

as the standard for BovReg reference pipelines

Jose Espinosa-Carrasco¹, Björn E. Langer¹, Philip A. Ewels⁴, Harshil Patel⁴, Peter Harrison³, Cedric Notredame^{1,2}

1 Centre for Genomic Regulation (CRG), Spain
 2 Universitat Pompeu Fabra (UPF), Spain
 3 European Bioinformatics Institute (EMBL-EBI), UK.
 4 Seqera Labs, Spain

Abstract

Analysis pipelines of research genomic consortia, used to process often large quantities of data, tend to be difficult to re-use, due to poor computational practices, leading to irreproducible analyses. To prevent this, the BovReg project chose nf-core as the computational standard, guaranteeing reproducibility, portability, interoperability and a unified minimal functionality.



Reduced environmental impact

Disease resistance

Training

Biological efficiency

Precision breeding

Research Methodology

Feeding a growing population

Communication and Dissemination

nextflow

Data-driven computational pipelines

Nextflow¹ enables scalable and reproducible scientific workflows using software containers. It allows the adaptation of pipelines written in the most common scripting languages. Its fluent DSL simplifies the implementation and the deployment of complex parallel and reactive workflows on clouds and clusters.

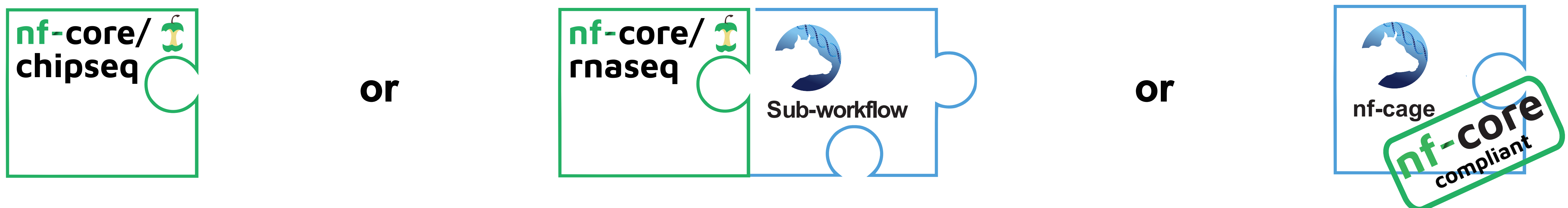
nf-core

A curated set of Nextflow analysis pipelines - a community effort.

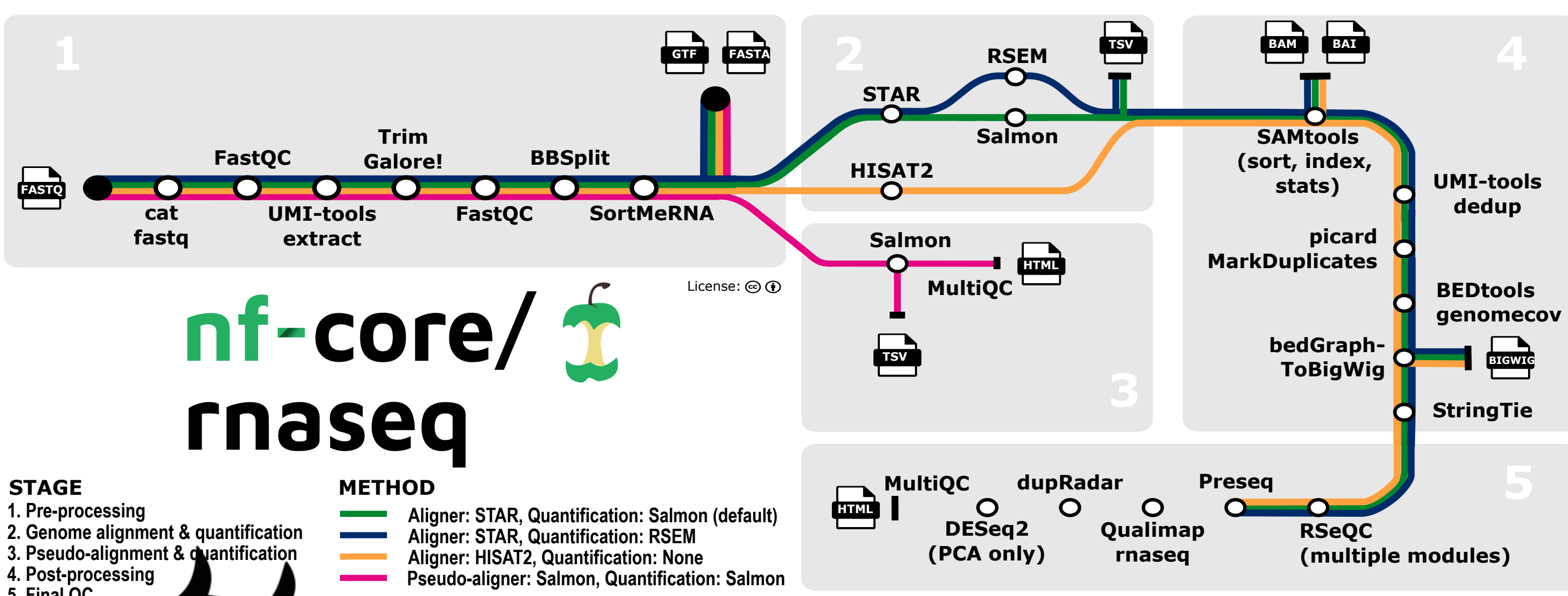
Best-practice pipelines are available to be deployed on virtually any computational infrastructure. Community-built tools help pipeline developers to create new pipelines and adhere to nf-core guidelines. Slack, Twitter and events such as hackathons allow both users and developers to actively participate in the nf-core² community.



BovReg partners typically use or extend existing nf-core pipelines even though the development of novel, nf-core compliant pipelines is supported



Example pipeline



BovReg analysis and reference pipelines

Analysis	In nf-core?	Note
ChIP-seq	✓	-
ATAC-seq	✓	no support for control
RNA-seq	✓	no StringTie transcript identification and quantification; no lncRNA annotation
CAGE-seq	✗	under active development
circRNA	✗	under active development
Hi-C	✓	-
smRNA-seq	✓	-
...		

Computational best practices like the nf-core pipeline development standard aid reference pipelines produced by BovReg to become a long-lasting resource for the community. Other EuroFAANG projects like



and adopted this approach, resulting in a collegiate developers' working group.

Conclusion

References

1) Di Tommaso et al. 2017. "Nextflow Enables Reproducible Computational Workflows." Nature Biotechnology 35 (4): 316-19.
 2) Ewels, Philip A. et al. 2020. "The Nf-Core Framework for Community-Curated Bioinformatics Pipelines." Nature Biotechnology 38 (3): 276-78.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815668.

