

IARC nextflow pipelines: towards efficient cancer genomics analyses

Tiffany Delhomme

PhD student at IARC(WHO)

22 nov. 2018

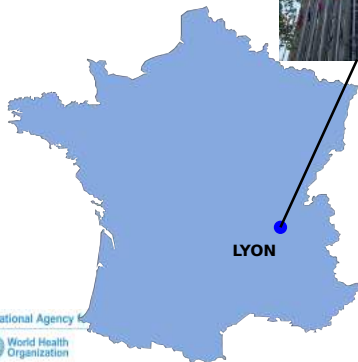
Who Am I?

- 1 / 9

Presentation

IARC Lyon

International Agency for Research on Cancer (IARC)



Presentation

IARC Lyon

International Agency for Research on Cancer (IARC)

- Created in 1965, intergovernmental agency forming part of the World Health Organization



Presentation

IARC Lyon

International Agency for Research on Cancer (IARC)



- Created in 1965, intergovernmental agency forming part of the World Health Organization
- Mostly known for:



GLOBOCAN: cancer statistics



**BLUE BOOKS:
classification of tumors**



Cancer genomics at IARC

Multiple teams, multiple projects

International Agency for Research on Cancer



World Health
Organization

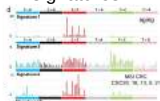
Cancer genomics at IARC

Multiple teams, multiple projects

**Mechanisms of
Carcinogenesis**

Molecular
Mechanisms and
Biomarkers

environmental
exposures
and mutational
signatures

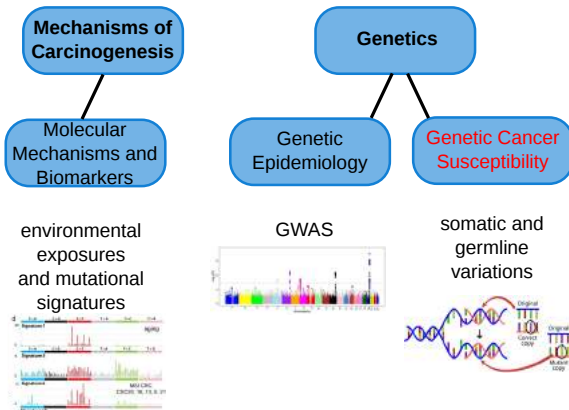


International Agency for Research on Cancer



Cancer genomics at IARC

Multiple teams, multiple projects

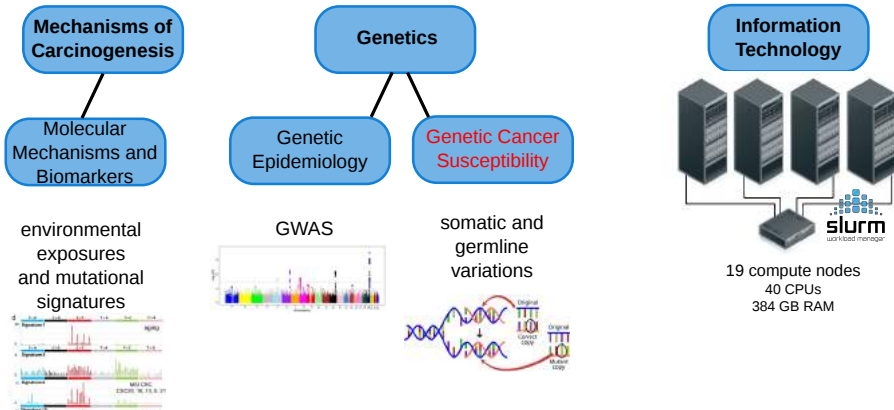


International Agency for Research on Cancer



Cancer genomics at IARC

Multiple teams, multiple projects



International Agency for Research on Cancer



Why nextflow?

For cancer genomics analyses

Nextflow enables scientists to write workflow emphasizing on:

- **efficiency**: parallel computations \Rightarrow analyses of hundreds of cancer genomes

International Agency for Research on Cancer



Why nextflow?

For cancer genomics analyses

Nextflow enables scientists to write workflow emphasizing on:

- **efficiency**: parallel computations \Rightarrow analyses of hundreds of cancer genomes
- **robustness**: scalable computations \Rightarrow multiple HPC environments + local machines

Why nextflow?

For cancer genomics analyses

Nextflow enables scientists to write workflow emphasizing on:

- **efficiency**: parallel computations \Rightarrow analyses of hundreds of cancer genomes
- **robustness**: scalable computations \Rightarrow multiple HPC environments + local machines
- **reproducibility**: GitHub + Docker + Singularity \Rightarrow n-uplication of analyses, validations

Why nextflow?

For cancer genomics analyses

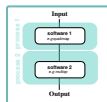
Nextflow enables scientists to write workflow emphasizing on:

- **efficiency**: parallel computations \Rightarrow analyses of hundreds of cancer genomes
- **robustness**: scalable computations \Rightarrow multiple HPC environments + local machines
- **reproducibility**: GitHub + Docker + Singularity \Rightarrow n-uplication of analyses, validations
- **user-friendliness**: 1 command line, no software installation \Rightarrow user profile palette is large (biologists, students, pathologists,...)

Implementation pattern

Efficiency, robustness, user-friendliness and reproducibility

nextflow



source code

source code is stored and
versioned on  GitHub



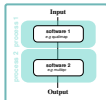
user runs the pipeline

`$ nextflow run pipeline.options`

Implementation pattern

Efficiency, robustness, user-friendliness and reproducibility

nextflow



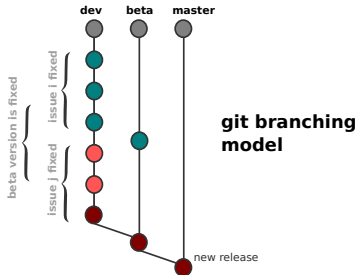
source code

source code is stored and versioned on  GitHub



user runs the pipeline

\$ nextflow run pipeline options

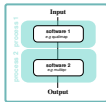


git branching model

Implementation pattern

Efficiency, robustness, user-friendliness and reproducibility

nextflow



source code

source code is stored and versioned on  GitHub



user runs the pipeline

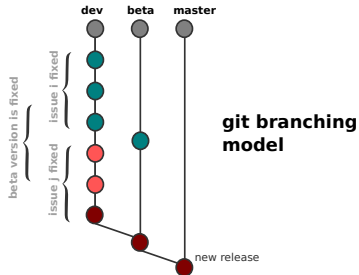
\$ nextflow run pipeline options



dockerhub

SINGULARITYHUB

containerization



git branching model

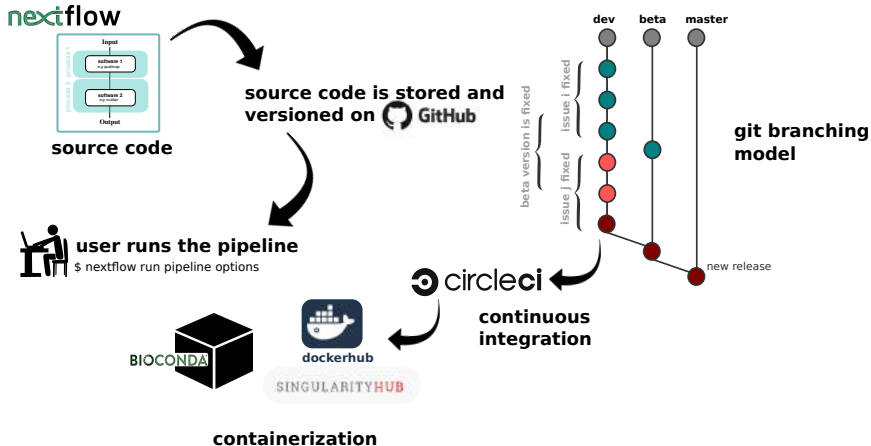
International Agency for Research on Cancer



World Health Organization

Implementation pattern

Efficiency, robustness, user-friendliness and reproducibility



International Agency for Research on Cancer



GitHub IARCbionfo

open-source pipelines

► [GitHub webpage](#)

International Agency for Research on Cancer



World Health
Organization

ITH pipeline example

Intra-tumor heterogeneity leads to multiple subclones

Genetic variation within individual tumors (ITH) provides insights into:

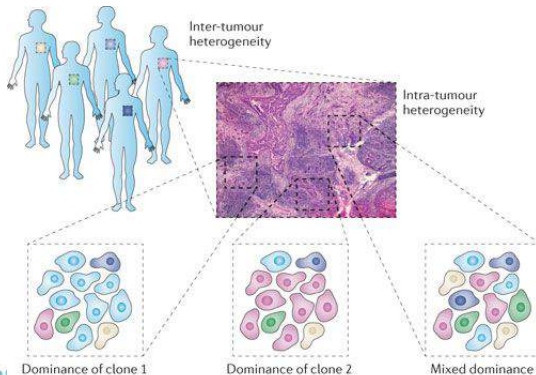
- cancer evolutionary trajectories
- cancer resistance to therapies
- cancer molecular classifications

ITH pipeline example

Intra-tumor heterogeneity leads to multiple subclones

Genetic variation within individual tumors (ITH) provides insights into:

- cancer evolutionary trajectories
- cancer resistance to therapies
- cancer molecular classifications



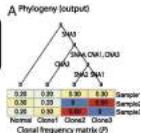
International Agency for Research on



Nature Reviews | Cancer

In practice (with L. Soudade and I. Lboukili)

In practice (with L. Soudade and I. Lboukili)



Acknowledgements

My supervisor **Matthieu Foll**



IARC/GCS group

Nicolas Alcala

Aurelie Gabriel
Noemie Leblay
Catherine Voegele



part. Alex Peltzer

IARC/other groups

Maxime Vallee

Alexis Robitaille
Vincent Cahais
Claire Renard
Nicolas Tardy

nextflow community

Paolo Di Tommaso
Evan Floden



La Ligue Nationale
Contre le Cancer
(PhD fellowship
2015-2018)



International Agency for Research on Cancer



delhommet@students.iarc.fr



@tm_delhomme



tdelhomme