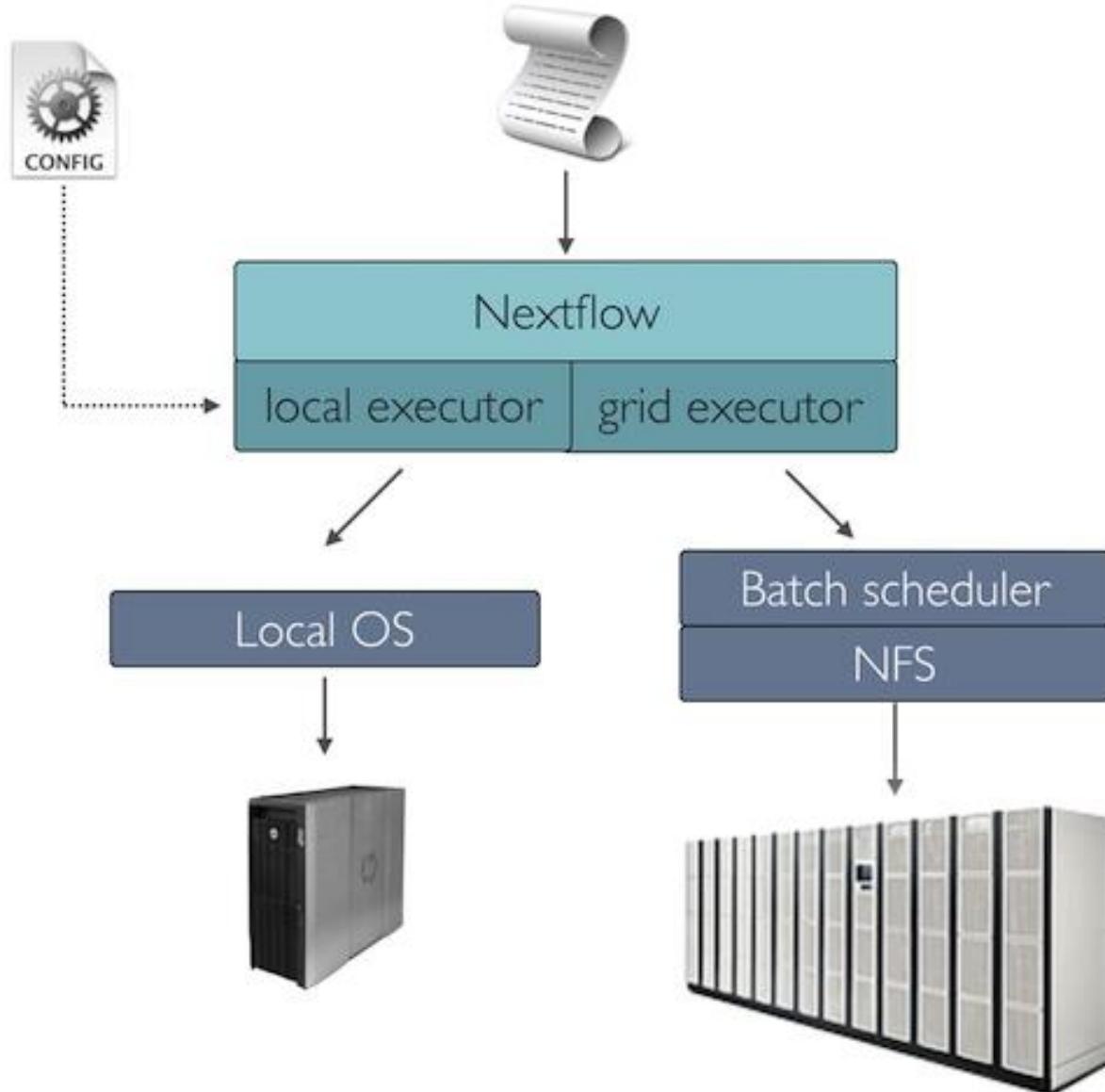
```
debug {
```

```
    cleanup = false
```

```
}
```

<https://github.com/nf-core/configs/blob/master/conf/tufts.config>

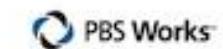
Running nf-core pipelines on Tufts HPC



UNIVA



Platform Computing
an IBM Company



```
#!/bin/bash
```

```
#SBATCH --time=00-48:00:00
#SBATCH -p batch
#SBATCH -N 1
#SBATCH -n 1
#SBATCH -c XX
#SBATCH --mem=XXG
#SBATCH --job-name nf-core
#SBATCH --output=%x-%J-%u.out
#SBATCH --error=%x-%J-%u.err
#SBATCH --mail-type=ALL
#SBATCH --mail-user=XXX@tufts.edu
```

```
module load nextflow
```

```
module load singularity
```

```
export NXF_SINGULARITY_CACHEDIR=/cluster/tufts/biocontainers/nf-core/singularity-images
```

```
nextflow run /cluster/tufts/biocontainers/nf-core/pipelines/nf-core-rnaseq/3.14.0/3_14_0/ \
    --input samplesheet.csv --outdir output \
    --fasta ref.fasta --gtf ref.gtf --aligner star_salmon \
    -profile singularity \
    --max_memory XXGB --max_cpus XX
```

Local mode

```
#!/bin/bash

#SBATCH --time=00-48:00:00
#SBATCH -p batch
#SBATCH -N 1
#SBATCH -n 1
#SBATCH -c 2 ## This is the parent script used for submitting children slurm jobs, 2 cores are enough
#SBATCH --job-name nf-core
#SBATCH --output=%x-%J-%u.out
#SBATCH --error=%x-%J-%u.err
#SBATCH --mail-type=ALL
#SBATCH --mail-user=XXX@tufts.edu

module load nextflow
module load singularity

export NXF_SINGULARITY_CACHEDIR=/cluster/tufts/biocontainers/nf-core/singularity-images

nextflow run /cluster/tufts/biocontainers/nf-core/pipelines/nf-core-rnaseq/3.14.0/3_14_0/ \
    --input samplesheet.csv --outdir output \
    --fasta ref.fasta --gtf ref.gtf \
    --aligner star_salmon \
    -profile tufts
```

Tufts profile

```
#!/bin/bash
```

Other partitions

```
#SBATCH --time=00-48:00:00
#SBATCH -p batch
#SBATCH -N 1
#SBATCH -n 2 ## This is the parent script used for submitting children slurm jobs, 2 cores are enough
#SBATCH --job-name nf-core
#SBATCH --output=%x-%J-%u.out
#SBATCH --error=%x-%J-%u.err
#SBATCH --mail-type=ALL
#SBATCH --mail-user=XXX@tufts.edu
```

```
module load nextflow
module load singularity
```

```
export NXF_SINGULARITY_CACHEDIR=/cluster/tufts/biocontainers/nf-core/singularity-images
```

```
nextflow run /cluster/tufts/biocontainers/nf-core/pipelines/nf-core-rnaseq/3.14.0/3_14_0/ \
    --input samplesheet.csv --outdir output \
    --fasta ref.fasta --gtf ref.gtf \
    --aligner star_salmon \
    -profile tufts --partition preempt
```

nf-core pipelines as modules

----- /cluster/tufts/hpc/tools/module -----			
nf-core/2.13.1	nf-core/2.14.1 (D)	----- /cluster/tufts/biocontainers/modules -----	
nf-core-ampliseq/2.8.0		nf-core-mag/2.5.4	nf-core-rnasplice/1.0.2
nf-core-ampliseq/2.9.0		nf-core-mag/3.0.0	nf-core-rnasplice/1.0.3
nf-core-ampliseq/2.10.0		nf-core-mag/3.0.2	nf-core-rnasplice/1.0.4 (D)
nf-core-ampliseq/2.11.0	(D)	nf-core-mag/3.1.0	nf-core-sarek/3.4.0
nf-core-atacseq/2.1.2		nf-core-metatdenovo/1.0.0	nf-core-sarek/3.4.1
nf-core-bacass/2.2.0		nf-core-metatdenovo/1.0.1	nf-core-sarek/3.4.3
nf-core-bacass/2.3.1	(D)	nf-core-methylseq/2.6.0	nf-core-sarek/3.4.4 (D)
nf-core-bamtofastq/2.1.1		nf-core-multiplesequencealign/1.0.0	nf-core-scnanoseq/1.0.0
nf-core-chipseq/2.0.0		nf-core-nanoseq/3.1.0	nf-core-scrnaseq/2.5.1
nf-core-chipseq/2.1.0	(D)	nf-core-nanostring/1.2.1	nf-core-scrnaseq/2.7.0
nf-core-denovotranscript/1.0.0		nf-core-nanostring/1.3.0	nf-core-scrnaseq/2.7.1 (D)
nf-core-detaxizer/1.0.0		nf-core-pairgenomealign/1.0.0	nf-core-smrnaseq/2.3.0
nf-core-differentialabundance/1.4.0		nf-core-pangenome/1.1.0	nf-core-smrnaseq/2.3.1 (D)
nf-core-differentialabundance/1.5.0	(D)	nf-core-pangenome/1.1.1	nf-core-taxprofiler/1.1.5
nf-core-eager/2.5.1		nf-core-pangenome/1.1.2	nf-core-taxprofiler/1.1.6
nf-core-fetchngs/1.11.0		nf-core-proteinifold/1.1.0	nf-core-taxprofiler/1.1.7
nf-core-fetchngs/1.12.0	(D)	nf-core-raredisease/2.0.1	nf-core-taxprofiler/1.1.8
nf-core-funcscan/1.1.4		nf-core-rnafusion/3.0.1	nf-core-taxprofiler/1.2.0 (D)
nf-core-funcscan/1.1.5	(D)	nf-core-rnafusion/3.0.2	nf-core-viralrecon/2.6.0
nf-core-hic/2.1.0		nf-core-rnaseq/3.14.0	
nf-core-mag/2.5.2		nf-core-rnaseq/3.16.0	

```
[yzhang85@login-prod-01 ~]$ module show nf-core-rnaseq/3.14.0
```

```
-----  
/cluster/tufts/biocontainers/modules/nf-core-rnaseq/3.14.0:
```

```
module-whatis    nf-core rnaseq pipeline  
module-whatis    https://nf-co.re/rnaseq  
prepend-path     PATH /cluster/tufts/biocontainers/tools/nf-core-rnaseq/3.14.0/bin  
-----
```

```
[yzhang85@login-prod-01 ~]$ more /cluster/tufts/biocontainers/tools/nf-core-rnaseq/3.14.0/bin/rnaseq  
#!/usr/bin/env bash
```

```
if [ ! $(command -v singularity) ]; then  
    module load singularity  
fi
```

```
VER=3.14.0
```

```
PKG=nf-core-rnaseq
```

```
export NXF_SINGULARITY_CACHEDIR=/cluster/tufts/biocontainers/nf-core/singularity-images  
nextflow run /cluster/tufts/biocontainers/nf-core/pipelines/nf-core-rnaseq/3.14.0/3_14_0 "$@"
```

```
[yzhang85@login-prod-01 ~]$ module load nf-core-rnaseq/3.14.0
```

```
[yzhang85@login-prod-01 ~]$ rnaseq --help
```

```
Nextflow 23.10.1 is available - Please consider updating your version to it
```

```
N E X T F L O W ~ version 23.10.0
```

```
Launching `/cluster/tufts/biocontainers/nf-core/pipelines/nf-core-rnaseq/3.14.0/3_14_0/main.nf` [cranky_hopper] DSL2 - revision: 74  
6820de9b
```



Run pipelines easily with modules

```
#!/bin/bash

#SBATCH --time=00-48:00:00
#SBATCH -p batch
#SBATCH -N 1
#SBATCH -n 1
#SBATCH -c 2 ## This is the parent script used for submitting children slurm jobs, 2 cores are enough
#SBATCH --job-name nf-core
#SBATCH --output=%x-%J-%u.out
#SBATCH --error=%x-%J-%u.err
#SBATCH --mail-type=ALL
#SBATCH --mail-user=XXX@tufts.edu

module load nf-core-rnaseq/3.14.0

rnaseq --input samplesheet.csv --outdir output \
    --fasta ref.fasta --gtf ref.gtf \
    --aligner star_salmon \
    -profile tufts
```

Troubleshooting

Start small

-profile **test,tufts**

```
[yzhang85@p1cmp045 rnaseq]$ module load nf-core-rnaseq/3.16.0
[yzhang85@p1cmp045 rnaseq]$ rnaseq -profile test,tufts --outdir testout
Nextflow 24.04.4 is available - Please consider updating your version to it
N E X T F L O W ~ version 23.10.0
Launching `/cluster/tufts/biocontainers/nf-core/pipelines/nf-core-rnaseq/3.16.0/3_16_0/main.nf` [cheesy_jepson] DSL2 - revision: f68f604b04
WARN: Access to undefined parameter `monochromeLogs` -- Initialise it to a default value eg. `params.monochromeLogs = some_value`
```



nf-core/rnaseq v3.16.0

Core Nextflow options

```
runName          : cheesy_jepson
containerEngine  : singularity
launchDir        : /cluster/tufts/rt/yzhang85/test_tufts/nf-core/rnaseq
workDir          : /cluster/tufts/rt/yzhang85/test_tufts/nf-core/rnaseq/work
projectDir       : /cluster/tufts/biocontainers/nf-core/pipelines/nf-core-rnaseq/3.16.0/3_16_0
userName         : yzhang85
profile          : test,tufts
configFiles      : 
```

Input/output options

```
input            : https://raw.githubusercontent.com/nf-core/test-datasets/626c8fab639062eade4b
10747e919341cbf9b41a/samplesheet/v3.10/samplesheet_test.csv
outdir          : testout
```

Reference genome options

```
fasta            : https://raw.githubusercontent.com/nf-core/test-datasets/626c8fab639062eade4b
10747e919341cbf9b41a/reference/genome.fasta
gtf              : https://raw.githubusercontent.com/nf-core/test-datasets/626c8fab639062eade4b
10747e919341cbf9b41a/reference/genes_with_empty_tid.gtf.gz
```

Check the basics

- Whether nextflow version is too old
- Whether required modules are loaded (nextflow and singularity)
- Haven't run out of disk space (du -f)

Check the troubleshooting docs:

- <https://nf-co.re/docs/usage/troubleshooting>

Anatomy of a work directory

- **.command.out** - STUOUT from tool
- **.command.err** – STDERR from tool
- **.command.log** - STOUT and STDERR from tool
- **.command.run** – Wrapper script used to run the job
- **.command.sh** – Process command used for this tasks
- **.command.begin** – Created ASAP the jobs launches
- **.command.trac** – Logs of computer resource usage
- **.exitcode** – Created when the job ends, with exit code

Seek help from nextflow and nf-core communities



Join Nextflow on Slack

Start by entering the email address you use for work.

 @nextflow.io ▼

Continue

You can use any account with the domain:

- nextflow.io
- seqera.io

Don't have an email address from one of those domains?
Contact the workspace administrator at [Nextflow](#) for an invitation.



See what nf-core is up to

Slack is a messaging app that brings your whole team together.



Marcel Ribeiro-Dantas, Phil Ewels and 8,382 others have already joined

We suggest using [the email account you use for work](#).

Continue With Google

Continue With Apple

Continue With Email

April. 2024



See what nf-core is up to

Slack is where work happens for companies of all sizes.



Marcel Ribeiro-Dantas, Remi-Andre Olsen and 9,942 others have already joined.

We suggest using [the email account you use for work](#).

Continue With Google

Continue With Apple

Continue With Email

Oct. 2024



OPEN
 OnDemand

ondemand.pax.tufts.edu

Open OnDemand Files Jobs Clusters Interactive Apps Bioinformatics Apps Misc

Develop Help

NOTIFICATIONS and SUPPORT REQUEST

- Request Assistance: Email tts-research@tufts.edu for questions rega
- Upload/Download: Via OnDemand web interface is limited to 976MB
- Acknowledging Usage of NSF CC* Grant Resources on Tufts HPC
- Acknowledging Usage of Tufts HPC Cluster - [Click Here](#)

Home / My Interactive Sessions

Bioinformatics Apps

Apps

- AlphaFold
- CellProfiler
- CellProfiler GPU(beta)
- CellProfiler-Analyst
- FastQC
- Jupyter Bioinfo
- QualiMap
- RELION
- RStudio for bioinformatics
- RStudio for scRNA-Seq
- Shinyngs
- petiteFinder

Jupyter Notebook (7891)

Created at: 2024-10-16 18:3

Session ID: [0f7c3b54-7538-4](#)

For debugging purposes,

Jupyter Lab (7880131)

Created at: 2024-10-15 14:0

Session ID: [780747e3-7968-4](#)

For debugging purposes,

Jupyter Notebook (7833)

nf-core pipelines

- ampliseq
- atacseq
- bacass
- bamtofastq
- chipseq
- demo
- denovotranscript
- detaxizer
- differentialabundance
- eager
- fetchngs
- funcscan
- hic
- mag
- metatdenovo
- methylseq

more days

more days

Bioinformatics Apps
Apps
AlphaFold
CellProfiler
CellProfiler GPU(beta)
FastQC
Jupyter Bioinfo
QualiMap
RELION
RStudio for bioinformatics
RStudio for scRNA-Seq
Shinyngs
nf-core pipelines
ampliseq
atacseq
bacass
bamtofastq
chipseq
detaxizer
differentialabundance
eager
fetchngs
funcscan
hic
mag
metatdenovo
methylseq
nanoseq
nanostring

fetchngs

This app will launch the [fetchngs](#) pipeline developed by nf-core community.

Number of hours

2

Which nextflow executor to use?

slurm



With slurm, tasks will be distributed to different nodes, local means all tasks will run on a single node.

Partition

batch

NOTE: Please do not choose specific lab partitions if you are not a member of that lab.

Reservation for class, training, workshop

Default

If you don't know about specific reservation, select default.

Version

1.12.0

Working Directory

/cluster/tufts/workshop/yzhang85/fetchngs

Select your project directory; defaults to \$HOME

outdir

fetchngsOut

The output directory where the results will be saved. You have to use absolute paths to storage on Cloud infrastructure.

input

samplesheet.csv

File containing SRA/ENA/GEO/DDBJ identifiers one per line to download their associated metadata and FastQ files.

ena_metadata_fields

Comma-separated list of ENA metadata fields to fetch before downloading data.

Which nextflow executor to use?

local

With slurm, tasks will be distributed to different nodes, local means all tasks will run on a single node.

Partition

batch

NOTE: Please do not choose specific lab partitions if you are not a member of that lab.

Cores

24

Number of cores (up to 128) for a shared job. Non-shared jobs will have exclusive nodes and be charged at 128 cores per node requested

Amount of memory

64GB

Reservation for class, training, workshop

Default

If you don't know about specific reservation, select default.

2578439	OnDemand/+	batch	default	2	COMPLETED	0:0
2578451	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2578452	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578453	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2578454	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578455	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2578456	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578457	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2578458	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578459	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578460	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2578461	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2578462	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578477	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2578630	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2578692	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2578693	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2578710	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578711	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578712	nf-NFCORE+	batch	default	2	COMPLETED	0:0
2578786	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2578909	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2579166	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2579310	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2580524	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2580697	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2580704	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2583415	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2583416	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2583417	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2583418	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2583421	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2583422	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2583439	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2583453	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2583470	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2583483	nf-NFCORE+	batch	default	12	COMPLETED	0:0

←master job

Hands-on demo

[https://tuftsdatalab.github.io/tuftsWorkshops/2024_workshops/
2024_bioinformatics401/03_nfcore/](https://tuftsdatalab.github.io/tuftsWorkshops/2024_workshops/2024_bioinformatics401/03_nfcore/)