Harshil Patel
Bioinformatics & Biostatistics

NBIS Epigenomics Workshop
26th November 2020

Fountain of Knowledge

Pipelines

Pipeline Name	Stars	Latest Release	Released
diaproteomics	2	1.0.0	5 days ago
rnaseq	250	2.0	5 days ago
mag	29	1.1.1	1 week ago
nanoseq	21	1.1.0	2 weeks ago
bacass	15	1.1.1	2 weeks ago
ampliseq	49	1.1.3	2 weeks ago
eager	37	2.2.1	3 weeks ago
epitopeprediction	9	1.1.0	4 weeks ago
proteomicslfq	11	1.0.0	4 weeks ago
cageseq	3	1.0.0	1 months ago
mhcquant	12	1.6.0	1 months ago
hic	22	1.2.2	3 months ago
hlatyping	19	1.2.0	3 months ago
chipseq	62	1.2.1	4 months ago
atacseq	54	1.2.1	4 months ago
rnafusion	51	1.2.0	4 months ago
viralrecon	25	1.1.0	5 months ago
sarek	79	2.6.1	5 months ago
imcyto	7	1.0.0	6 months ago
slamseq	3	1.0.0	6 months ago
coproid	4	1.1	7 months ago
methylseq	50	1.5	7 months ago
bcellmagic	10	1.2.0	10 months ago
scrnaseq	14	1.0.0	11 months ago
smrnaseq	15	1.0.0	1 years ago
nascent	2	1.0	2 years ago
deepvariant	28	1.0	2 years ago

Released 27	Under development 14	Archived 4

Pipeline Name	Stars	Latest Release	Released
artic	3	dev	-
bactmap	2	dev	-
crisprvar	0	dev	-
ddamsproteomics	4	dev	-
demultiplex	3	dev	-
denovohybrid	2	dev	-
dualrnaseq	0	dev	-
gwas	2	dev	-
kmermaid	6	dev	-
lncpipe	18	dev	-
metaboigniter	0	dev	-
mnaseseq	4	dev	-
pangenome	5	dev	-
smartseq2	6	dev	-

nf-core/ 🍏 tools

nf-core/ 🍎
atacseq

nf-core/ 🍎
chipseq

nf-core/ 🍎
mnaseseq

nf-core/ 🍎
viralrecon

nf-core/ 🍎
nanoseq

nf-core/ 🍎
imcyto

Parameter Schema

- Standard JSON format

- Custom website integration to edit the Schema

- Define a parameter once and put the kettle on!

- "nf-core lint" against parameters defined in workflow

- Parameter validation e.g. "--resume" != "-resume"

- Launch directly via website / Nextflow Tower

- Used to render:
    - Pipeline "--help"
    - Summary log information at run-time
    - Documentation on website

🚀 Launch version 1.2.1

 https://github.com/nf-core/atacseq

AWS Megatests

- Automated CI tests for full-sized data on pipeline release

- Benchmark default resource settings

- Benchmark against "truth" full-sized data (if available)

- Benchmark between pipeline releases

- Auto-rendered on nf-core website as a persistent reference

- Incredibly trivial relaunch via Nextflow Tower

- Thank you AWS! May we have some more credits please?
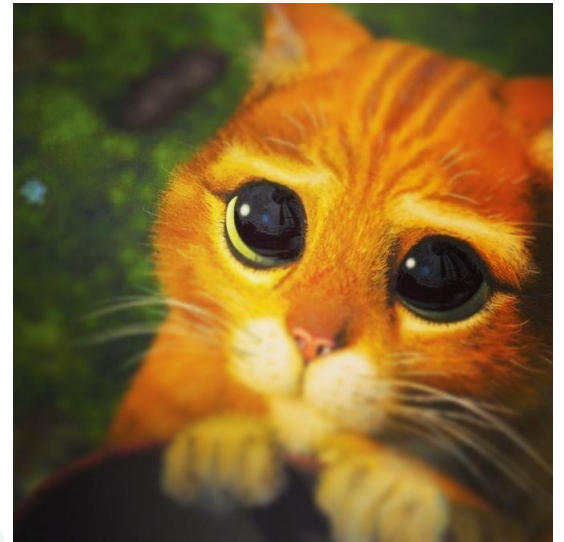
Full Test Data

```
nextflow run nf-core/chipseq -profile test_full,docker -r 1.2.1
```

```
params {
  config_profile_name        = 'Full test profile'
  config_profile_description = 'Full test dataset to check pipeline function'

  // Input data
  input   = 'https://raw.githubusercontent.com/nf-core/test-datasets/chipseq/design_full.csv'

  // Genome references
  genome = 'hg19'
}
```

Genomes

```
params {
  // illumina iGenomes reference file paths
  genomes {
    'GRCh37' {
      fasta       = "${params.igenomes_base}/Homo_sapiens/Ensembl/GRCh37/Sequence/WholeGenomeFasta/genome.fa"
      bwa         = "${params.igenomes_base}/Homo_sapiens/Ensembl/GRCh37/Sequence/BWAIndex/genome.fa"
      bowtie2     = "${params.igenomes_base}/Homo_sapiens/Ensembl/GRCh37/Sequence/Bowtie2Index/"
      star        = "${params.igenomes_base}/Homo_sapiens/Ensembl/GRCh37/Sequence/STARIndex/"
      bismark     = "${params.igenomes_base}/Homo_sapiens/Ensembl/GRCh37/Sequence/BismarkIndex/"
      gtf         = "${params.igenomes_base}/Homo_sapiens/Ensembl/GRCh37/Annotation/Genes/genes.gtf"
      bed12       = "${params.igenomes_base}/Homo_sapiens/Ensembl/GRCh37/Annotation/Genes/genes.bed"
      readme      = "${params.igenomes_base}/Homo_sapiens/Ensembl/GRCh37/Annotation/README.txt"
      mito_name   = "MT"
      macs_gsize  = "2.7e9"
      blacklist   = "${baseDir}/assets/blacklists/GRCh37-blacklist.bed"
    }
    'GRCm38' {
      fasta       = "${params.igenomes_base}/Mus_musculus/Ensembl/GRCm38/Sequence/WholeGenomeFasta/genome.fa"
      bwa         = "${params.igenomes_base}/Mus_musculus/Ensembl/GRCm38/Sequence/BWAIndex/genome.fa"
      bowtie2     = "${params.igenomes_base}/Mus_musculus/Ensembl/GRCm38/Sequence/Bowtie2Index/"
      star        = "${params.igenomes_base}/Mus_musculus/Ensembl/GRCm38/Sequence/STARIndex/"
      bismark     = "${params.igenomes_base}/Mus_musculus/Ensembl/GRCm38/Sequence/BismarkIndex/"
      gtf         = "${params.igenomes_base}/Mus_musculus/Ensembl/GRCm38/Annotation/Genes/genes.gtf"
      bed12       = "${params.igenomes_base}/Mus_musculus/Ensembl/GRCm38/Annotation/Genes/genes.bed"
      readme      = "${params.igenomes_base}/Mus_musculus/Ensembl/GRCm38/Annotation/README.txt"
      mito_name   = "MT"
      macs_gsize  = "1.87e9"
      blacklist   = "${baseDir}/assets/blacklists/GRCm38-blacklist.bed"
    }
    <TRUNCATED>
  }
}
```

AWS iGenomes
Common reference genomes hosted on AWS S3

amazon web services

refgenie

DSL2 has dropped!

🍏 Standardised, reproducible, self-contained, modular tool wrappers for re-use by the entire Nextflow community!

```
nf-core modules list                    ## List available nf-core/modules
```

```
nf-core modules install . fastqc        ## Install from nf-core/modules
```

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

nextflow.enable.dsl = 2
include { FASTQC } from './modules/nf-core/software/fastqc/main' addParams( options: [:] )

worflow {

  FASTQC(
    ch_reads
  )

}
```

nf-core/ 🍏
modules

nf-core/rnaseq

Pipeline Statistics

250	252	2100	45
Stars	Forks	Commits	Code contributors

Version history

2.0   6 days ago ▾
Released 12 Nov 2020 — `bc5fc76`   [Download .zip] [Download .tar.gz] [⬡ View release]

1.4.2	1 year ago ◂
1.4.1	1 year ago ◂
1.4	1 year ago ◂
1.3	2 years ago ◂
1.2	2 years ago ◂
1.1	2 years ago ◂
1.0	2 years ago ◂

↓ clones in last 1 year

24450

stars	watchers
250	65

last release	last updated
	13 hours ago

| open issues	pull requests
20	271

collaborators
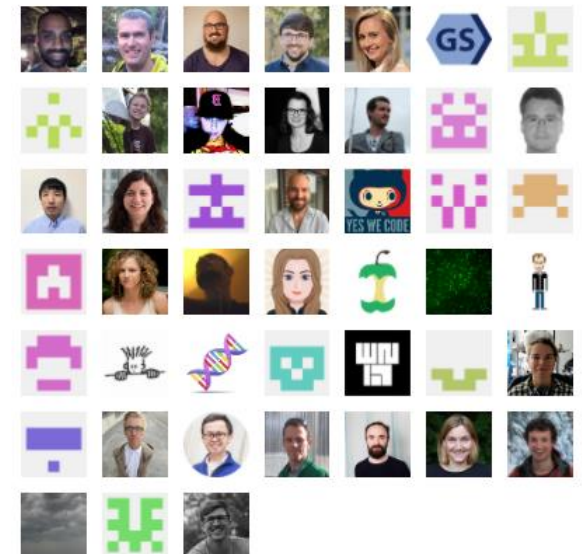
nf-core/rnaseq v2.0

- One process -> one container

- Exclusively using Biocontainers

- Retrieve metadata, download FastQ files, md5sum check and auto-create input samplesheet via SRA / ENA / GEO ids

```
nextflow run nf-core/rnaseq --public_data_ids ids.txt -profile docker
```

- Major new features:
  - Samplesheet input
  - RSEM / STAR support
  - UMI support
  - rRNA removal
  - Strand-specificity checks
  - Spike-in / transgene support

nf-core/rnaseq commits per week
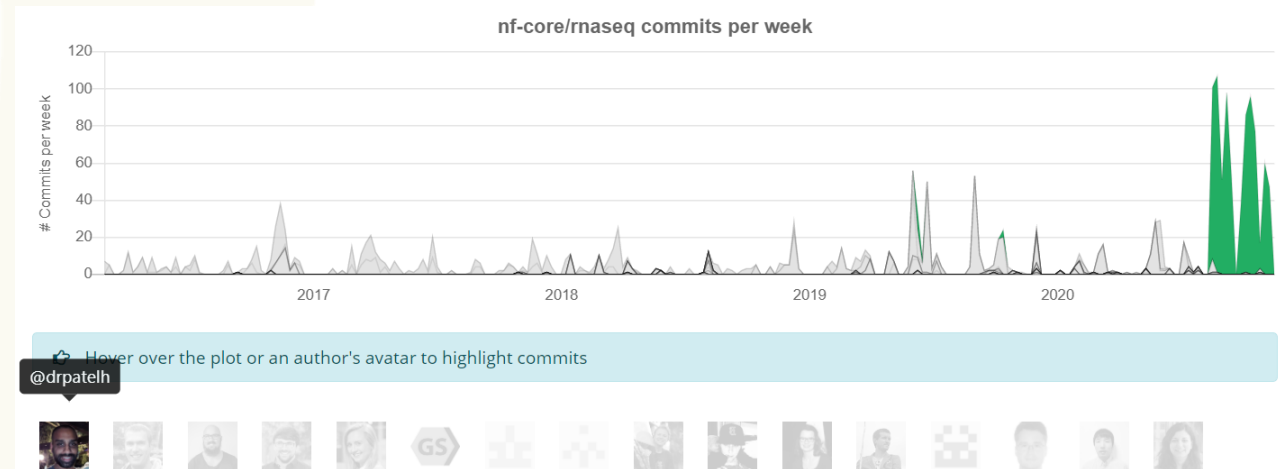
nf-core/chipseq

Pipeline Statistics

62	65	690	10
Stars	Forks	Commits	Code contributors

Version history

1.2.1 — 4 months ago
Released 29 Jul 2020 — `0f487ed`   [Download .zip] [Download .tar.gz] [⚙ View release]

1.2.0 — 5 months ago ◀

1.1.0 — 1 year ago ◀

1.0.0 — 1 year ago ◀

↓ clones in last 1 year
2433

stars	watchers
62	58

last release	last updated
	3 weeks ago

open issues	pull requests
6	105

collaborators

nf-core/atacseq

Pipeline Statistics

54	43	779	7
Stars	Forks	Commits	Code contributors

Version history

1.2.1   4 months ago ▼

Released 29 Jul 2020 — `1b3a832`   [Download .zip] [Download .tar.gz] [◯ View release]

1.2.0   5 months ago ◀

1.1.0   1 year ago ◀

1.0.0   2 years ago ◀

↓ clones in last 1 year

2993

stars	watchers
54	60
last release	last updated
	1 week ago
open issues	pull requests
9	73

collaborators

Usage

```
nextflow run nf-core/rnaseq      -profile test,docker -r 2.0   --input samplesheet.csv --genome GRCh37

nextflow run nf-core/chipseq     -profile test,docker -r 1.2.1 --input samplesheet.csv --genome GRCh37

nextflow run nf-core/atacseq     -profile test,docker -r 1.2.1 --input samplesheet.csv --genome GRCh37

nextflow run nf-core/viralrecon -profile test,docker -r 1.1.0 --input samplesheet.csv --genome 'MN908947.3'
```

```
module purge
module load Nextflow/19.10.0
module load Singularity/3.4.2

export NXF_SINGULARITY_CACHEDIR=~/nf-core/containers/

nextflow run nf-core/chipseq \
    --input samplesheet.csv \
    --genome GRCh37 \
    --single_end \
    --narrow_peak \
    --email harshil.patel@crick.ac.uk \
    -profile crick \
    -r 1.1.0
```

Communication

Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg JA, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S.
The nf-core framework for community-curated bioinformatics pipelines.
Nat Biotechnol. 2020 Feb 13. doi: 10.1038/s41587-020-0439-x.

Thank you!

nf-core

nextflow

harshil.patel@crick.ac.uk

drpatelh

aka_hpatel

THE FRANCIS CRICK INSTITUTE