nextflow

Nextflow Introduction

ScRNAseq in the Cloud: Pipelines to Insights

2024

Workflows

Using computers to collect, store, analyze, and disseminate data and information

Large files

> 100 GB for one raw human genome...

Many languages

Bash, Python, R, PERL...

Complex interactions

Networks of software and their dependencies...

Nextflow

Nextflow is a language, a runtime, and a community

Reproducible

Integration with code management tools, with versioned releases.

Portable

Docker, Singularity, Conda, works with most compute environments.

Scalable

5 samples on your laptop, 5k on an HPC or 5 million in the cloud.

Reproducibility

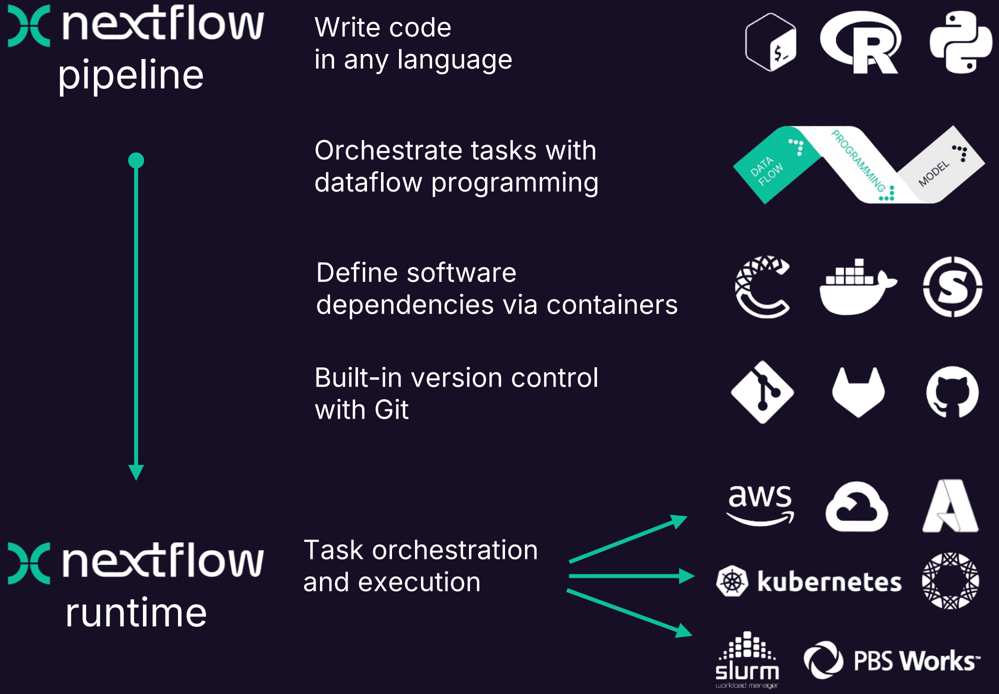
Hidden reproducibility issues are like an iceberg

> First, we tried to re-run the analysis with the code and data provided by the authors.
>
> Second, we rewrote the whole method in a Python package...

Experimenting with reproducibility:
a case study of robustness in bioinformatics
Kim et al., GigaScience (2018).
https://doi.org/10.1093/gigascience/giy077

Nextflow

Nextflow is a language, a runtime, and a community



nextflow pipeline

Write code
in any language

Orchestrate tasks with
dataflow programming

Define software
dependencies via containers

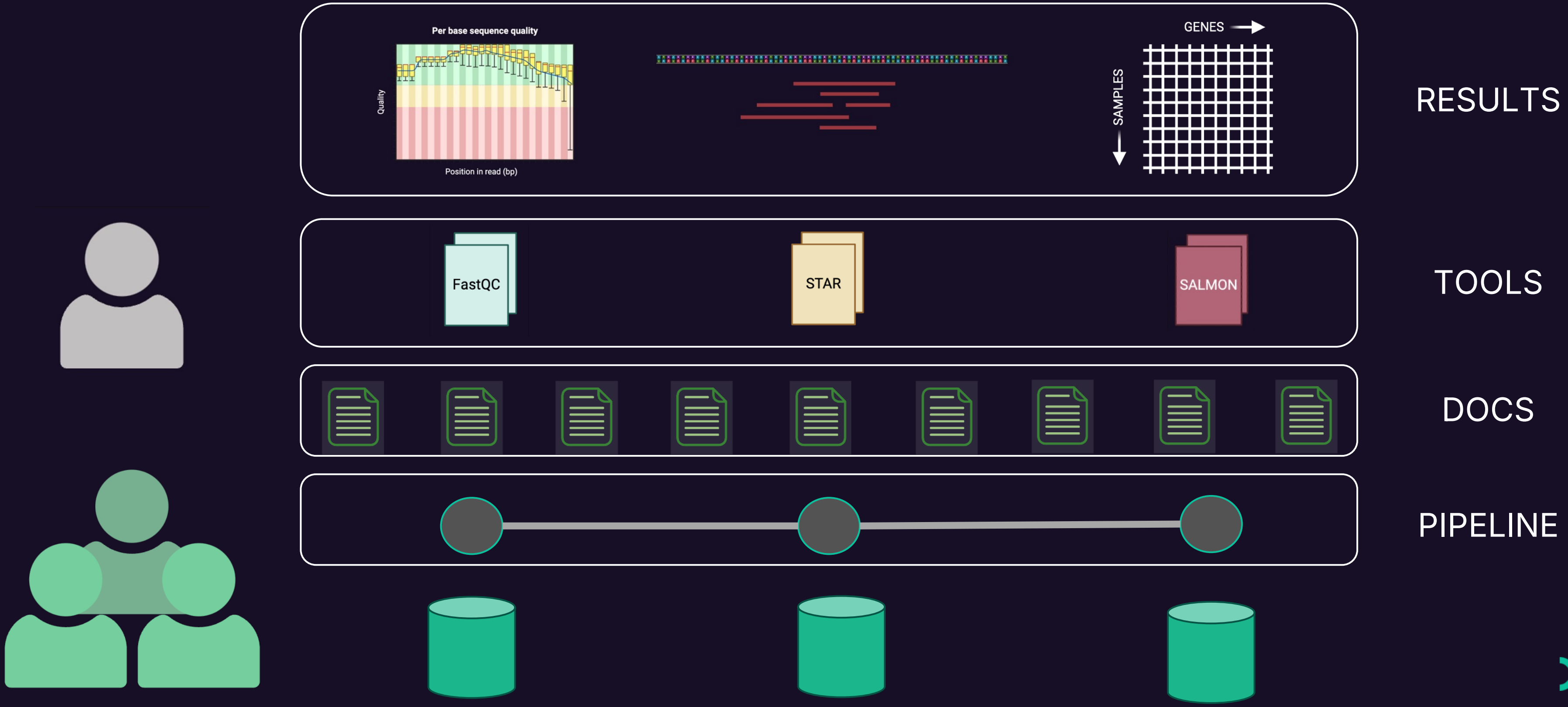
Built-in version control
with Git

nextflow runtime

Task orchestration
and execution

Nextflow

Nextflow is a language, a runtime, and a community



nf-core

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

8k+

Slack users

2k+

GitHub contributors

120+

GitHub repositories

40k

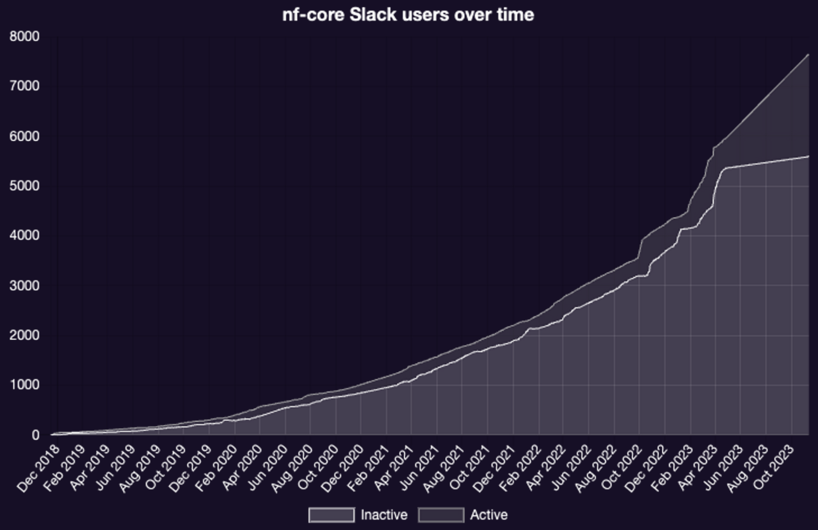
GitHub commits

16k+

Pull requests

7k+

GitHub issues



nf-core principles

Principles that guide the nf-core community

Cooperation

Develop with the community

Standards

Use a common template

Collaboration

No duplicate pipelines within nf-core

Helper Tools

Tools built for everyone

Compatibility

Tools work for any Nextflow pipeline

Components

Collaborate on component-level

nf-core components

Pick and choose which component you need

Pipelines

>95 pipelines and a base template

Subworkflows

>55 subworkflows

Modules

>1150 modules

Linting

Choose conventions to test for consistency

Schema

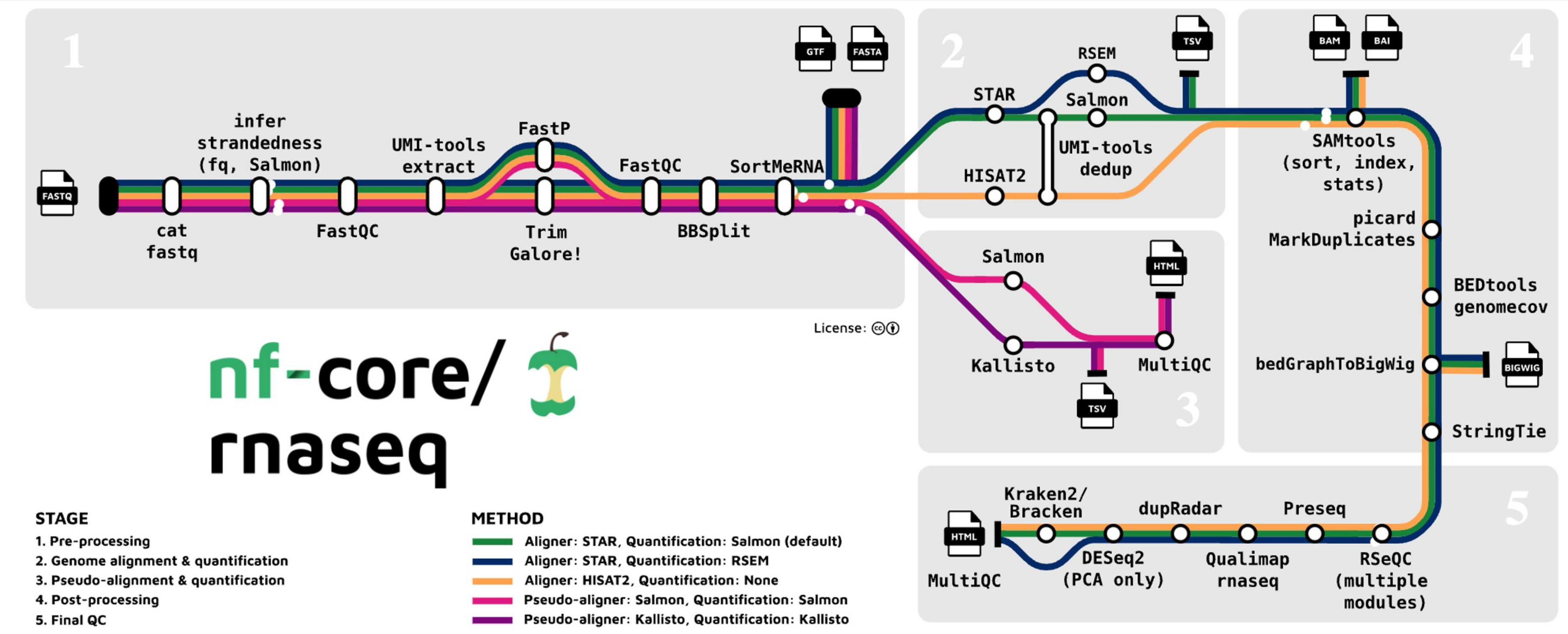
Validation, channels and user interface

Tooling

Development and deployment

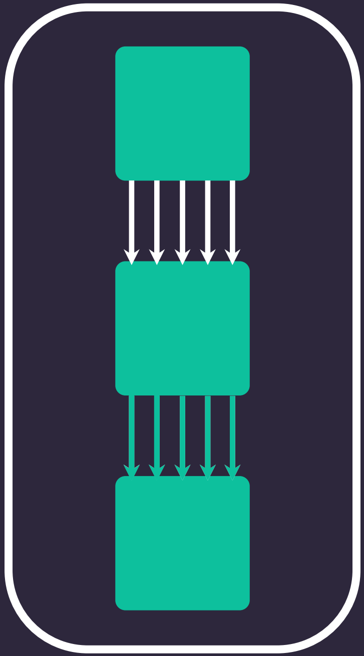
Nextflow

Managing modern workflows is complicated



Nextflow

A reactive workflow framework and a programming DSL



Parallelism

Reentrancy
(Resume partial runs)

Reproducibility

Nextflow

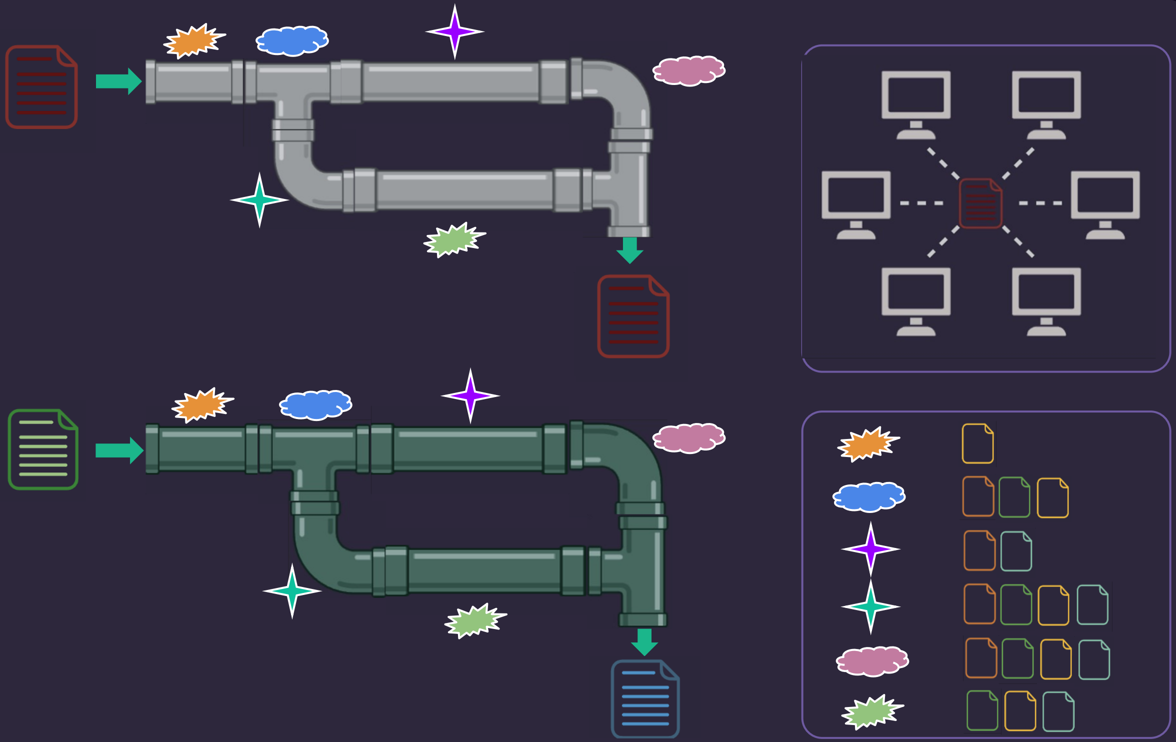
Use Case Studies - Reproducibility

Nextflow

Use Case Studies - Reentrancy

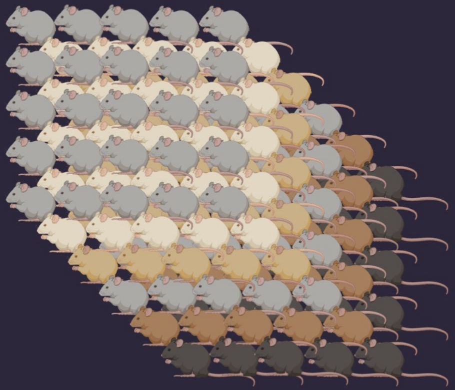
Custom Pipeline

Nextflow Pipeline



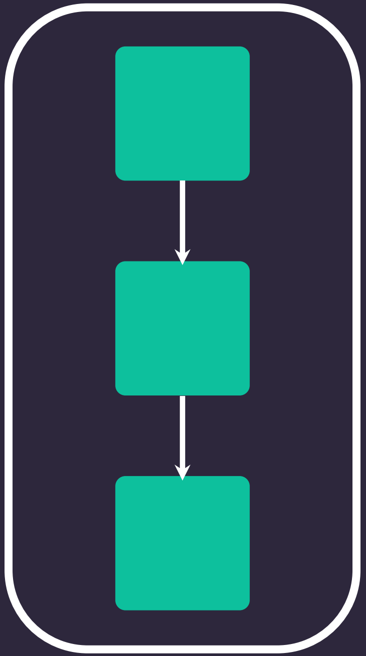
Nextflow

Use Case Studies - Parallelism



Nextflow

A reactive workflow framework and a programming DSL



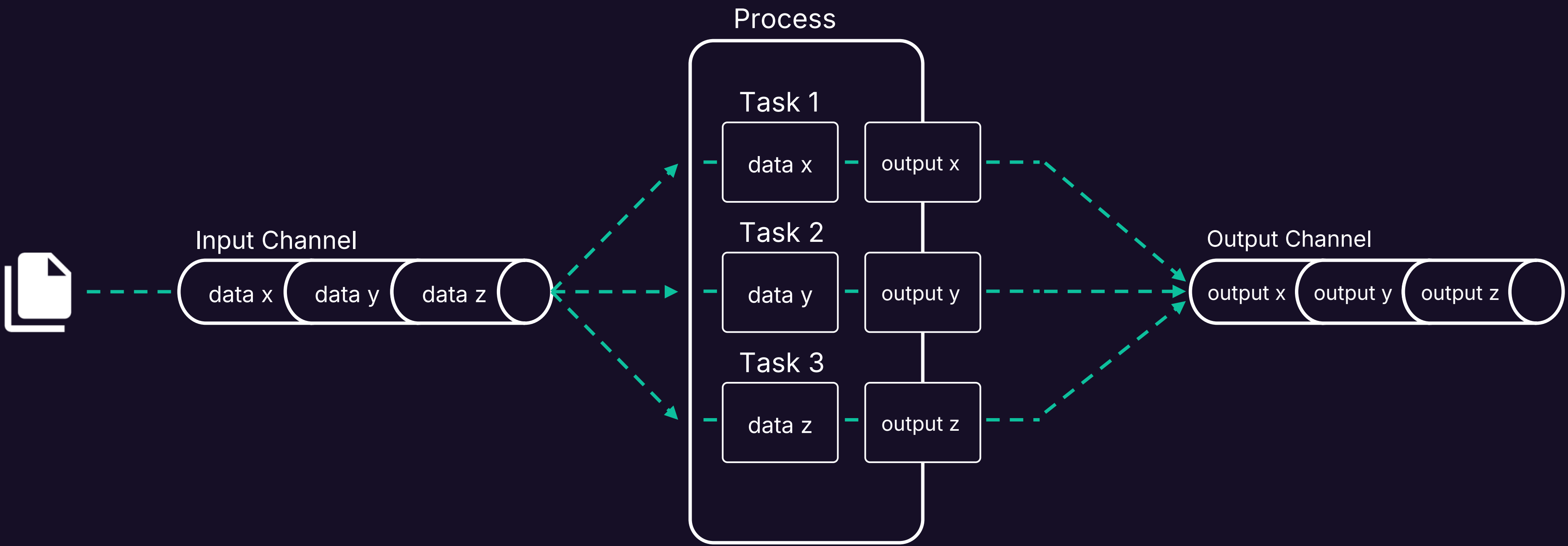
Processe
s

Channel
s

Workflow

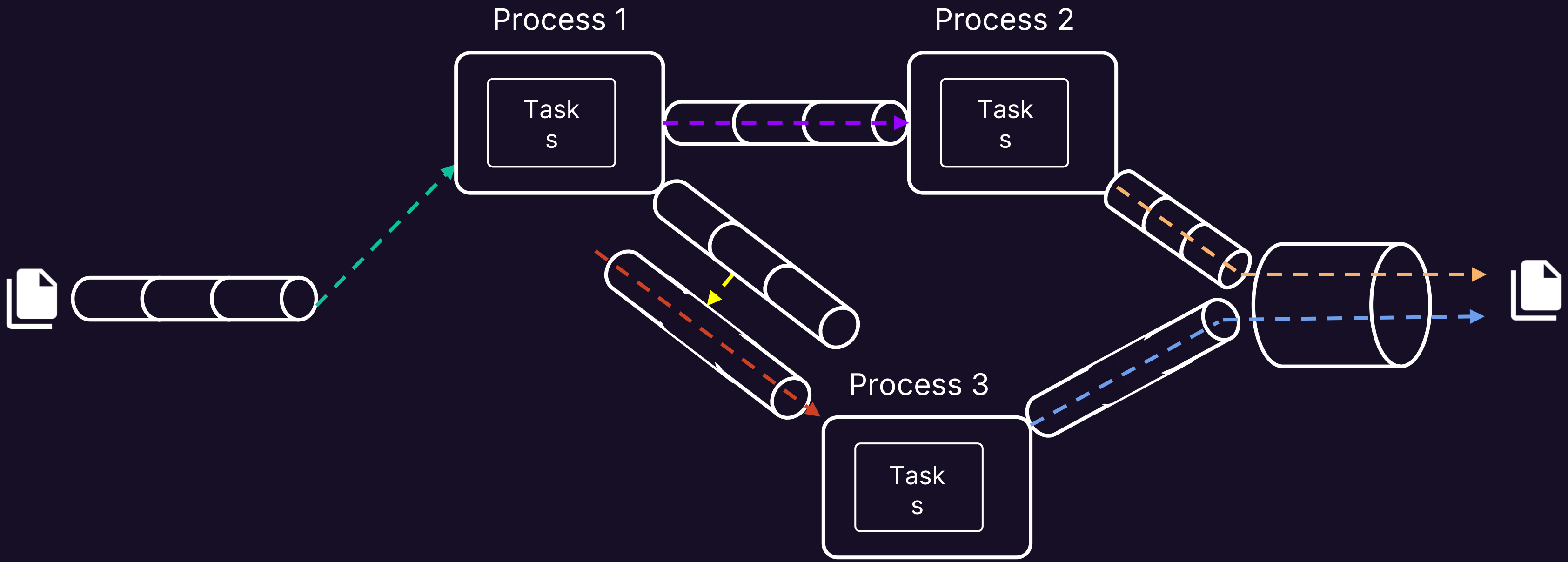
Nextflow

A reactive workflow framework and a programming DSL



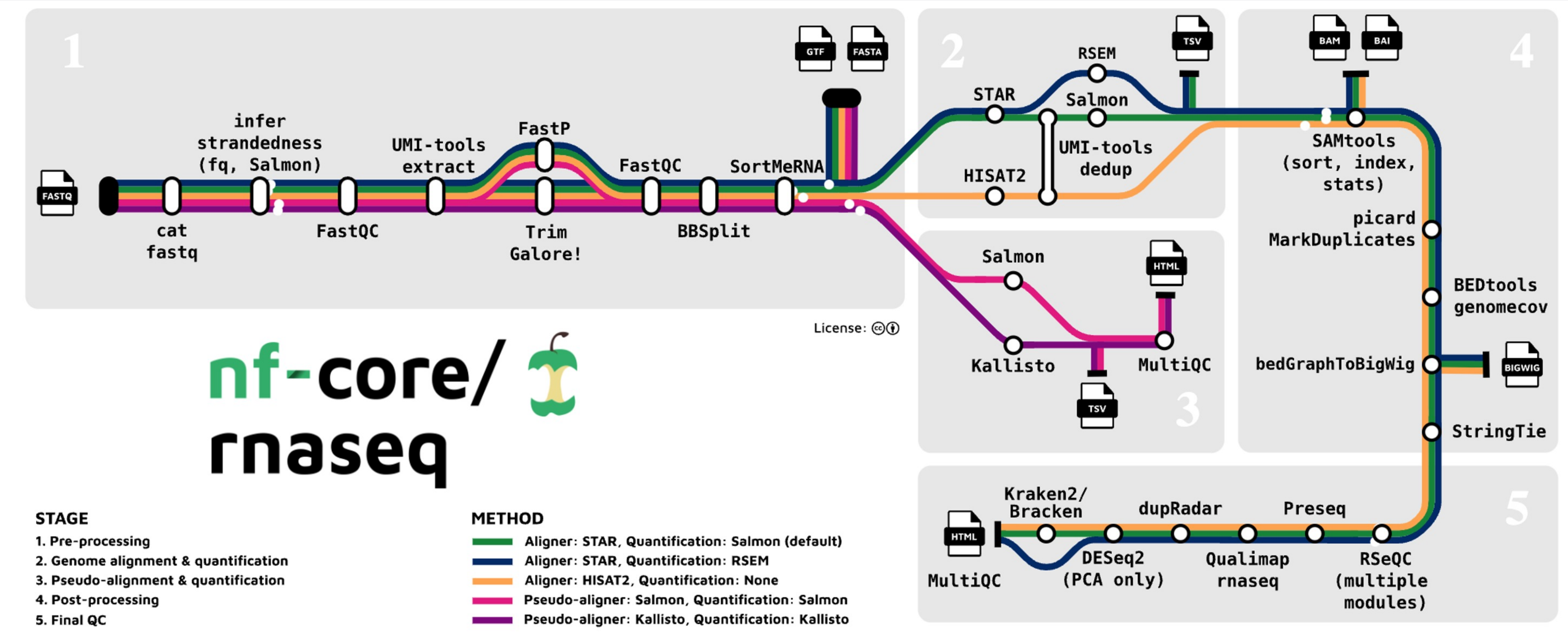
Nextflow

A reactive workflow framework and a programming DSL



Nextflow

Managing modern workflows is easier with Nextflow



nf-core/
rnaseq

STAGE

1. Pre-processing
2. Genome alignment & quantification
3. Pseudo-alignment & quantification
4. Post-processing
5. Final QC

METHOD

- 🟩 Aligner: STAR, Quantification: Salmon (default)
- 🟦 Aligner: STAR, Quantification: RSEM
- 🟧 Aligner: HISAT2, Quantification: None
- 🟥 Pseudo-aligner: Salmon, Quantification: Salmon
- 🟪 Pseudo-aligner: Kallisto, Quantification: Kallisto