Making reproducible workflows with

nextlow

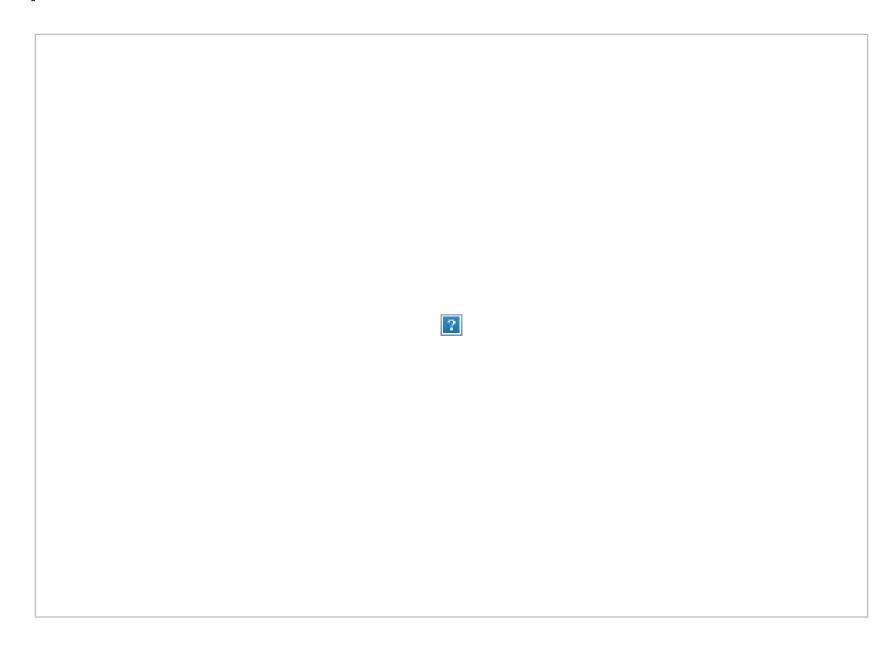


Nextflow features

- Generalisable
- Portable
- Scalable
- Platform-agnostic
- Based on Groovy and Java
- Large active community in e.g. nf-core



Concepts and nomenclature



- Channels contain data, e.g. input files
- Processes run some kind of code, e.g. a script or a command-line program

```
process GET_SRA_BY_ACCESSION {
  input:
  val(sample)
  output:
  path("${sample}.fastq.gz")
  script:
     fastq-dump ${sample} > ${sample}.fastq.gz
}
```



```
process GET_SRA_BY_ACCESSION {
    input:
    val(sample)
    output:
    tuple val(sample), path("${sample}.fastq.gz")
    script:
    inut
    fastq-dump ${sample} > ${sample}.fastq.gz
}
```



```
cpus 2
memory '8 GB'
input:
val(sample)

output:
tuple val(sample), path("${sample}.fastq.gz")

script:
####
fastq-dump ${sample} > ${sample}.fastq.gz
}
```



```
process GET_SRA_BY_ACCESSION {
    cpus 2
    memory '8 GB'

    conda 'sra-tools=2.11.0'
    container 'ncbi/sra-tools:2.11.0'

    input:
    val(sample)

    output:
    tuple val(sample), path("${sample}.fastq.gz")

    script:
    imm
    fastq-dump ${sample} > ${sample}.fastq.gz
}
```



```
process GET_SRA_BY_ACCESSION {
    cpus 2
    memory '8 GB'
    conda 'sra-tools=2.11.0'
    container 'ncbi/sra-tools:2.11.0'
    input:
    val(sample)
    output:
    tuple val(sample), path("${sample}.fastq.gz")
    script:
    imutiful restriction of the second o
```



Anatomy of a workflow



Anatomy of a workflow



Anatomy of a workflow



Executing Nextflow

Execute a workflow \$ nextflow run main.nf

Re-run using cached results \$ nextflow run main.nf -resume

Execute with a specific configuration file \$ nextflow run main.nf -c nextflow.config

Supply a custom parameter \$ nextflow run main.nf --my_param "my value"

Use Docker or Singularity \$ nextflow run main.nf -with-docker \$ nextflow run main.nf -with-singularity

Use a pre-defined configuration profile \$ nextflow run main.nf -profile uppmax



Differences between Snakemake and Nextflow

	Snakemake	Nextflow
Language	Python	Groovy
Data	Everything is a file	Can use both files and values
Execution	Working directory	Each job in its own directory
Philosophy	"Pull"	"Push"
Dry-runs	Yes	No
Track code changes	No	Yes

- Question: But, which one is the best?
- Answer: Both it's mostly up to personal preference!



Questions?

