



Phil Ewels, PhD

phil.ewels@seqera.io

 seqera labs

Funded by

Chan Zuckerberg
Initiative



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



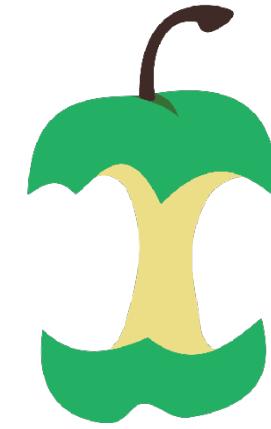
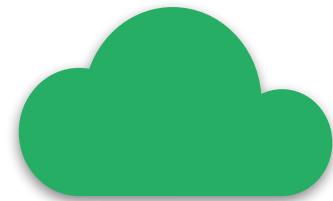






Strategies for **Containing Cloud Computing
Costs** in Healthcare and Life Sciences Research

December 15th 2022



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



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

<https://nf-co.re>

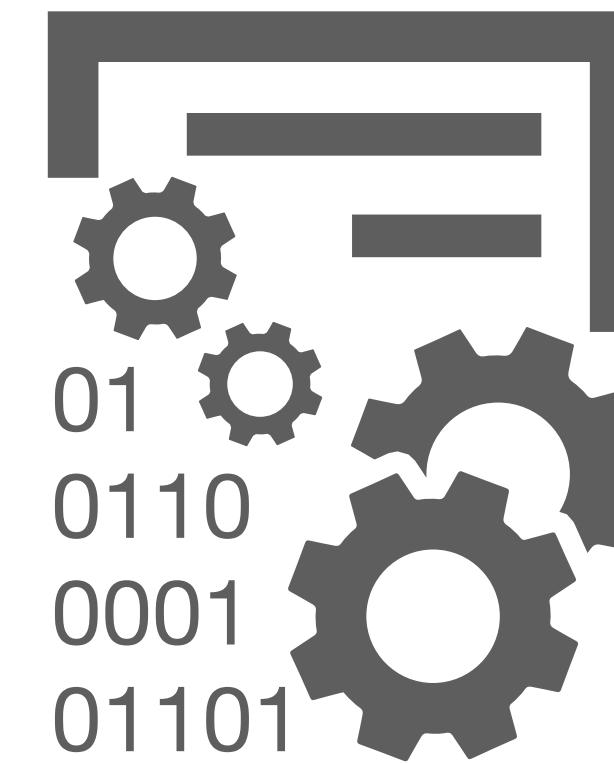
nf-core



73

PIPELINES

<https://nf-co.re>



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

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	Triggered via release 5 months ago	Status	Total duration	Artifacts
Summary	drpatelh published 2.1	Success	2m 23s	-

Jobs

- Run AWS full tests (illumina)
- Run AWS full tests (nanopore)

awsfulltest.yml
on: release

Matrix: Run AWS full tests

2 jobs completed

Cloud testing



aws nf-core-awsmegatests / viralrecon
/ [results-2ebae61442598302c64916bd5127cf23c8ab5611](#) / platform_illumina

[Copy Bucket S3 URL](#)

> command
`nextflow run nf-core/viralre`

↓ clones in last 2 years
13054

stars
49

watchers
16

last release
1 month ago

last updated

open issues
9

pull requests
168

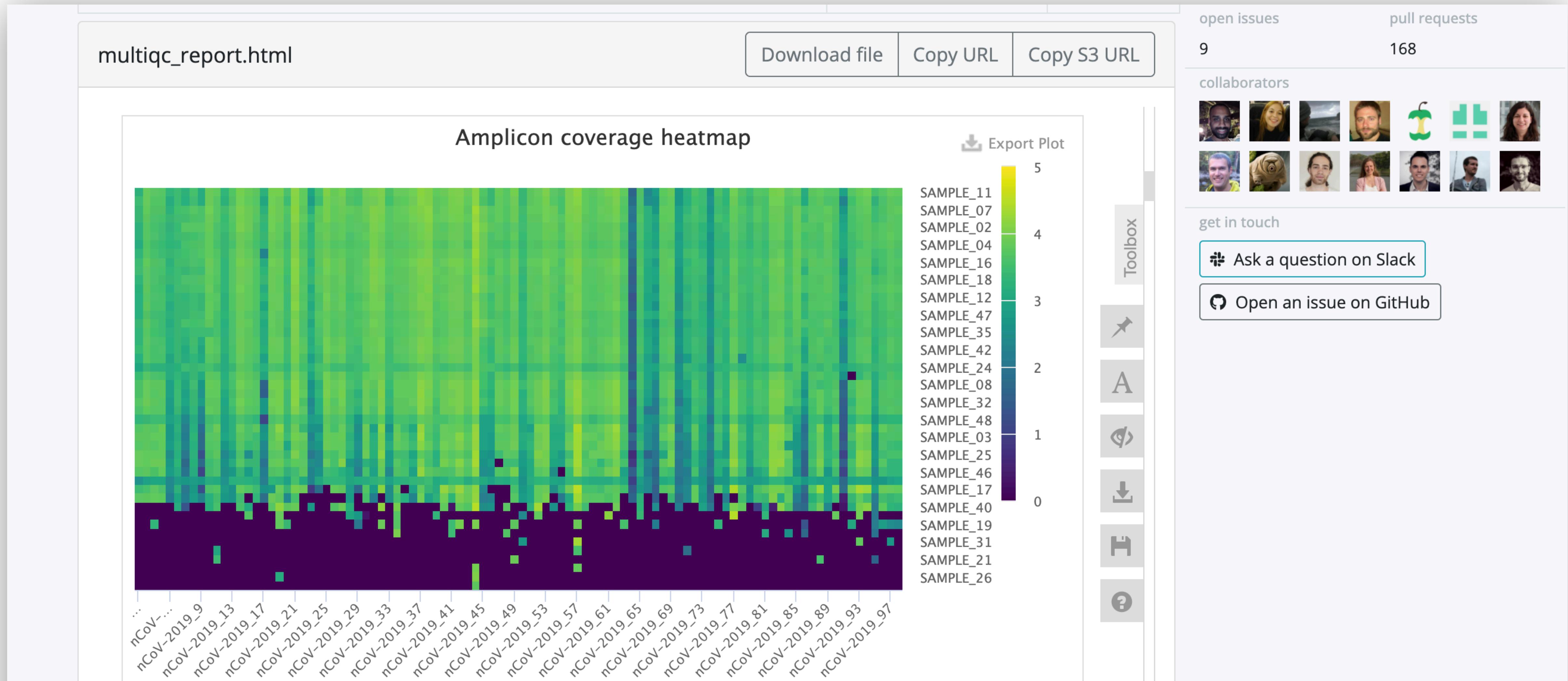
collaborators

get in touch

Ask a question on Slack

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

Cloud testing

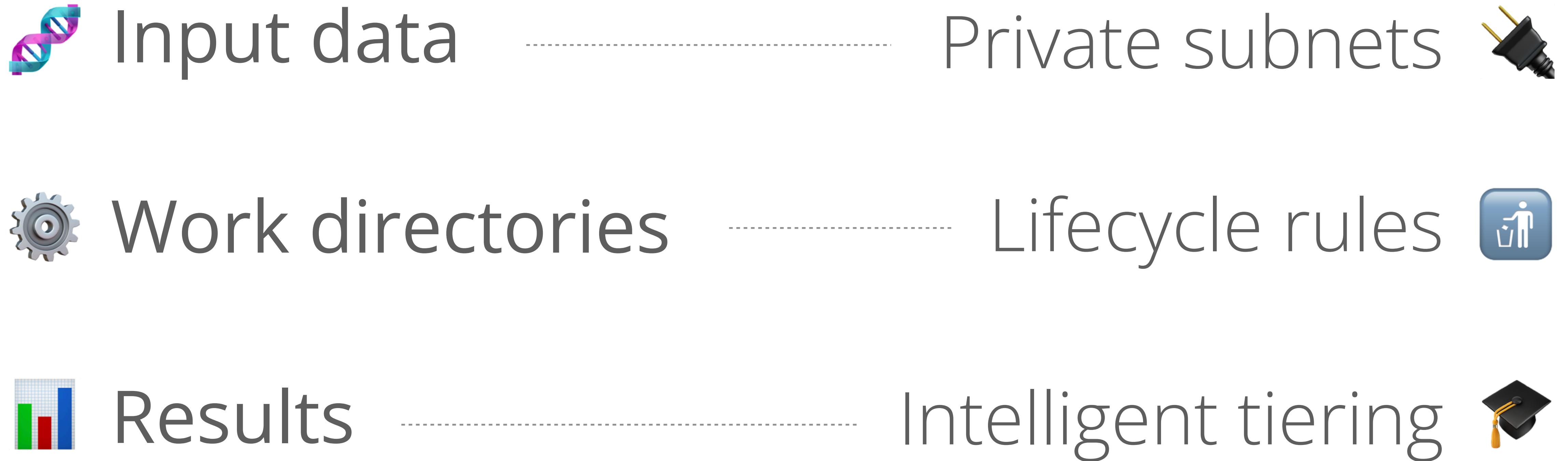


Every pipeline, every release:

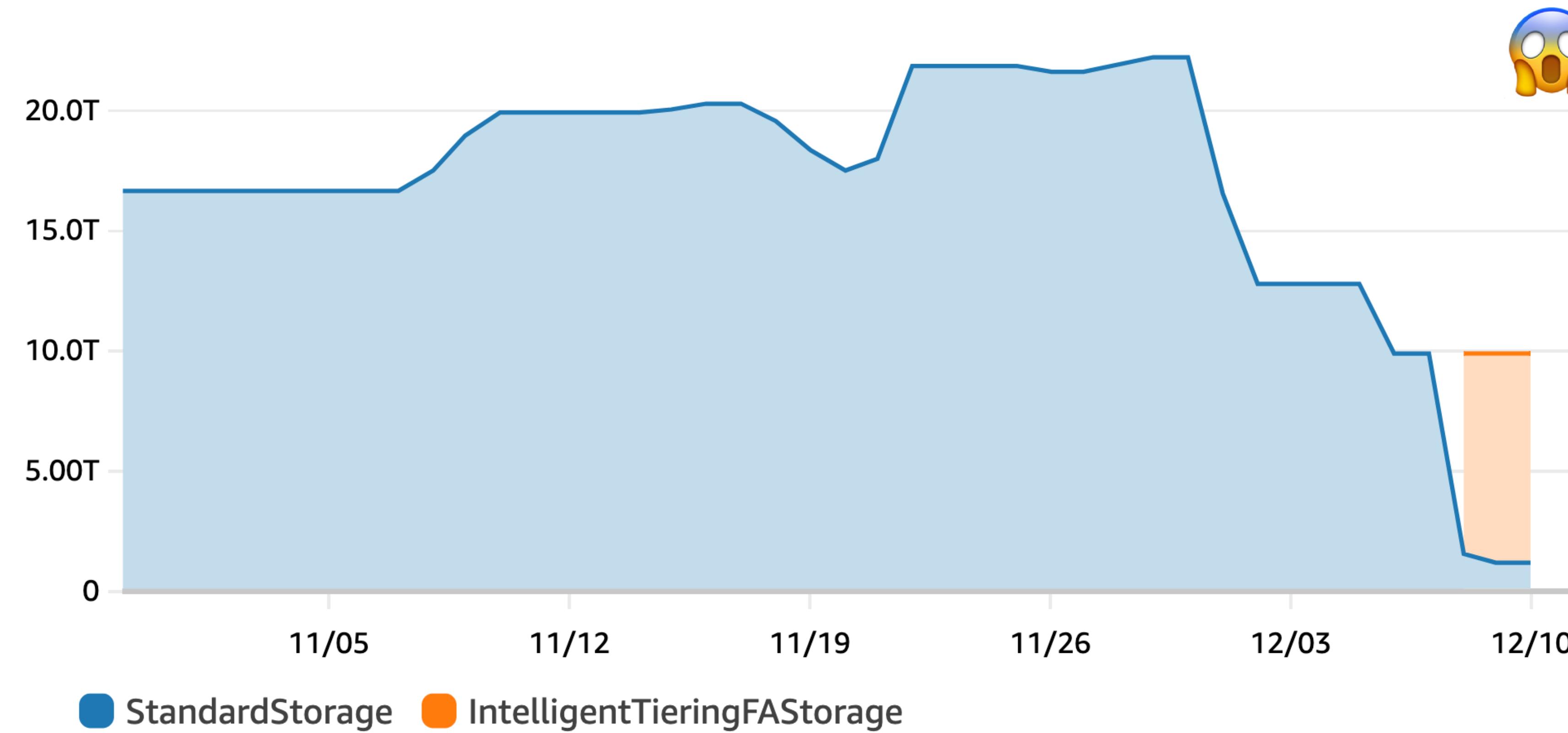


Know thy buckets

Know thy buckets



Know thy buckets

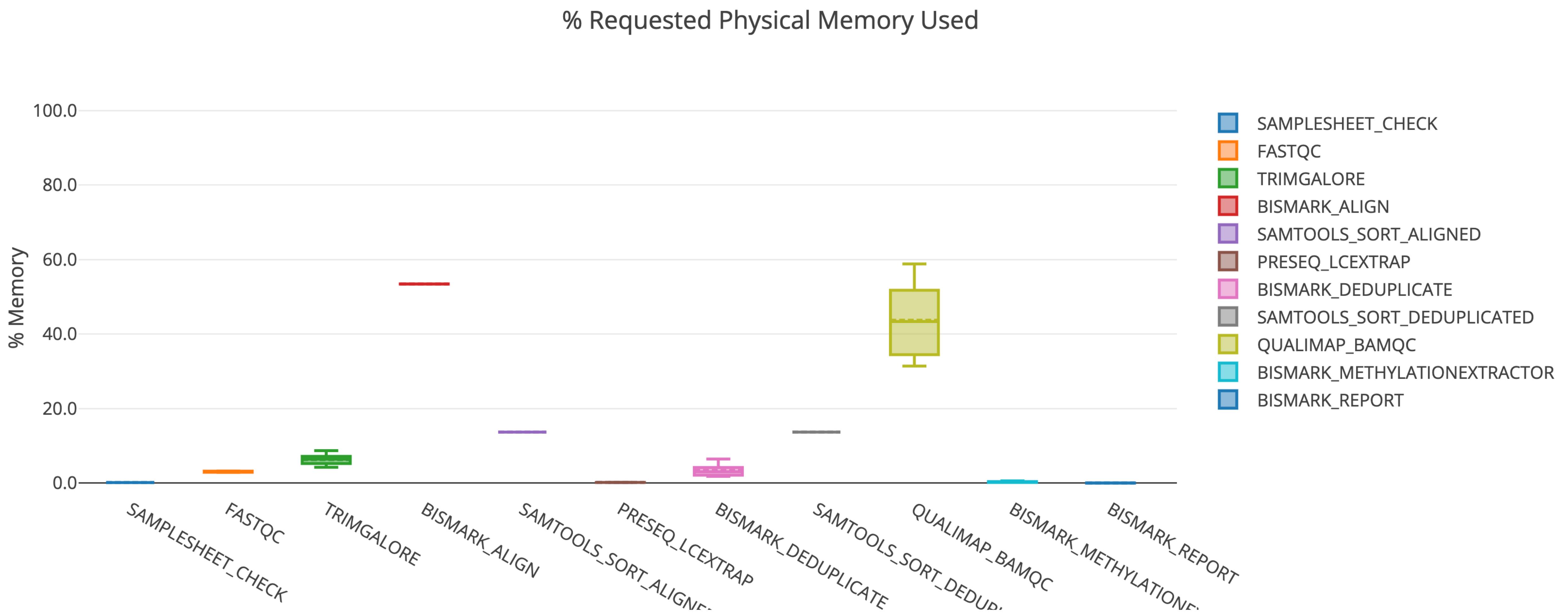


Optimising resource requests

Optimising resource requests

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

Optimising resource requests



Optimising resource requests

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

[Launchpad](#) [Runs](#) [Actions](#) [Datasets](#) [Compute Environments](#) [Credentials](#) [Secrets](#) [Participants](#) [Settings](#)

< 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
```

[Launchpad](#) [Runs](#) [Actions](#) [Datasets](#) [Compute Environments](#) [Credentials](#) [Secrets](#) [Participants](#) [Settings](#)

< View Workflow Run

curious_sanger nf-core/rnaseq

[azure-batch](#)[spot-instances](#)[Command line](#)[Parameters](#)[Configuration](#)[Datasets](#)[Execution log](#)[Reports](#)

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

Optimization available

[Launch](#)



Launchpad

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_IDXSTATS'
        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
        memory = 1.GB
    }
}
```

Re-launch

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

X SUMMIT 2022



Venkat Malladi

Microsoft

Deep dive into Nextflow on Azure

X SUMMIT 2022



Hatem Nawar

Google

Pipeline economics on Cloud

X SUMMIT 2022

Phil Ewels

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

phil@seqera.io

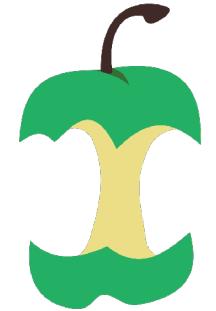


<https://seqera.io>

Chan Zuckerberg
Initiative



nextflow

nf-core The nf-core logo icon is a stylized green 'X' shape with a small stem and leaf at the top right corner.

nextflow tower

Nextflow / nf-core training

6-10 March 2023

nf-core hackathon

20-24 March 2023

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