

NEXTFLOW & NF-CORE

Phil Ewels

National Genomics Infrastructure
SciLifeLab Sweden

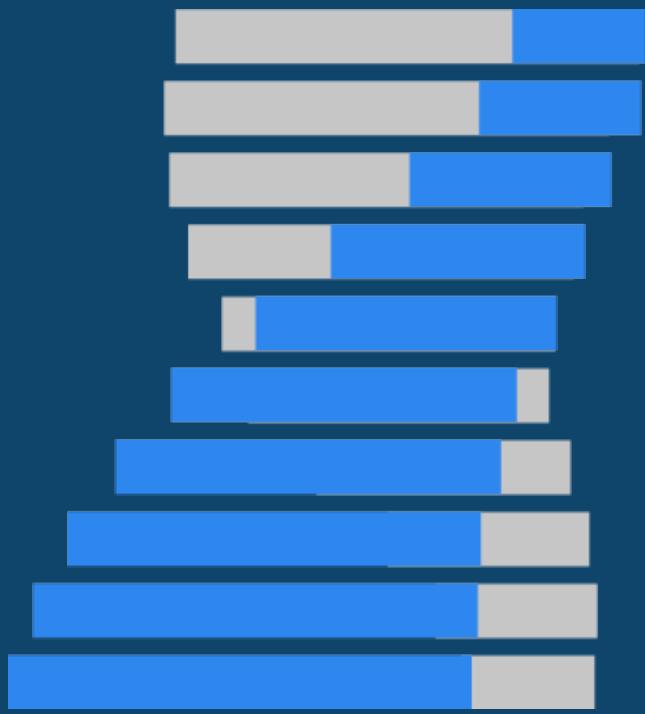
phil.ewels@scilifelab.se



NATIONAL
GENOMICS
INFRASTRUCTURE

SciLifeLab

<https://scilifelab.se>
<https://ngisweden.scilifelab.se>



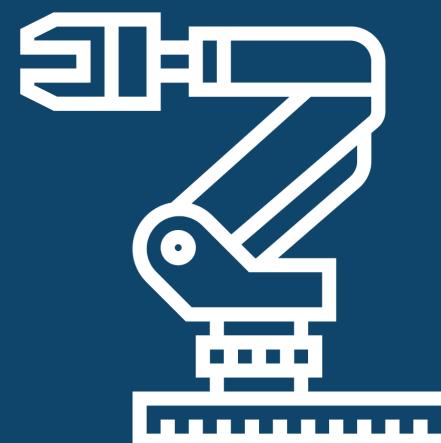
NATIONAL GENOMICS INFRASTRUCTURE



Guidelines
and support



Available to all
Swedish researchers



State-of-the-art
infrastructure



NATIONAL GENOMICS INFRASTRUCTURE



Meetings



QC & Library
Preparation

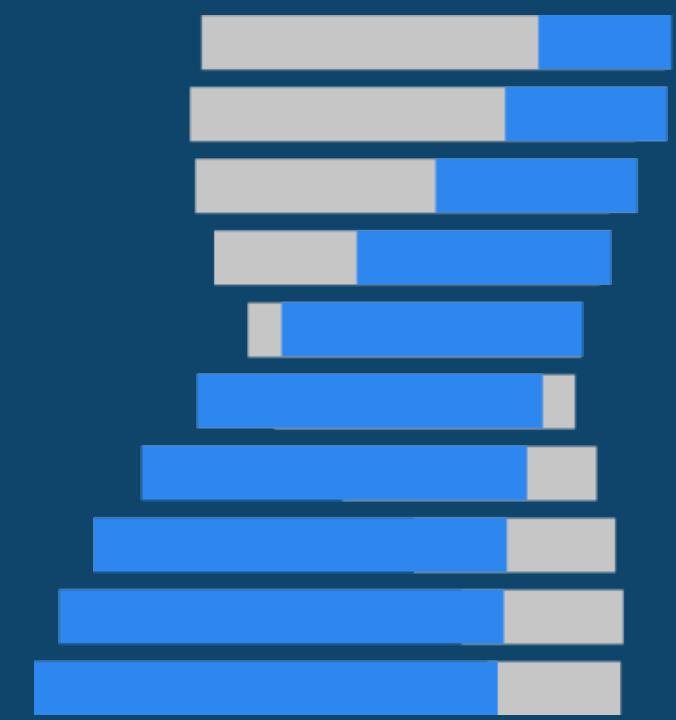


Sequencing
Genotyping

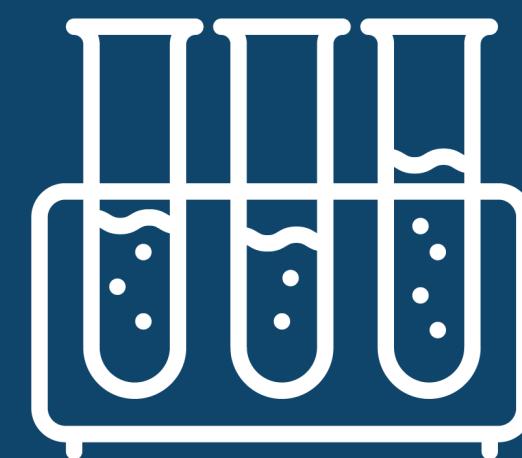


Data analysis
& delivery

- RNA-seq
- DNA-seq (inc. ChIP)
- Methylation
- ATAC-seq
- Hi-C
- Modifications
- Enrichment



NATIONAL GENOMICS INFRASTRUCTURE



QC & Library
Preparation

- RNA-seq

<https://ngisweden.scilifelab.se>

- DNA-seq (inc. ChIP)

support@ngisweden.se

- Methylation

- ATAC-seq

- Hi-C

- Modifications

- Enrichment

NATIONAL GENOMICS INFRASTRUCTURE

HOME SERVICES ▾ RESOURCES ▾ NEWS ▾ ABOUT US ▾ CONTACT NEW ORDER 

Home / Applications / Epigenetics

Epigenetics

Methods to characterise genetic or transcriptomic differences other than the nucleotide composition, e.g. 3D conformation of chromatin or levels of methylation and other DNA/RNA base modifications.

Applications 4

Methods 9

Bioinformatics 5

DNA Methylation 
Methods to determine the proportion of Cytosine methylation (5mC), the most common type of base modification in DNA.

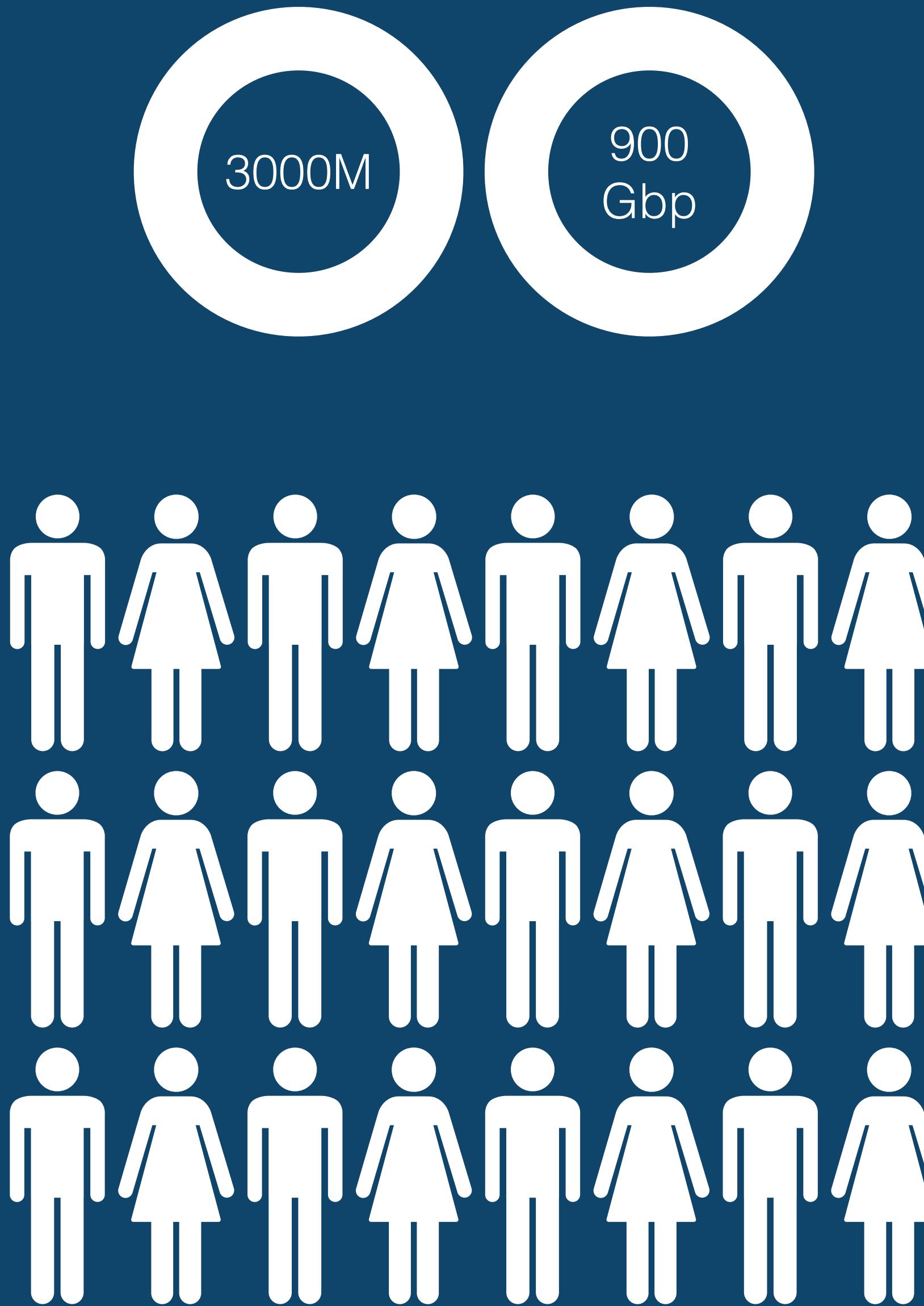
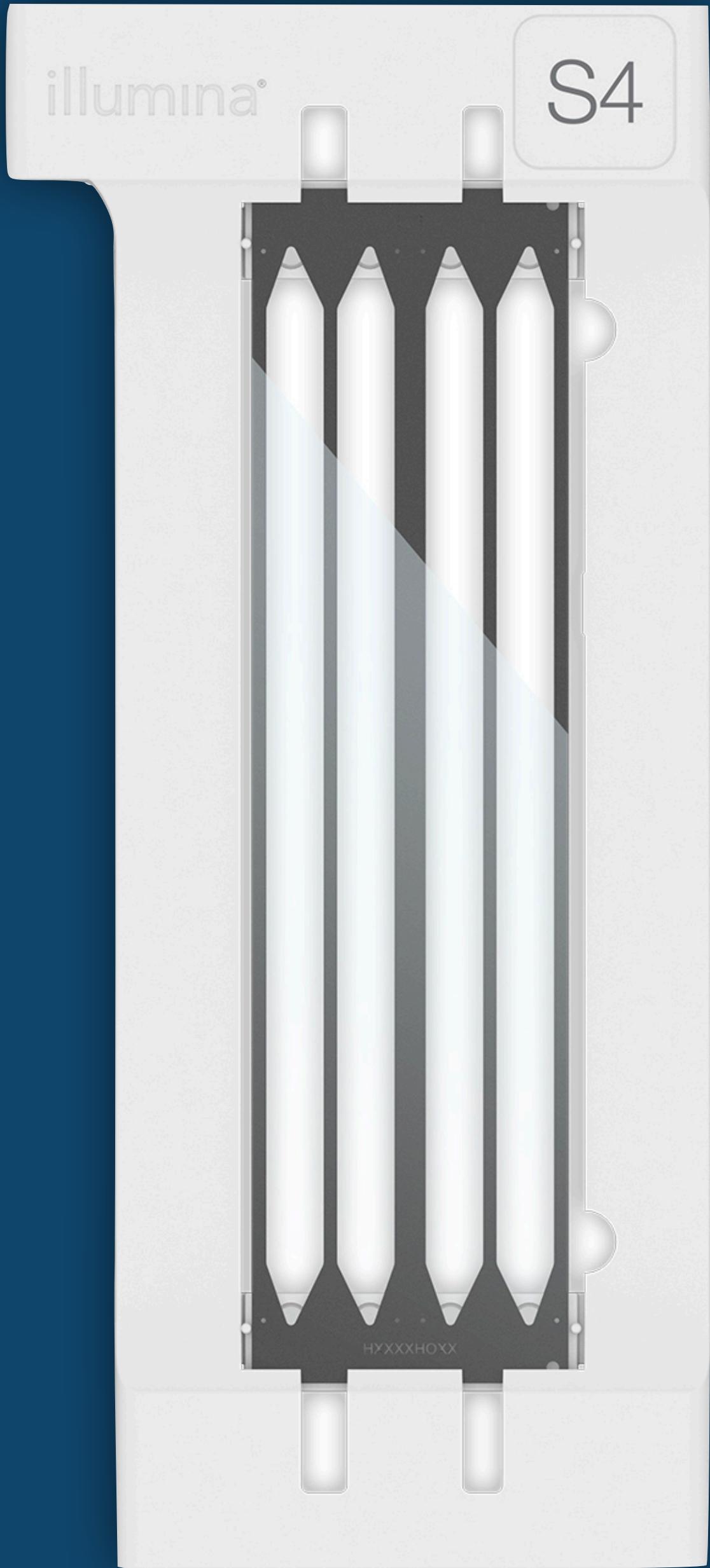
ChIP-Sequencing 
Chromatin immunoprecipitation sequencing (ChIP-seq) is used to analyze protein interactions with DNA across the genome.

HiC 
A method to characterise genome-wide chromatin contacts. Can be used to investigate the 3D genome organisation in the nucleus, e.g. to study topologically-associated functional domains (TADs).

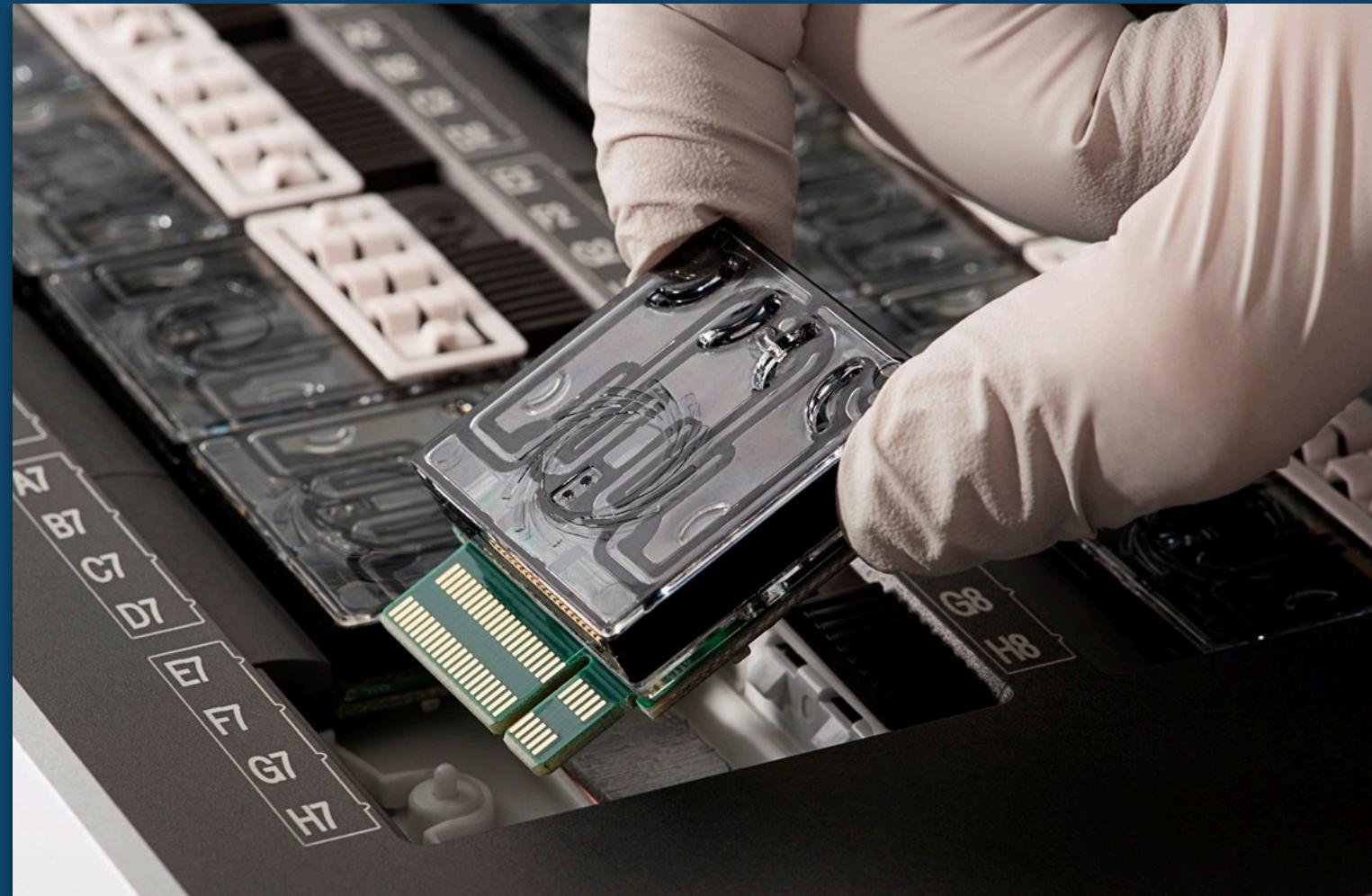
ATAC-Sequencing 
The Assay for Transposase Accessible Chromatin (ATAC-seq) gives information on the open chromatin regions in a genome. We offer both bulk and single-cell ATAC-seq.







~24 genomes
30X



2500K

290
Gbp



Direct calling of base modifications

Long-read phasing of epigenetic marks

Isoforms / fusions

Adaptive / selective sequencing?

Multi-contact HiC?



NATIONAL
GENOMICS
INFRASTRUCTURE



The background of the image is a photograph of a vast, misty mountain range with a dense forest of evergreen trees in the foreground. The sky is a clear, pale blue.

nextflow

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

nextflow

```
#!/usr/bin/env nextflow
input = Channel.fromFilePairs(params.reads)

process fastqc {
    input:
        file reads from input

    output:
        file "*_fastqc.{zip,html}" into results

    script:
    """
        fastqc -q $reads
    """

}
```



SGE



Google Cloud



LSF

PBS

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

nextflow

```
#!/usr/bin/env nextflow
input = Channel.fromFilePairs(params.reads)

process fastqc {
    input:
        file reads from input

    output:
        file "*_fastqc.{zip,html}" into results

    script:
    """
        fastqc -q $reads
    """

}
```



Singularity



kubernetes



SGE



Google Cloud



docker



CONDA



LSF

PBS

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

nextflow

```
#!/usr/bin/env nextflow
input = Channel.fromFilePairs(params.reads)

process fastqc {
    input:
        file reads from input

    output:
        file "*_fastqc.{zip,html}" into results

    script:
    """
        fastqc -q $reads
    """

}
```



LSF

PBS

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



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

<https://nf-co.re>



Guidelines



Tools



Pipelines

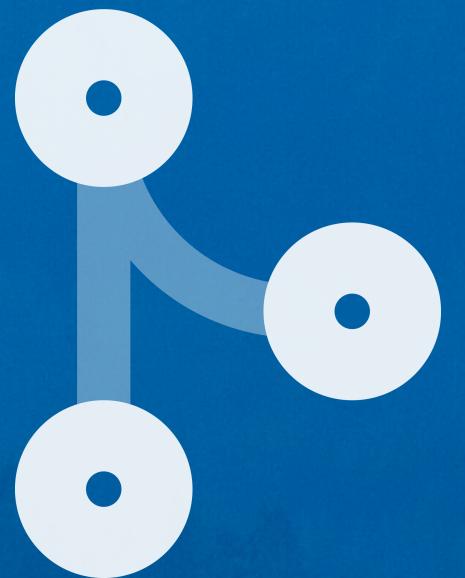
nf-core



Develop with
the community

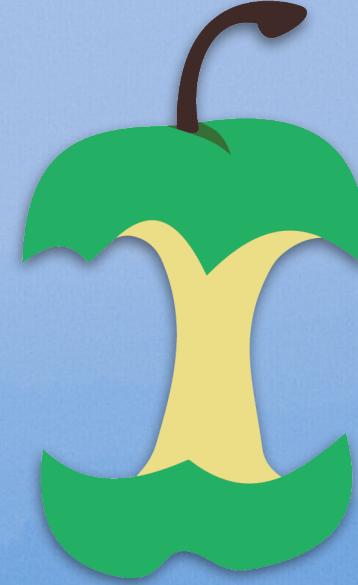


Start from the
template



Collaborate,
don't duplicate

nf-core



16
UNDER DEVELOPMENT



5
ARCHIVED

Available Pipelines

Can you think of another pipeline that would fit in well? [Let us know!](#)

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/imcyto ✓

☆ 5

cytometry image-analysis image-processing image-segmentation

Image Mass Cytometry analysis pipeline

Version **1.0.0**

Published 1 month ago

nf-core/coproid ✓

☆ 2

adna ancient-dna coprolite microbiome

Sample-level identification pipeline

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/slamseq ✓

☆ 3

differential-expression quantseq slamseq transcriptomics

SLAMSeq processing and analysis pipeline

Version **1.0.0**

Published 2 months ago

nf-core/mhcquant ✓

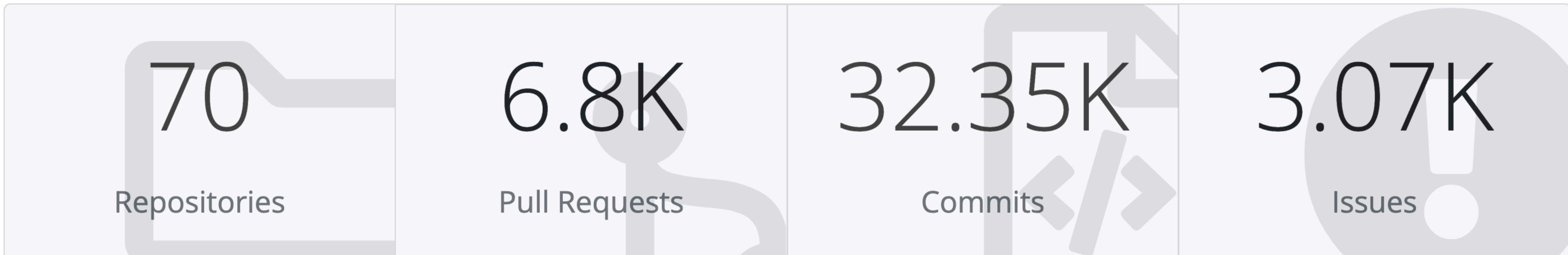
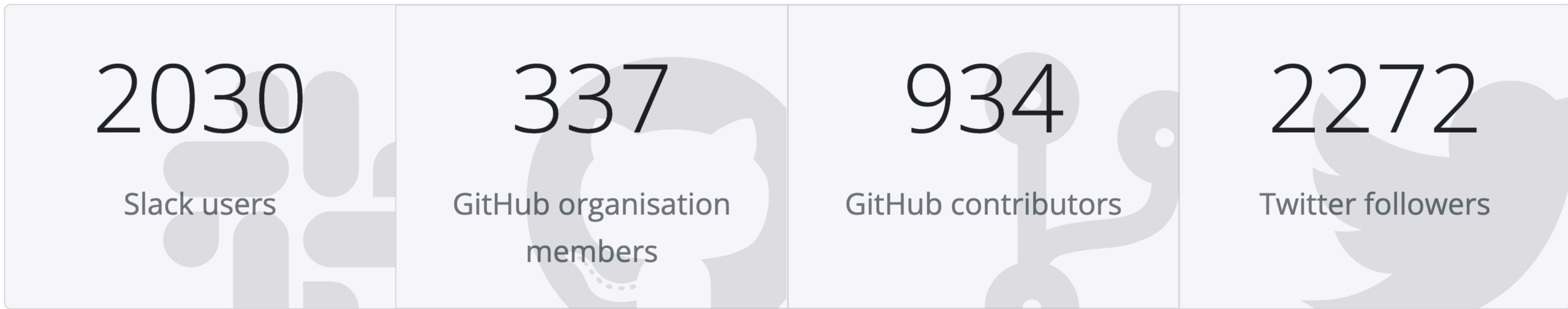
☆ 12

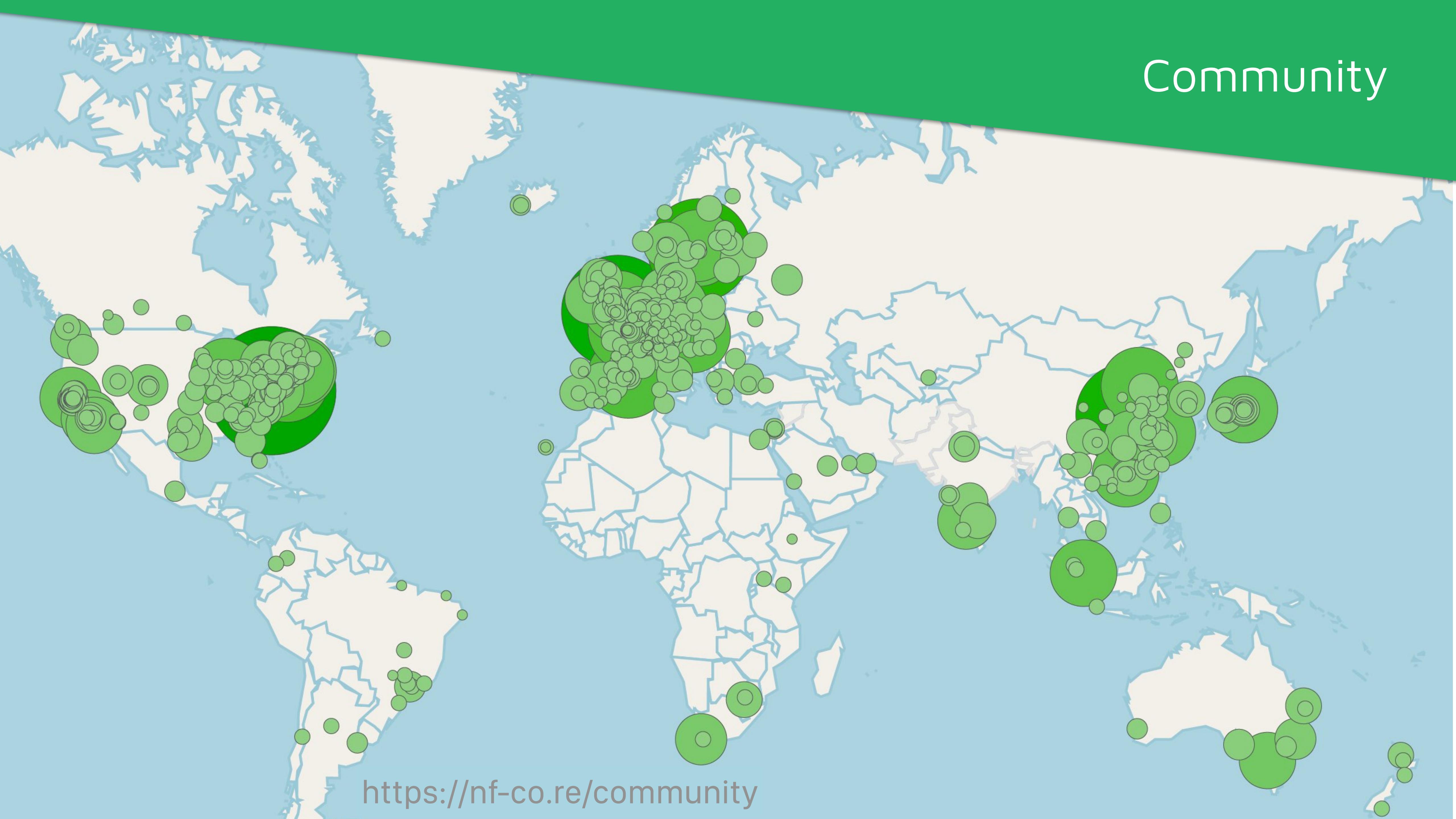
mass-spectrometry mhc peptides

Identify and quantify MHC peptides from mass-spectrometry data

Community

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





Community

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

Correspondence | Published: 13 February 2020

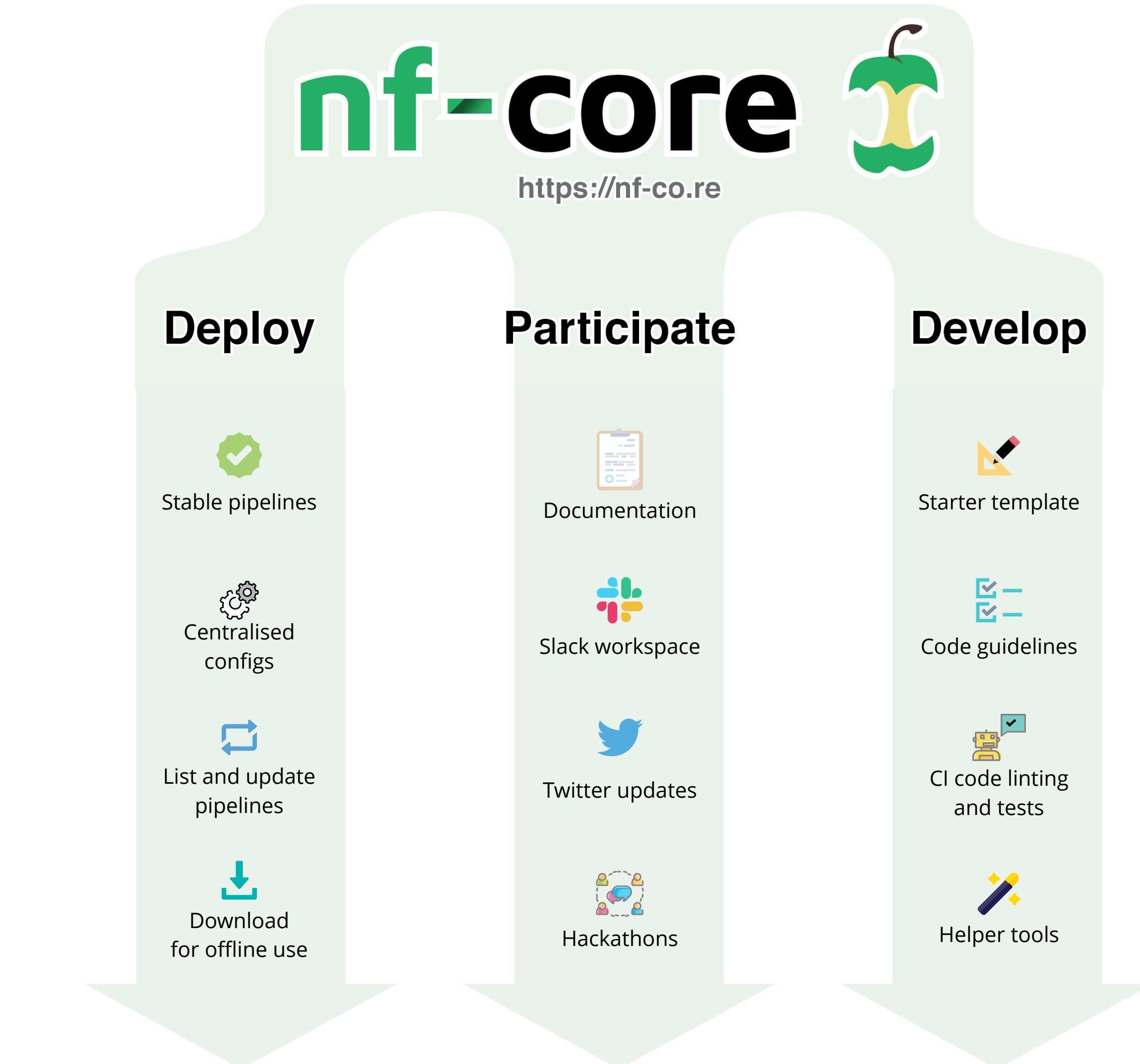
The nf-core framework for community-curated bioinformatics pipelines

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

3253 Accesses | 3 Citations | 172 Altmetric | Metrics

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.



Get involved



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

Nextflow

nextflow

```
curl -s https://get.nextflow.io | bash
```

```
module load nextflow
```



```
conda install nextflow
```

BIOCONDA®

Nextflow



~/.bashrc

```
# Nextflow singularity image cachedir
export NXF_SINGULARITY_CACHEDIR=/proj/MYPROJ/nobackup/phil/singularity-images

# Nextflow environment variables
export NXF_OPTS=' -Xms1g -Xmx4g'
export NXF_HOME=/proj/MYPROJ/nobackup/phil/nxf-home
export NXF_TEMP=${SNIC_TMP:-$HOME/glob/nxftmp}
```

nf-core/tools

nf-core/

tools

```
pip install nf-core
```

```
module load nf-core
```

```
conda install nf-core
```



BIOCONDA[®]

Running a pipeline

Let Nextflow fetch the pipeline for you!

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

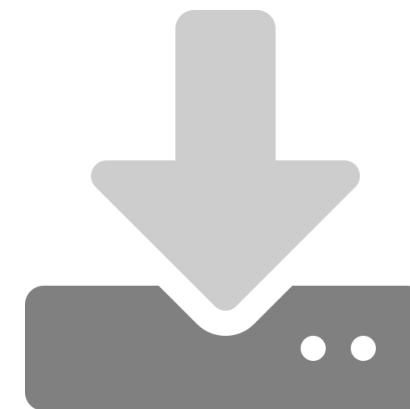
GitHub pipeline name
Clones to `~/.nextflow`

Pipeline release tag
Repository branch

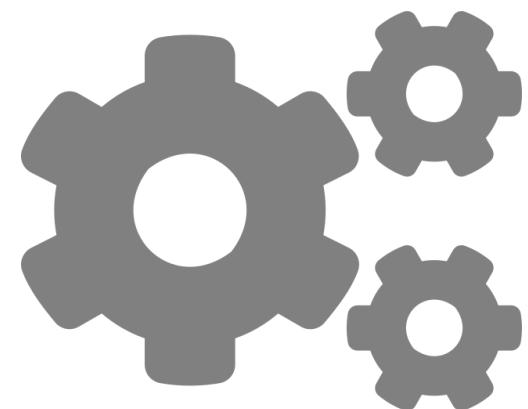
Offline usage

Offline? Fetch using nf-core helper tool

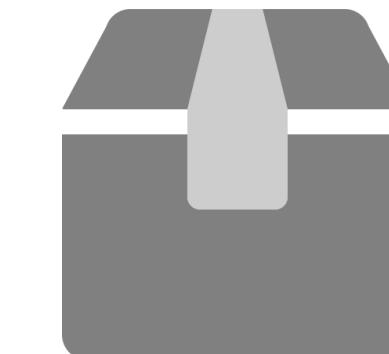
```
nf-core download nf-core/<pipeline> -r <version>
```



Downloads pipeline



Downloads configs



Fetches Singularity containers

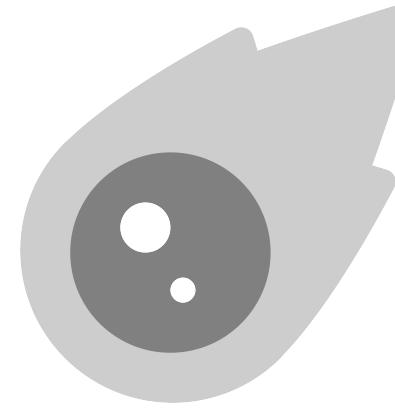


Creates archive

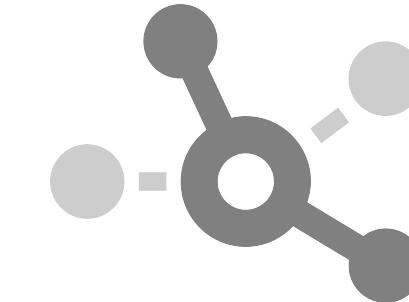
Central location of all nf-core pipelines

```
module load df-core-pipelines
```

```
nextflow run $NF_CORE_PIPELINES/<pipeline>/<version>/workflow
```



All pipelines, all versions
available out of the box



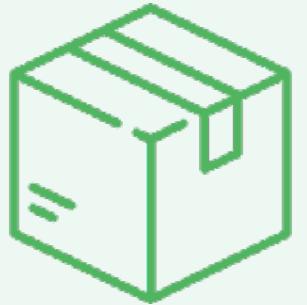
Uses central store
of Singularity containers



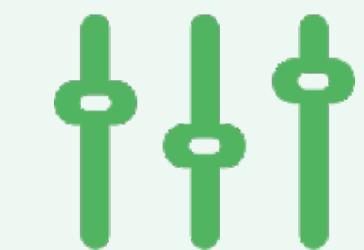
Available on Bianca 😊



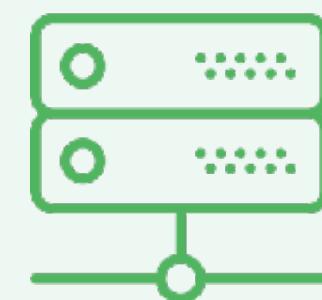
Very new!
Please report problems



Default 'base' config (always loaded)



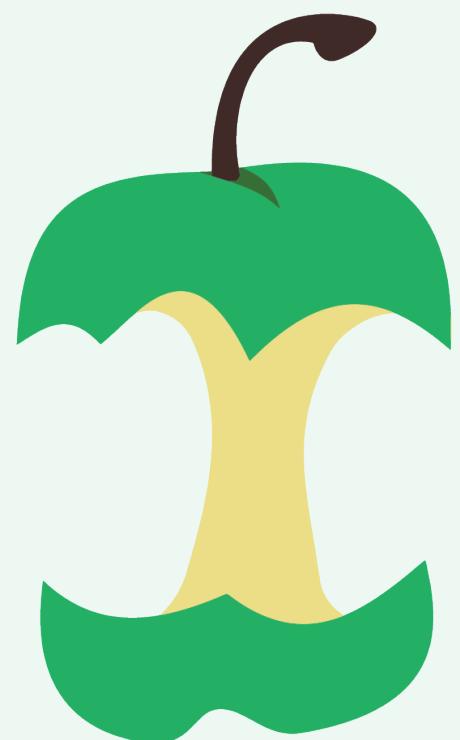
Core profiles (eg. docker, conda, test)

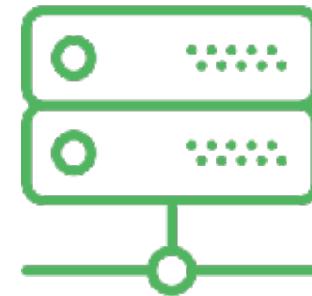


Institutional profiles (nf-core/configs)

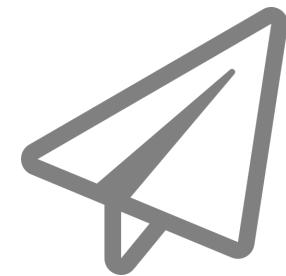


Your local config files (eg. the -c flag)





Institutional profiles : UPPMAX



SLURM, with parameters for UPPMAX projects



Singularity, automatic settings for resource limits



Local copies of iGenomes references

```
nextflow run <pipeline> -profile uppmax --project XXX
```



Running a pipeline

```
nextflow run <pipeline> --option_one
```

params.option_one

```
nextflow run <pipeline> -c myrun.config
```

```
nextflow run <pipeline> -params-file myrun.json
```

Running a pipeline

```
nextflow run nf-core/<pipeline> [params]
```

Remember to detach from your terminal!



-bg flag

screen or tmux

Submit Nextflow as a job

Running a pipeline

```
nf-core launch <pipeline>
```



Help / validation for pipeline parameters
Use online form or offline command-line wizard

nf-core



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

nf-core

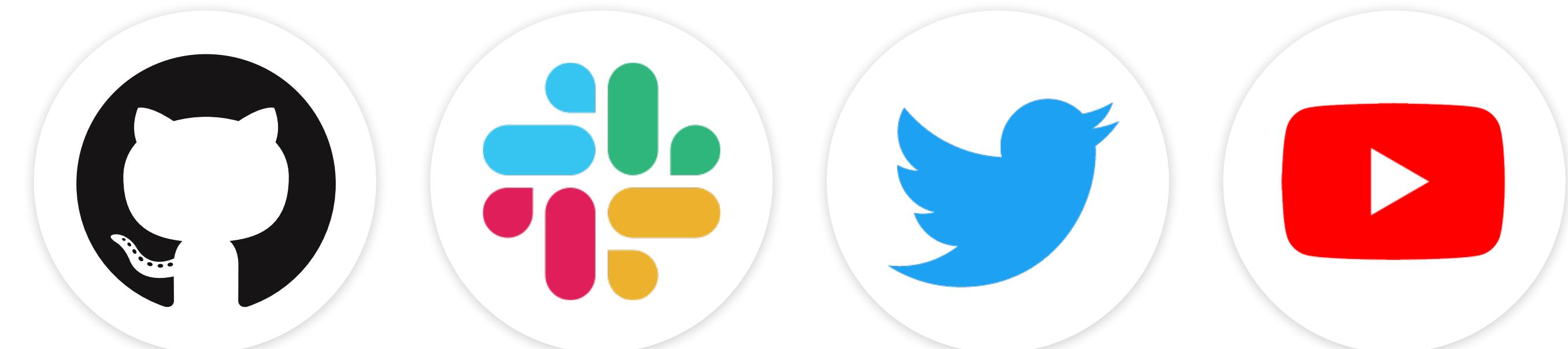


Phil Ewels

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



<https://ngisweden.scilifelab.se>



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