seqeralabs

Phil Ewels
phil@seqera.io

Reproducible bioinformatics for everyone:

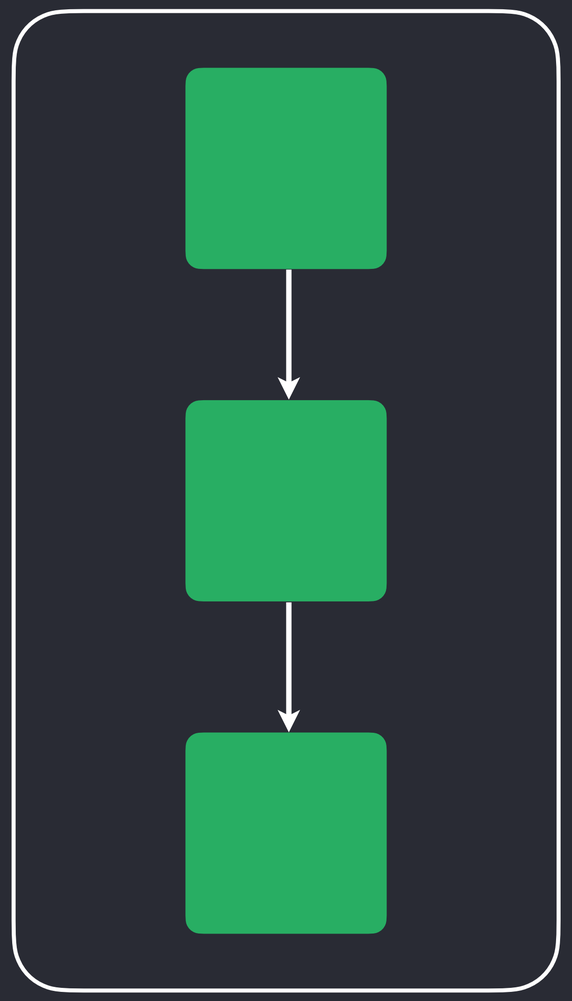
Nextflow & nf-core


nextflow

nextflow

Language

nextflow



Processes

Channels

Workflows

nextflow

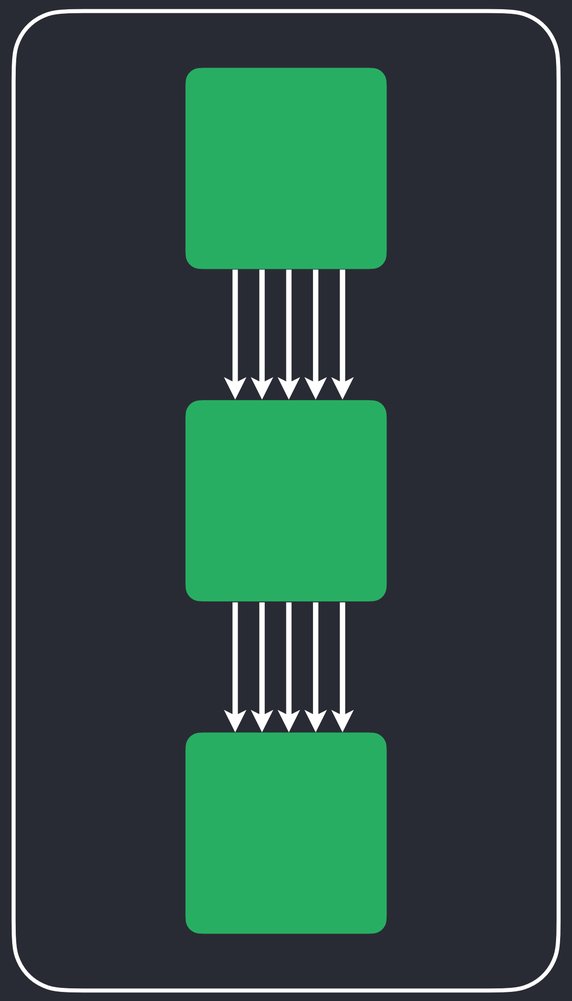
```
#!/usr/bin/env nextflow
process fastqc {
  input:
  path input

  output:
  path "*_fastqc.{zip,html}"

  script:
  """
  fastqc -q $input
  """
}

workflow {
  Channel.fromPath("*.fastq.gz") | fastqc
}
```

nextflow



Parallelisation

Reentrancy

Reusability

nextflow

Language

Software

Compute

nextflow

Language

Singularity

Compute

nextflow

Language

docker

Singularity

CONDA

SGE

Microsoft Azure

slurm
workload manager

openstack

aws

LSF

PBS

Google Cloud

kubernetes

nextflow

git

GitHub

Bitbucket

GitLab

Gitea

AWS CodeCommit

Azure Repos

docker

Singularity

CONDA

SGE

Microsoft Azure

slurm workload manager

openstack

aws

LSF

PBS

Google Cloud

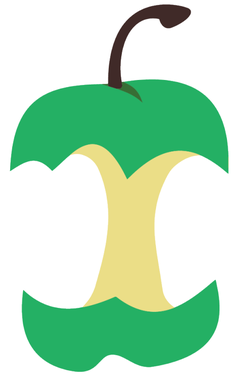
kubernetes

nextflow

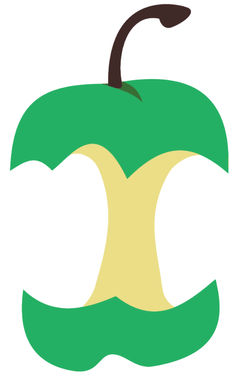
Reproducible

Portable

nf-core



nf-core



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

https://nf-co.re

nf-core

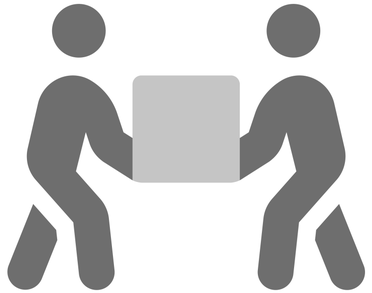
Guidelines

Tools

Pipelines

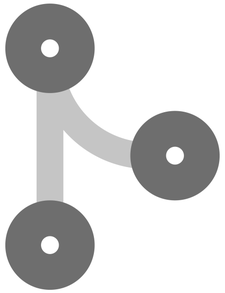
https://nf-co.re

nf-core



Develop <u>with</u> the community

Use a common template



Collaborate, don’t duplicate

https://nf-co.re

nf-core

37

RELEASED

24

UNDER DEVELOPMENT

6

ARCHIVED

https://nf-co.re

Available Pipelines

Can you think of another pipeline that would fit in well? Let us know!

Search keywords | Filter: Released 25 | Under development 14 | Archived 3 | Sort: Last Release | Alphabetical | Stars | Display:

nf-core/viralrecon

☆ 18

amplicon assembly covid-19 covid19 illumina metagenomics sars-cov-2 variant-calling viral virus

Assembly and intrahost/low-frequency variant calling for viral samples

Version **1.1.0** Published 7 days ago

nf-core/sarek

☆ 49

cancer germline somatic variant-calling

Analysis pipeline to detect germline or somatic variants from WGS / targeted sequencing

Version **2.6.1** Published 1 week ago

nf-core/imcyto

☆ 5

cytometry image-analysis image-processing image-segmentation

Image Mass Cytometry analysis pipeline

Version **1.0.0** Published 1 month ago

nf-core/slamseq

☆ 3

differential-expression quantseq slamseq transcriptomics

SLAMSeq processing and analysis pipeline

Version **1.0.0** Published 2 months ago

nf-core/coproid

☆ 2

adna ancient-dna coprolite microbiome

Coprolite host Identification pipeline

nf-core/mhcquant

☆ 12

mass-spectrometry mhc peptides

Identify and quantify MHC eluted peptides from mass spectrometry raw

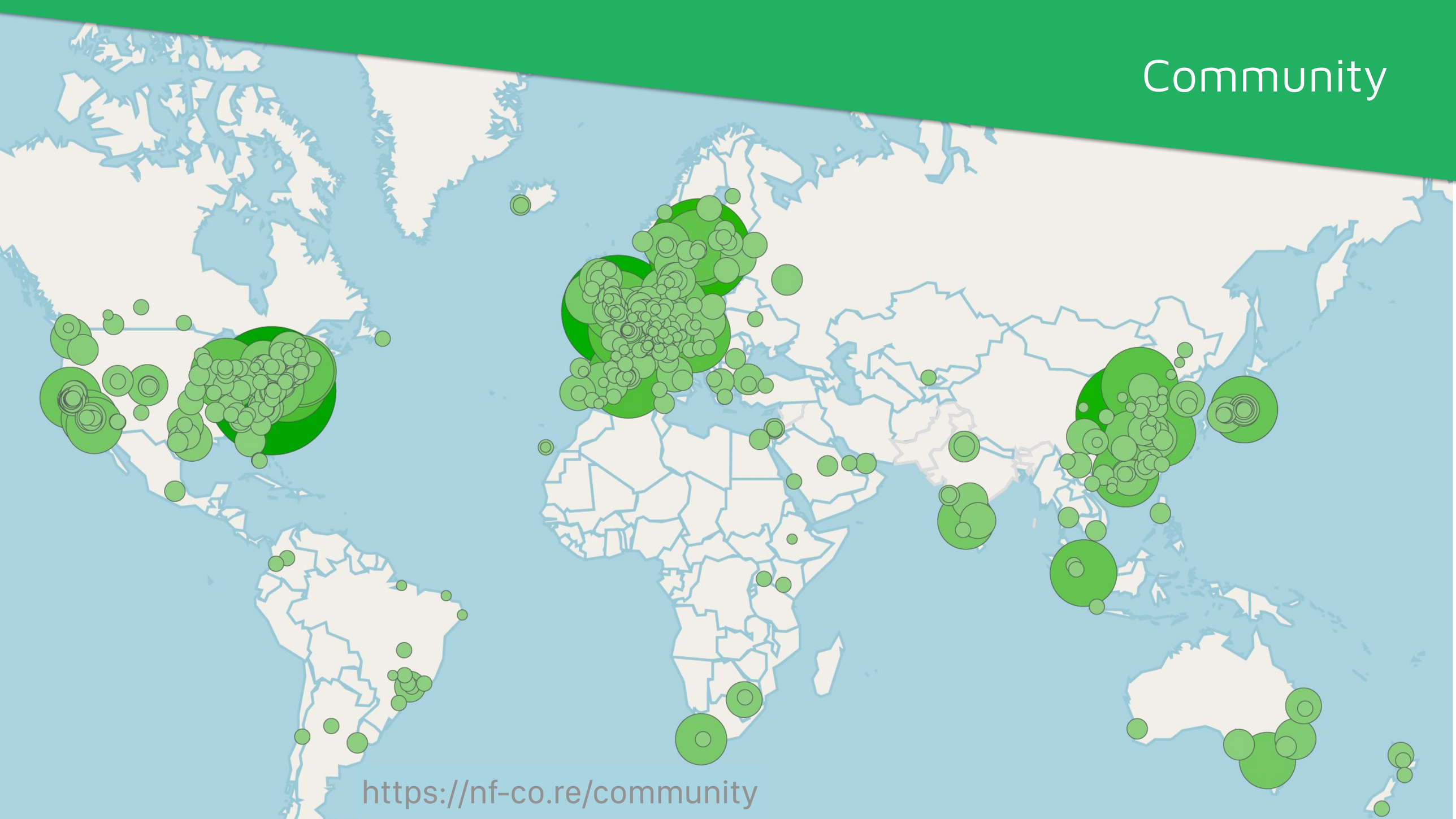
Community

https://nf-co.re/stats

3067	440	1240	2825
Slack users	GitHub organisation members	GitHub contributors	Twitter followers

82	9.39K	39.79K	4.2K
Repositories	Pull Requests	Commits	Issues

Community



https://nf-co.re/community

nature biotechnology

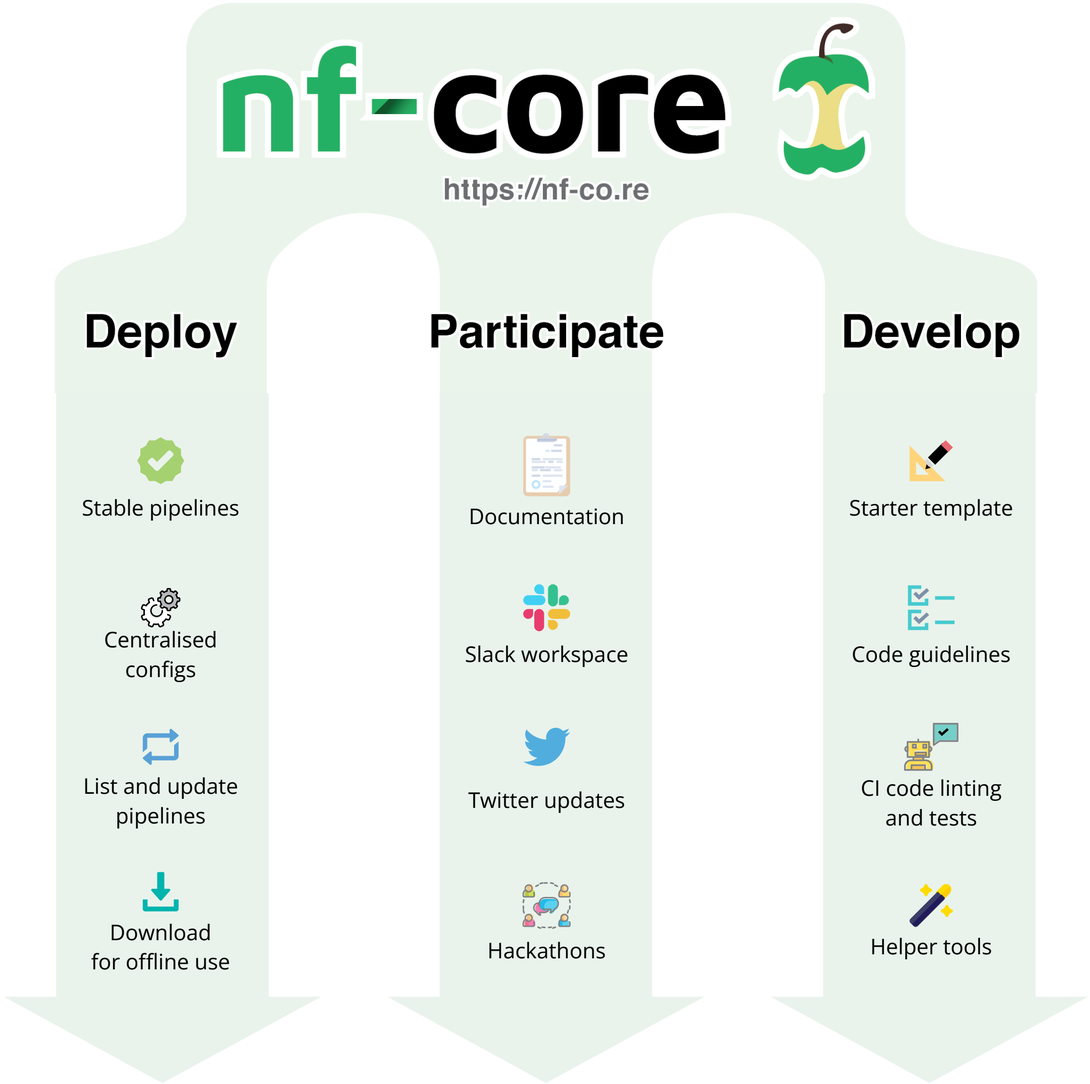
Correspondence | Published: 13 February 2020

The nf-core framework for community-curated bioinformatics pipelines

Philip A. Ewels, Alexander Peltzer, Sven Fillinger, Harshil Patel, Johannes Alneberg, Andreas Wilm, Maxime Ulysse Garcia, Paolo Di Tommaso & Sven Nahnsen

Nature Biotechnology **38**, 276–278(2020) | Cite this article

To the Editor — The standardization, portability and reproducibility of analysis pipelines are key issues within the bioinformatics community. Most bioinformatics pipelines are designed for use on-premises; as a result, the associated software dependencies and execution logic are likely to be tightly coupled with proprietary computing environments. This can make it difficult or even impossible for others to reproduce the ensuing results, which is a fundamental requirement for the validation of scientific findings. Here, we introduce the nf-core framework as a means for the development of collaborative, peer-reviewed, best-practice analysis pipelines (Fig. 1). All nf-core pipelines are written in Nextflow and so inherit the ability to be executed on most computational infrastructures, as well as having native support for container technologies such as Docker and Singularity. The nf-core community (Supplementary Fig. 1) has developed a suite of tools that automate pipeline creation, testing, deployment and synchronization. Our goal is to provide a framework for high-quality bioinformatics pipelines that can be used across all institutions and research facilities.



Documentation

https://nf-co.re/rnaseq

Introduction | Results | Usage docs | Parameter docs | Output docs | Releases & Statistics | 3.4

Parameters

Input/output options

Define where the pipeline should find input data and save output data.

`--input`	Path to comma-separated file containing information about the samples in the experiment.	Help
`--outdir`	Path to the output directory where the results will be saved.	`default: './results'`
`--email`	Email address for completion summary.	Help
`--multiqc_title`	MultiQC report title. Printed as page header, used for filename if not otherwise specified.	
`--save_merged_fastq`	Save FastQ files after merging re-sequenced libraries in the results directory.	

UMI options

Options for processing reads with unique molecular identifiers

On this page

Parameters

- Input/output options
- UMI options
- Read filtering options
- Reference genome options
- Read trimming options
- Alignment options
- Process skipping options

Show all help | Show hidden params

Back to top

Launch wizard

https://nf-co.re/launch

Read trimming options

Launch

--save_trimmed True False

Save the trimmed FastQ files in the results directory.

Alignment options

Options to adjust parameters and filtering criteria for read alignments.

--aligner star_salmon

Specifies the alignment algorithm to use - available options are 'star_salmon', 'star_rsem' and 'hisat2'.

--pseudo_aligner [Select an option]

Specifies the pseudo aligner to use - available options are 'salmon'. Runs in addition to '--aligner'.

--bam_csi_index True False

Create a CSI index for BAM files instead of the traditional BAI index. This will be required for genomes with larger chromosome sizes.

--star_ignore_sjdbgtf True False

When using pre-built STAR indices do not re-extract and use splice junctions from the GTF file.

--salmon_quant_libtype

On this page

Show hidden params

Back to top

Launch wizard

https://nf-co.re/launch

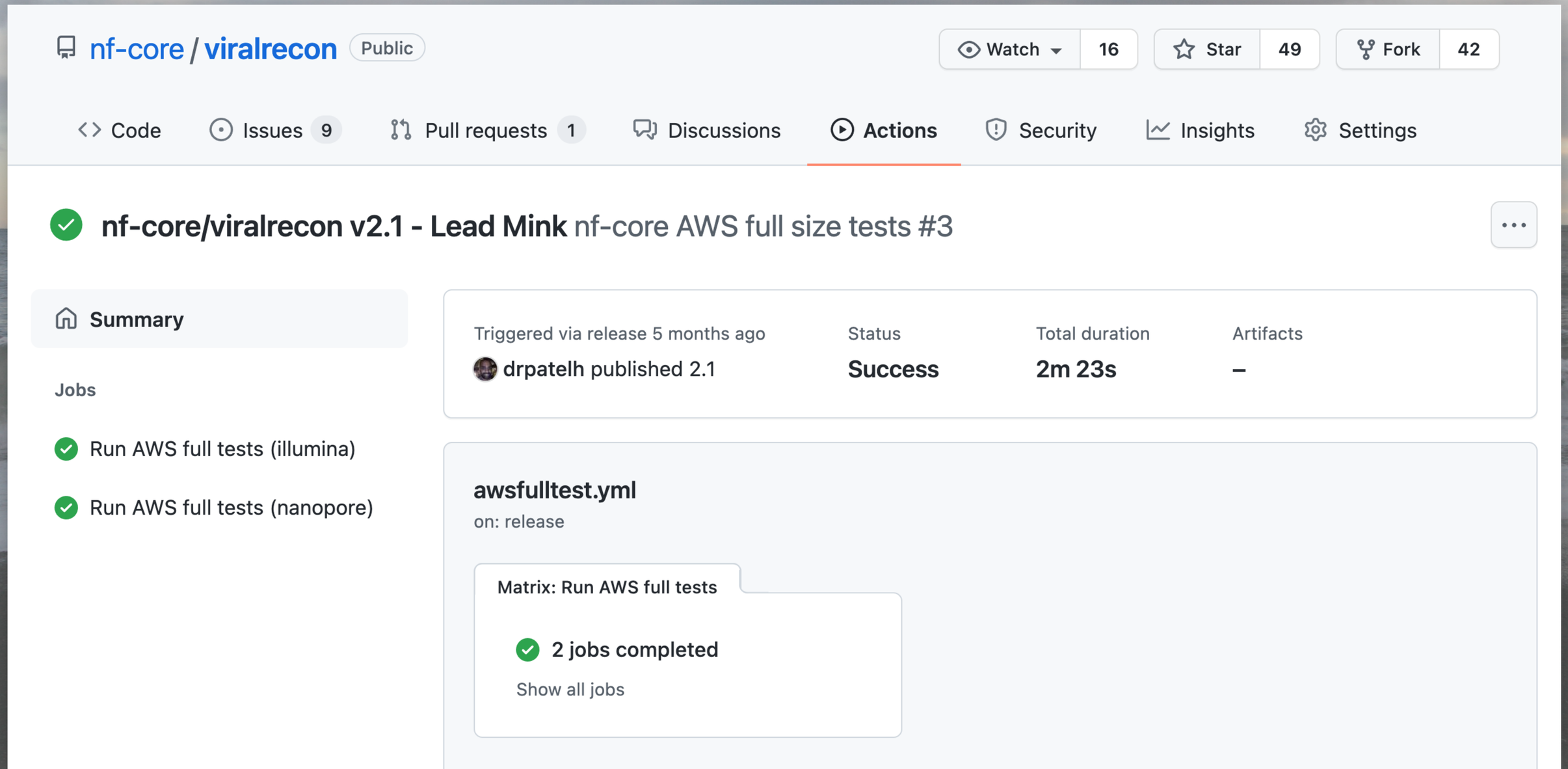
```
nf-core launch --id 1637063024_d92e0f1632c2
```

```
nextflow run nf-core/rnaseq -params-file nf-params.json
```

Running in the cloud

Running in the cloud

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



Running in the cloud

https://nf-co.re/viralrecon

Introduction | aws Results | Usage docs | Parameter docs | Output docs | Releases & Statistics | 2.2

aws nf-core-awsmegatests / viralrecon
/ results-2ebae61442598302c64916bd5127cf23c8ab5611 / platform_illumina

Copy Bucket S3 URL

Name	Last Modified	Size
..		
assembly/		
fastp/		
fastqc/		
kraken2/		
multiqc/		
pipeline_info/		
variants/		

>_ command

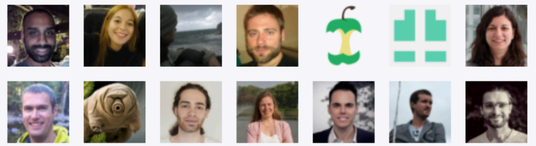
```
nextflow run nf-core/viralre
```

↓ clones in last 2 years

13054

stars	watchers
49	16
last release	last updated
	1 month ago
open issues	pull requests
9	168

collaborators



get in touch

Ask a question on Slack

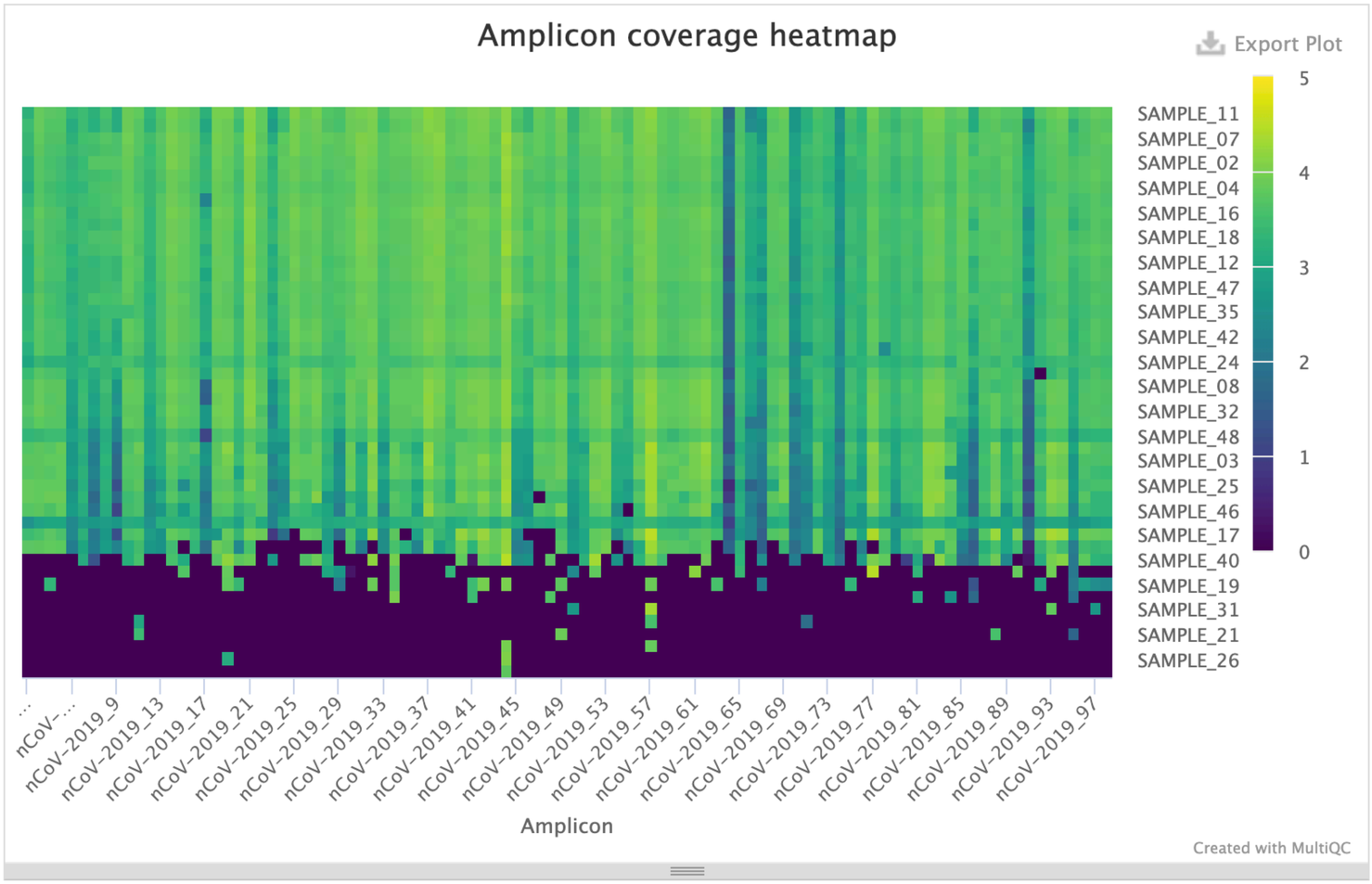
Open an issue on GitHub

Running in the cloud

https://nf-co.re/viralrecon

multiqc_report.html

Download file | Copy URL | Copy S3 URL

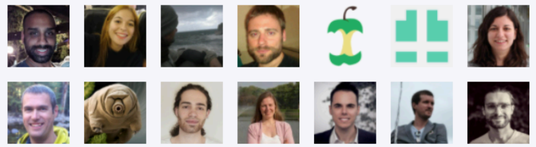


Toolbox

open issues 9

pull requests 168

collaborators



get in touch

Ask a question on Slack

Open an issue on GitHub

nf-core/modules

Modular design gives clearer pipeline code

Proper unit testing of individual steps in each pipeline

Different pipelines can reuse tool wrappers and software images

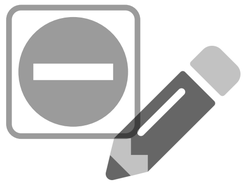
Library of tool wrappers makes building a new pipeline fast

nf-core/modules

```
$ nf-core modules list
```

```
$ nf-core modules install fastqc
```

Copies module files in to pipeline



CI tests check that module files have not been edited

Metadata file tracks git hash of modules repo

Update command fetches latest versions

nf-core/ modules

Nextflow DSL2

```
⌥⌘1
❯ nf-core modules install bwa/aln

         ___     __   __   __   ___
   |\ | |__  __ /  ` /  \ |__) |__         }  {
   | \| |       \__, \__/ |  \ |___     \`-._,-`-,
                                          `._,._,'

    nf-core/tools version 2.5.dev0 - https://nf-co.re


INFO     Installing 'bwa/aln'
INFO     Downloaded 2 files to ./modules/nf-core/modules/bwa/aln
INFO     Include statement: include { BWA_ALN } from '../modules/nf-core/modules/bwa/aln/
```

nf-core/ modules

Nextflow DSL2

⌥⌘1

Module: bwa/aln

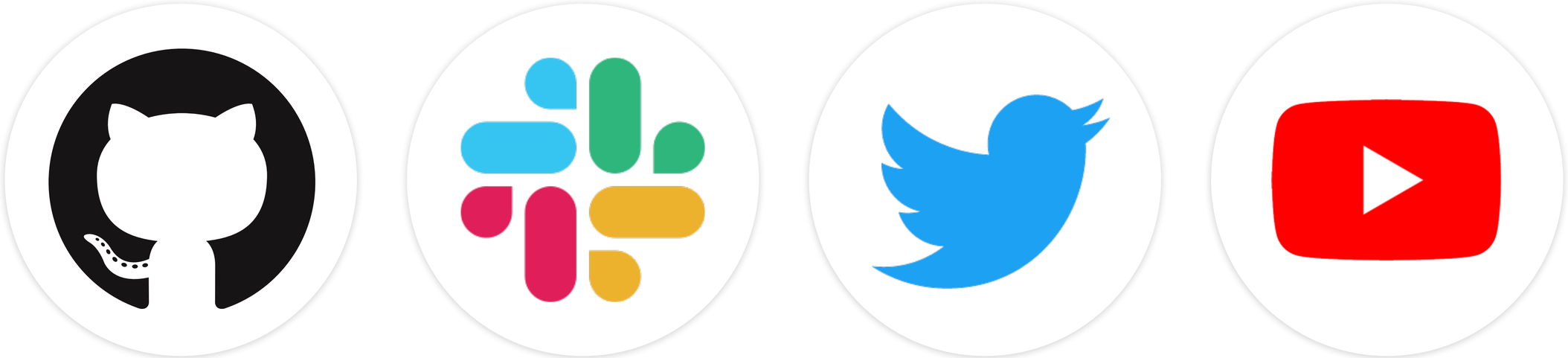
Location: ./modules/nf-core/modules/bwa/aln

🔧 Tools: bwa

📖 Description: Find SA coordinates of the input reads for bwa short-read mapping

📥 Inputs	Description	Pattern
meta *(map)*	Groovy Map containing sample information e.g. [id:'test', single_end:false]	
reads *(file)*	List of input FastQ files of size 1 and 2 for single-end and paired-end data, respectively.	
index *(file)*	BWA genome index files	Directory containing BWA index *.{amb,ann,bwt,pac,sa}

Join the community

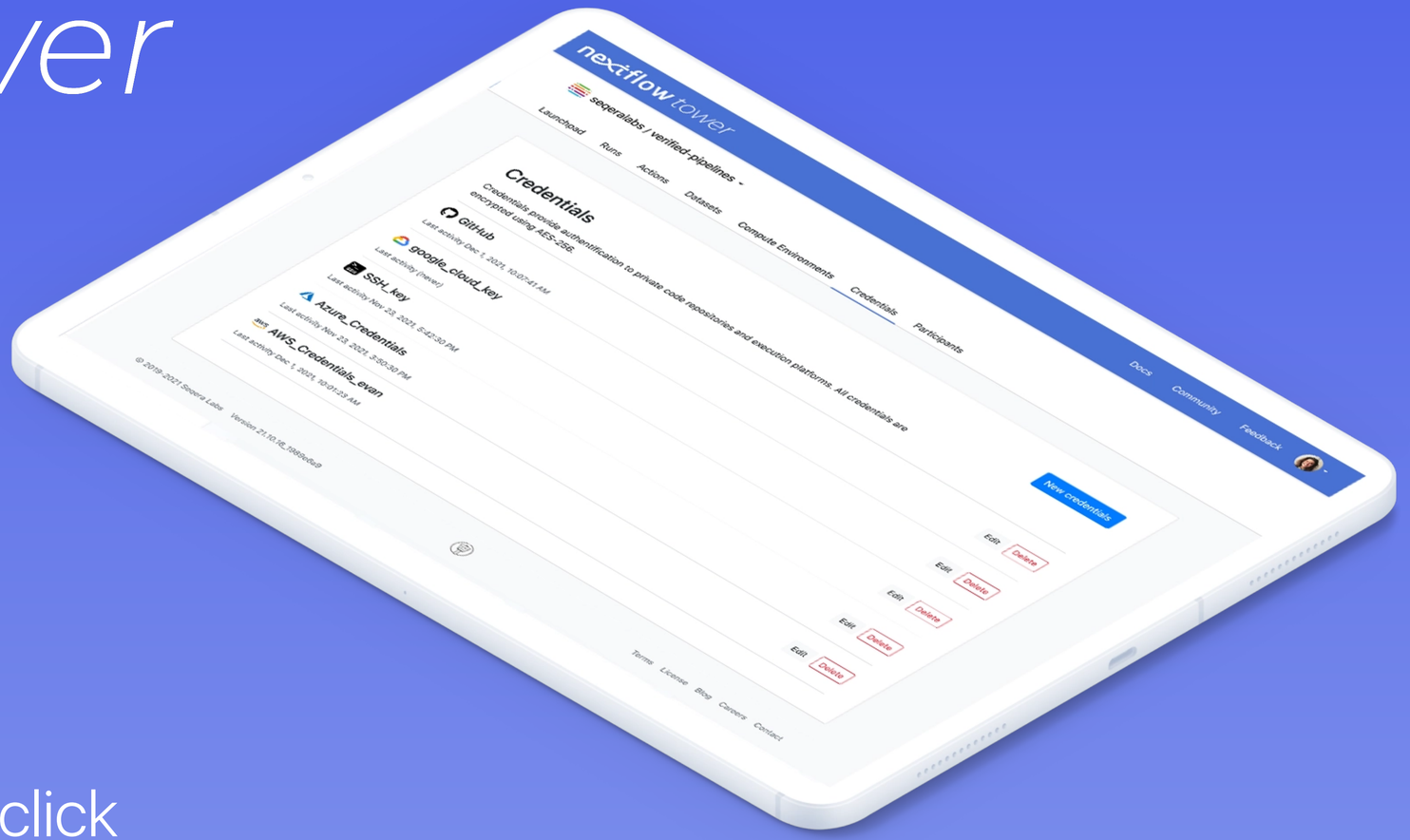


https://nf-co.re/join

seqeralabs

https://tower.nf

nextflow *tower*



Intuitive launchpad interface

Launch, manage, and monitor

Share runs and work in teams

Create cloud infrastructure with a click

Phil Ewels

https://phil.ewels.co.uk
phil@seqera.io

seqeralabs
https://seqera.io

Chan Zuckerberg Initiative

nextflow

nf-core

nextflow *tower*

https://nf-co.re/join