



Phil Ewels, PhD
phil.ewels@seqera.io



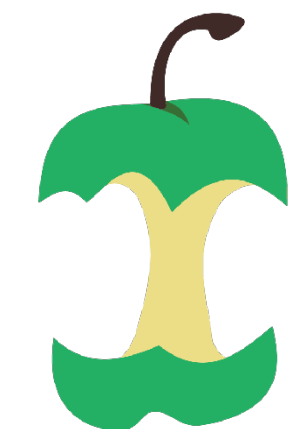
Funded by **Chan Zuckerberg Initiative** 

Experiences with cloud compute in the nf-core open source community

nextflow



nf-core



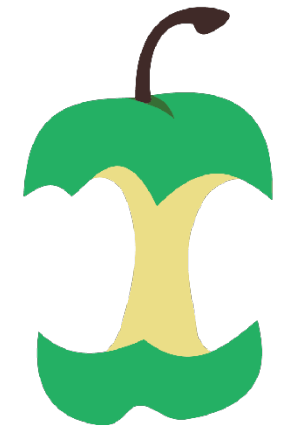
Strategies for **Containing Cloud Computing Costs** in Healthcare and Life Sciences Research

December 15th 2022

nextflow



nf-core



Key points:

- Test, test, test
- Know thy buckets
- Optimising resource requests

Test, test, test

 Mistakes with big data can be expensive

 Make them with small tests instead

 Better still: **let someone else make them instead**

nf-core

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

<https://nf-co.re>

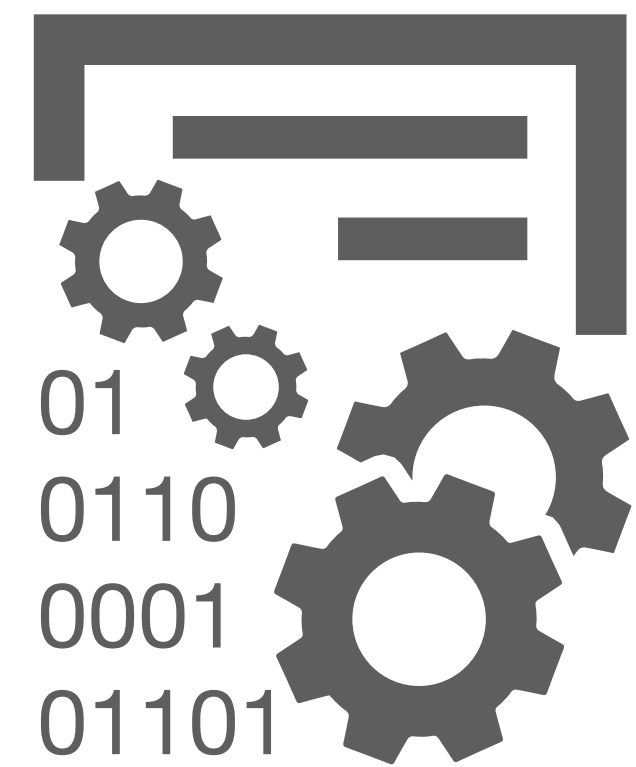
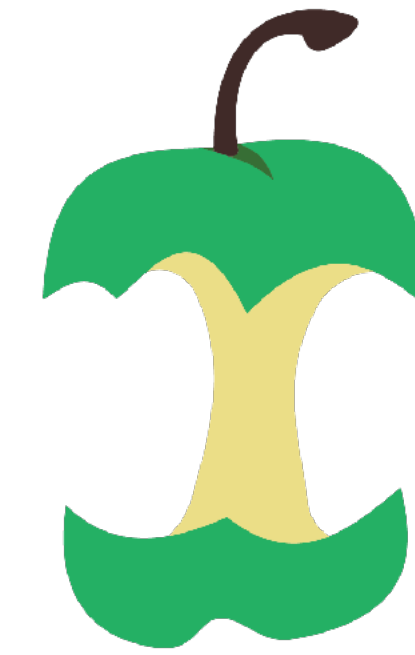
nf-core



73
PIPELINES

<https://nf-co.re>

nf-core



752

MODULES

29

SUB-WORKFLOWS

<https://nf-co.re>

4305

Slack users

544

GitHub organisation
members

1632

GitHub contributors

3383

Twitter followers

90

Repositories

12.1K

Pull Requests

42.45K

Commits

5.46K

Issues

<https://nf-co.re>

Cloud testing



nf-core / viralrecon Public

Watch 16

Star 49

Fork 42

Code

Issues 9

Pull requests 1

Discussions

Actions

Security

Insights

Settings

nf-core/viralrecon v2.1 - Lead Mink nf-core AWS full size tests #3

Summary

Jobs

Run AWS full tests (illumina)

Run AWS full tests (nanopore)

Triggered via release 5 months ago

drpatelh published 2.1

Status

Success

Total duration

2m 23s

Artifacts

-

awsfulltest.yml

on: release

Matrix: Run AWS full tests

2 jobs completed

Cloud testing



→ 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

>_ command

```
nextflow run nf-core/viralrecon
```

↓ clones in last 2 years

13054



stars

watchers

49

16

last release

last updated

1 month ago

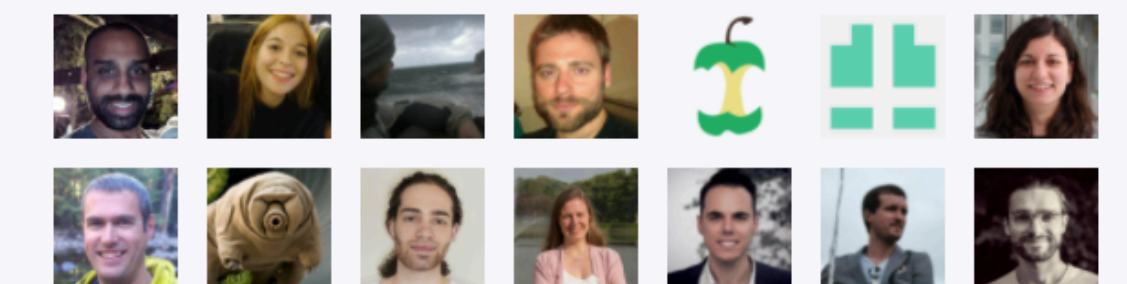
open issues

pull requests

9

168

collaborators



get in touch

Ask a question on Slack

Name	↑↓	Last Modified ↑↓	Size ↑↓
..			
_assembly/			
_fastp/			
_fastqc/			
_kraken2/			
_multiqc/			
_pipeline_info/			

Cloud testing



multiqc_report.html

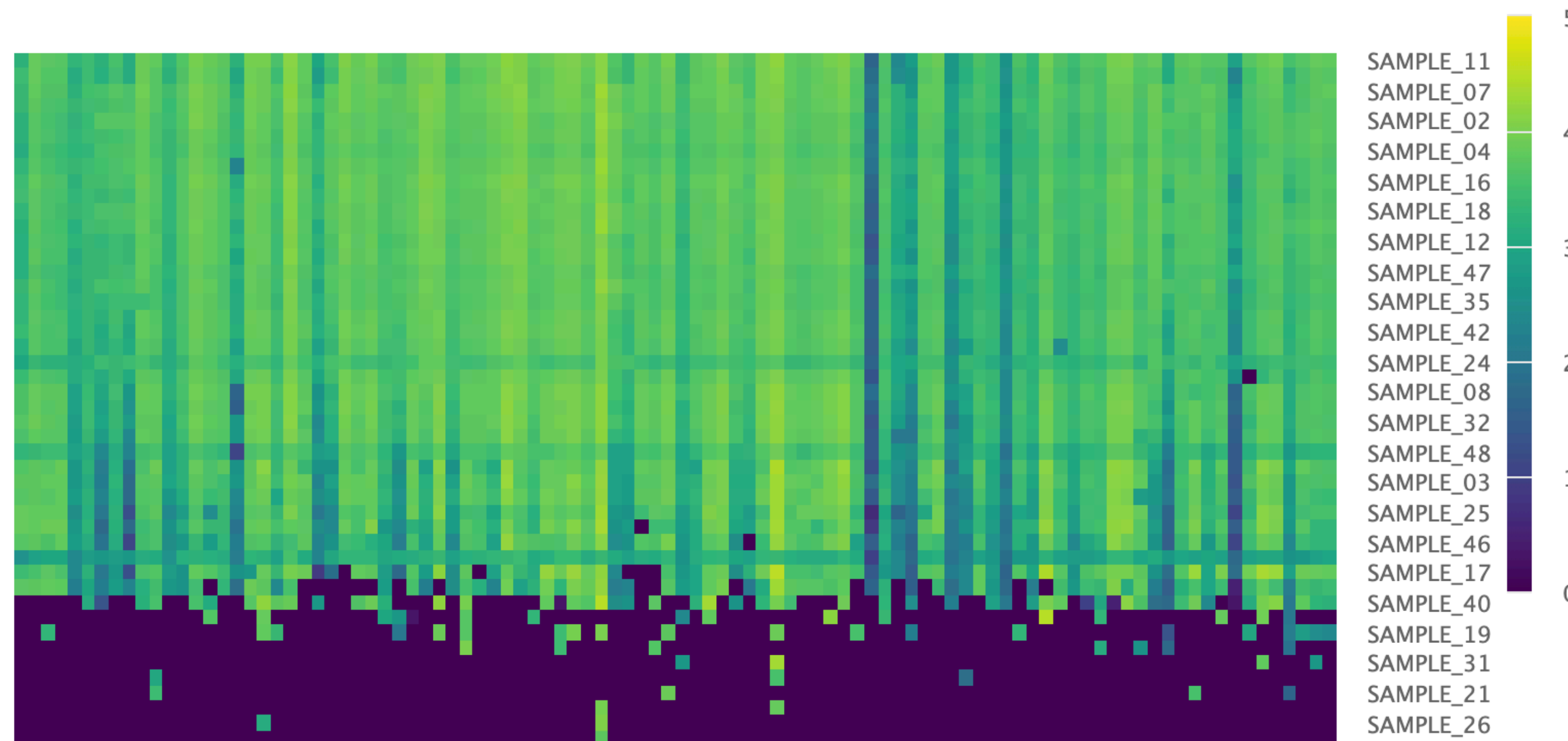
Download file

Copy URL

Copy S3 URL

Amplicon coverage heatmap

Export Plot



Toolbox



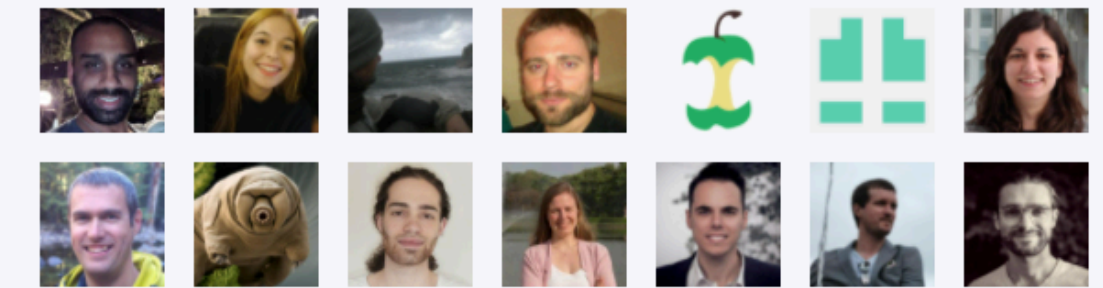
open issues

9

pull requests

168

collaborators



get in touch

Ask a question on Slack

Open an issue on GitHub

Cloud testing

Every pipeline, every release:



Know thy buckets

Know thy buckets



Input data



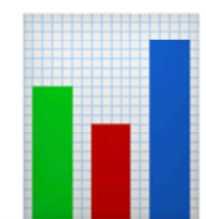
Private subnets



Work directories



Lifecycle rules



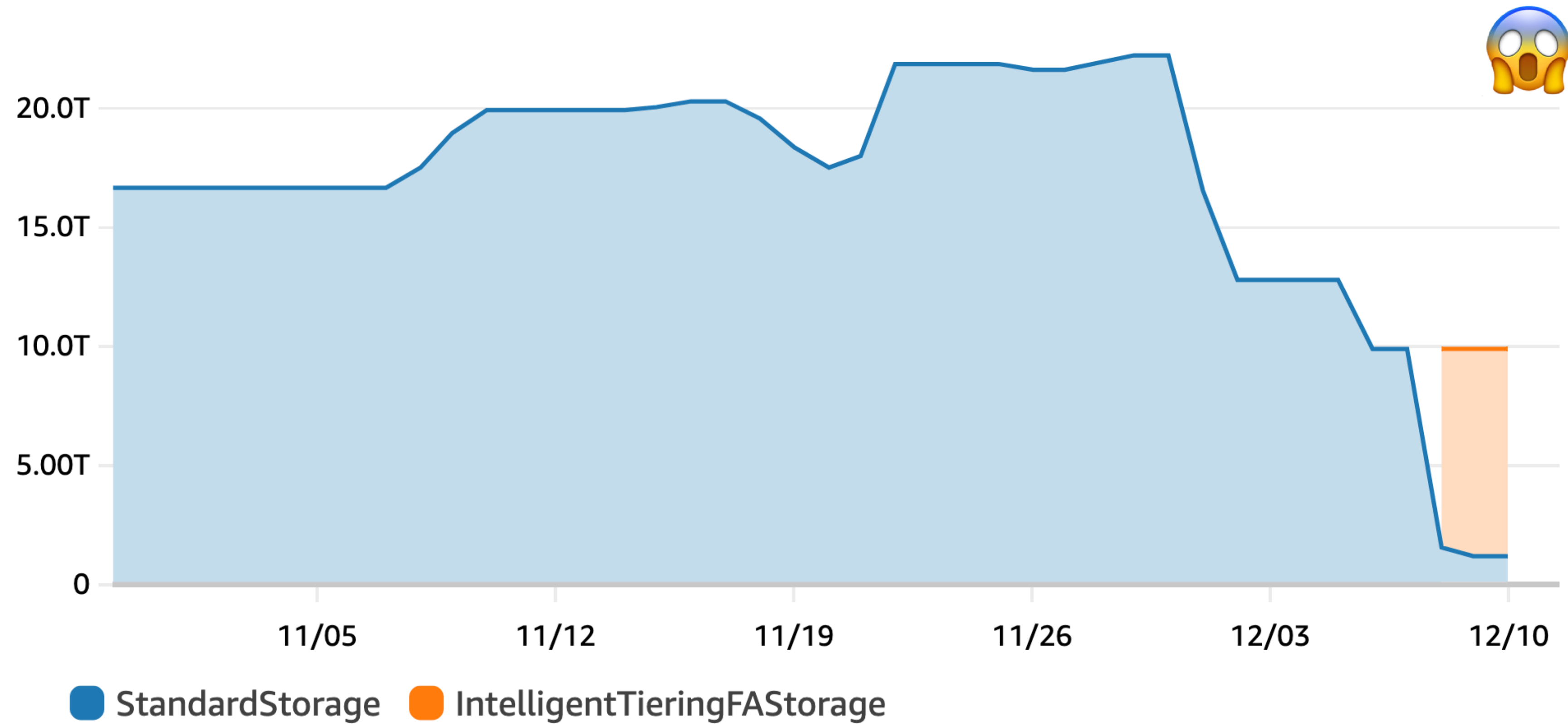
Results



Intelligent tiering



Know thy buckets



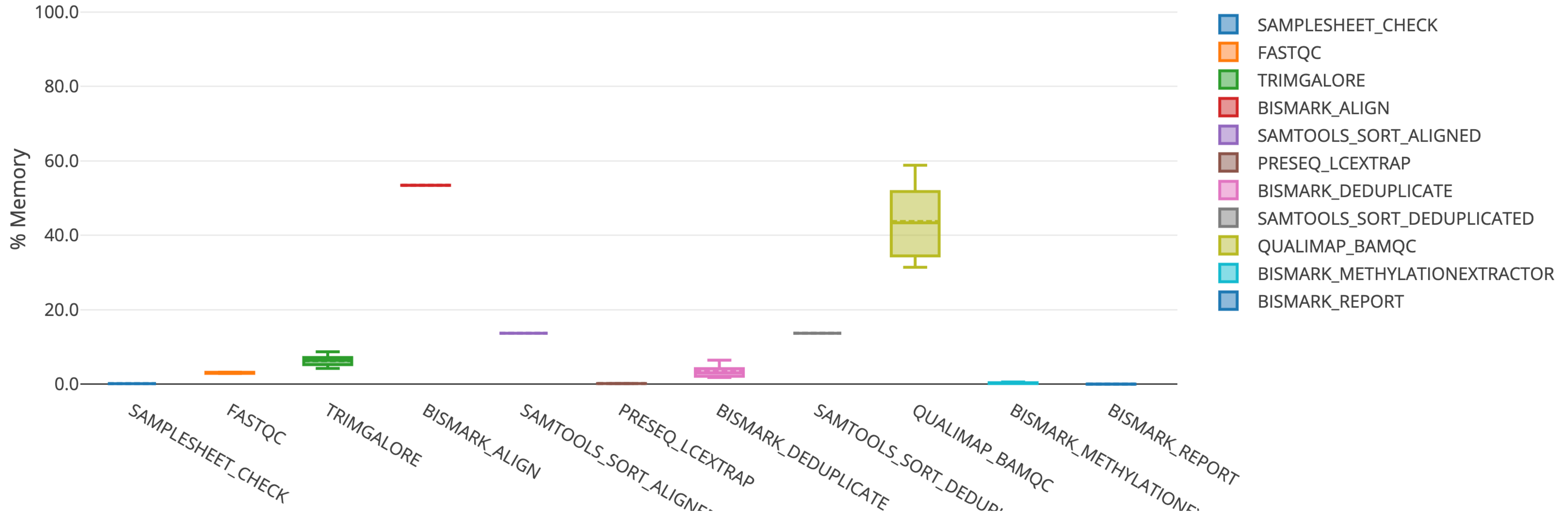
Optimising resource requests

Optimising resource requests

nf-core pipelines are built for everyone's data. Not *your* data.

Optimising resource requests

% Requested Physical Memory Used



Optimising resource requests

```
process {  
    withName: 'NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:STAR_ALIGN' {  
        memory = 100.GB  
    }  
}
```

< View Workflow Run

 **curious_sanger** nf-core/rnaseq

azure-batch

spot-instances



Command line

Parameters

Configuration

Datasets

Execution log

Reports

Optimization available

Re-launch 

```
nextflow run 'https://github.com/nf-core/rnaseq'  
  -name curious_sanger  
  -params-file 'https://api.tower.nf/ephemeral/Qztjwy1QcxfXT-ZzdKD-HQ.yaml'  
  -with-tower  
  -profile test
```



< View Workflow Run

✔ curious_sanger nf-core/rnaseq

azure-batch spot-instances 



Command line

Parameters

Configuration

Datasets

Execution log

Reports

Launch 

Optimization available

```
nextflow run 'https://github.com/nf-core/rnaseq'  
  -name curious_sanger  
  -params-file 'https://api.tower.nf/ephemeral/Qztjwy1QcxfXT-ZzdK'  
  -with-tower  
  -profile test
```



View optimized configuration



```
process {
  withName: 'NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:BAM_SORT_SAMTOOLS:BAM_STATS_SAMTOOLS:SAMTOOLS_FLAGSTA'
    cpus = 2
    memory = 1.GB
  }
  withName: 'NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:BAM_SORT_SAMTOOLS:BAM_STATS_SAMTOOLS:SAMTOOLS_IDXSTAT'
    cpus = 2
    memory = 1.GB
  }
  withName: 'NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:BAM_SORT_SAMTOOLS:BAM_STATS_SAMTOOLS:SAMTOOLS_STATS'
    cpus = 2
    memory = 1.GB
  }
  withName: 'NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:BAM_SORT_SAMTOOLS:SAMTOOLS_INDEX' {
    cpus = 2
    memory = 1.GB
  }
  withName: 'NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:BAM_SORT_SAMTOOLS:SAMTOOLS_SORT' {
    cpus = 3
    memory = 1.GB
  }
  withName: 'NFCORE_RNASEQ:RNASEQ:ALIGN_STAR:STAR_ALIGN' {
    cpus = 2
```


nextflow SUMMIT 2022

<https://summit.nextflow.io>

Talks from all major
cloud providers on
YouTube



Google Cloud



Angel Pizarro

Amazon Web Services

Price/performance of different cloud storage options for Nextflow workflows

 SUMMIT 2022



Venkat Malladi

Microsoft

Deep dive into Nextflow on Azure

 SUMMIT 2022



Hatem Nawar

Google

Pipeline economics on Cloud

 SUMMIT 2022

Phil Ewels

<https://phil.ewels.co.uk>

phil@seqera.io



seqeralabs

<https://seqera.io>

**Chan Zuckerberg
Initiative**



nextflow

nf-core 

nextflow *tower*

Nextflow / nf-core training

6-10 March 2023

nf-core hackathon

20-24 March 2023

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