



IARC nextflow pipelines: towards efficient cancer genomics analyses

Tiffany Delhomme

PhD student at IARC(WHO)

22 nov. 2018

Presentation

Who Am I?

- M.D Evolution, Bioinfo, Biomaths from Universite Lyon 1, France (2014)
- M.D Theoretical Computer Science - Complex Systems from ENS Lyon and IXXI institute, France (2015)
- **current:** 4th year PhD student in Bioinformatics and Cancer Genomics at IARC, Lyon, France
 - *Dealing with NGS errors to produce efficient variant calling.
Application to early cancer detection.*

Presentation

IARC Lyon

International Agency for Research on Cancer (IARC)



International Agency for



Presentation

IARC Lyon

International Agency for Research on Cancer (IARC)



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



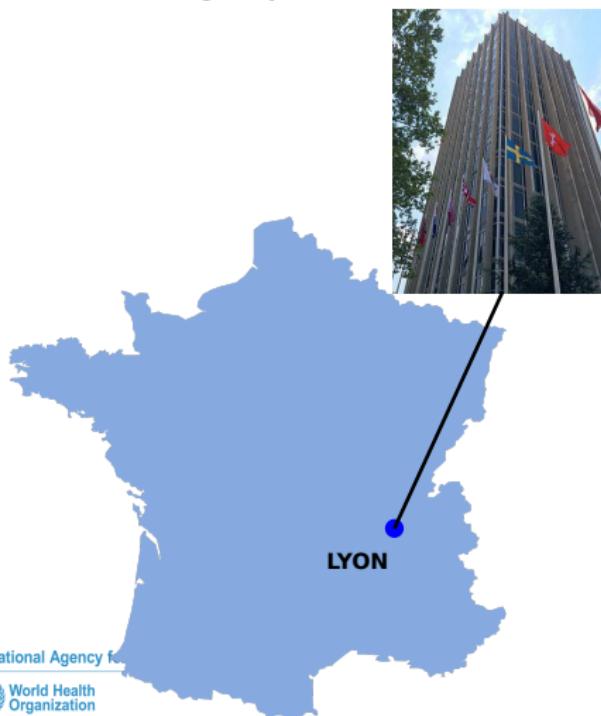
International Agency for



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:



International Agency for

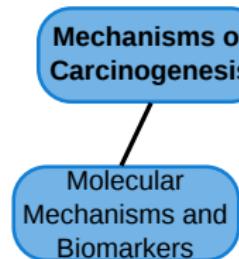


Cancer genomics at IARC

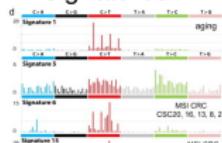
Multiple teams, multiple projects

Cancer genomics at IARC

Multiple teams, multiple projects



environmental
exposures
and mutational
signatures

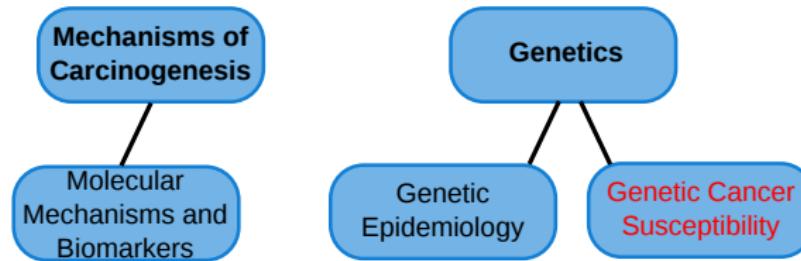


International Agency for Research on Cancer

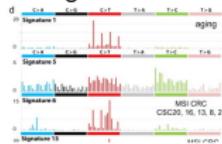


Cancer genomics at IARC

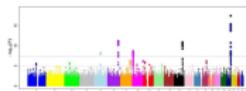
Multiple teams, multiple projects



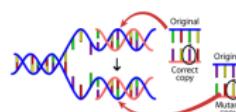
environmental
exposures
and mutational
signatures



GWAS

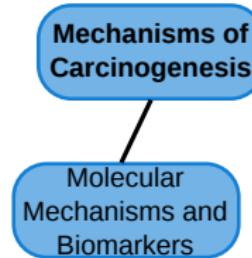


somatic and
germline
variations

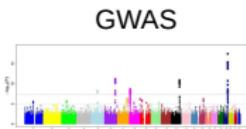
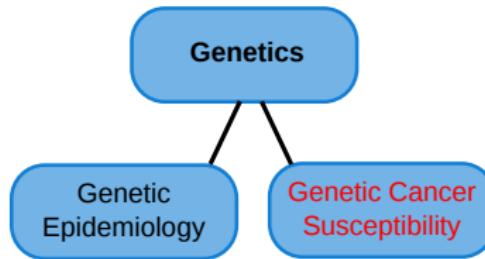
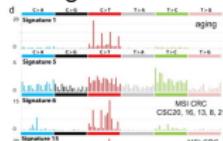


Cancer genomics at IARC

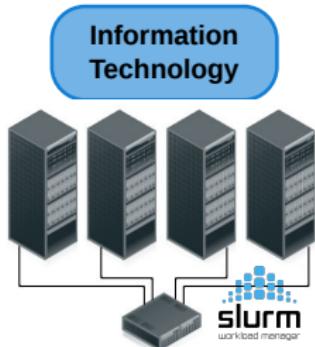
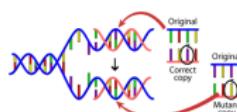
Multiple teams, multiple projects



environmental exposures and mutational signatures



somatic and germline variations



19 compute nodes
40 CPUs
384 GB RAM

Why nextflow?

For cancer genomics analyses

Nextflow enables scientists to write workflow emphasizing on:

- **efficiency:** parallel computations ⇒ analyses of hundreds of cancer genomes

Why nextflow?

For cancer genomics analyses

Nextflow enables scientists to write workflow emphasizing on:

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

Why nextflow?

For cancer genomics analyses

Nextflow enables scientists to write workflow emphasizing on:

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

Why nextflow?

For cancer genomics analyses

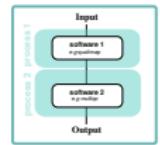
Nextflow enables scientists to write workflow emphasizing on:

- **efficiency**: parallel computations ⇒ analyses of hundreds of cancer genomes
- **robustness**: scalable computations ⇒ multiple HPC environments + local machines
- **reproducibility**: GitHub + Docker + Singularity ⇒ n-uplication of analyses, validations
- **user-friendliness**: 1 command line, no software installation ⇒ 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

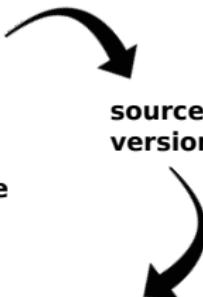
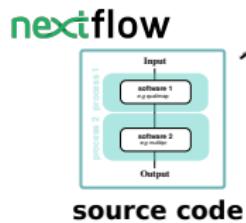


user runs the pipeline

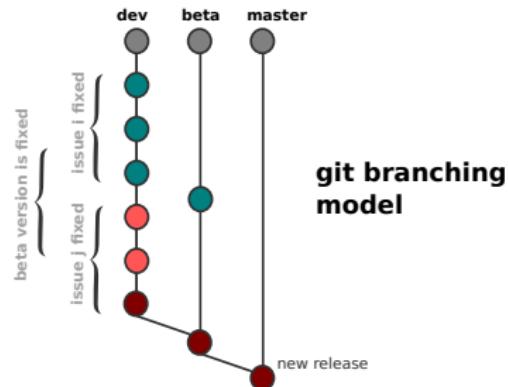
\$ nextflow run pipeline options

Implementation pattern

Efficiency, robustness, user-friendliness and reproducibility

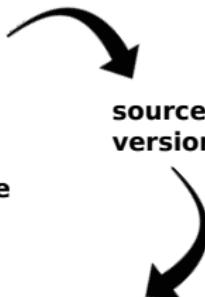
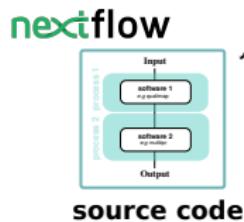


 user runs the pipeline
\$ nextflow run pipeline options



Implementation pattern

Efficiency, robustness, user-friendliness and reproducibility



source code is stored and versioned on GitHub



user runs the pipeline
\$ nextflow run pipeline options



SINGULARITY HUB

containerization

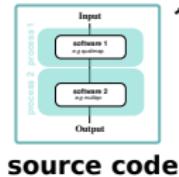
International Agency for Research on Cancer



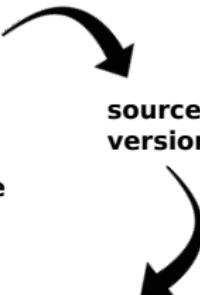
Implementation pattern

Efficiency, robustness, user-friendliness and reproducibility

nextflow



source code



source code is stored and versioned on GitHub

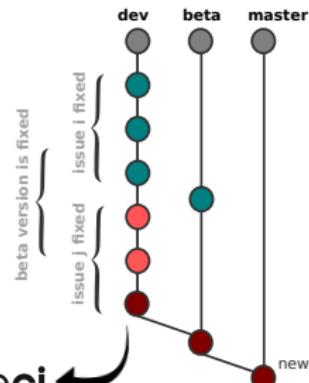


user runs the pipeline

\$ nextflow run pipeline options



continuous integration



git branching model



SINGULARITY HUB

containerization

International Agency for Research on Cancer



World Health Organization

GitHub IARCbionfo

open-source pipelines

► GitHub webpage

International Agency for Research on Cancer



ITH pipeline example

Intra-tumor heterogeneity leads to multiple subclones

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

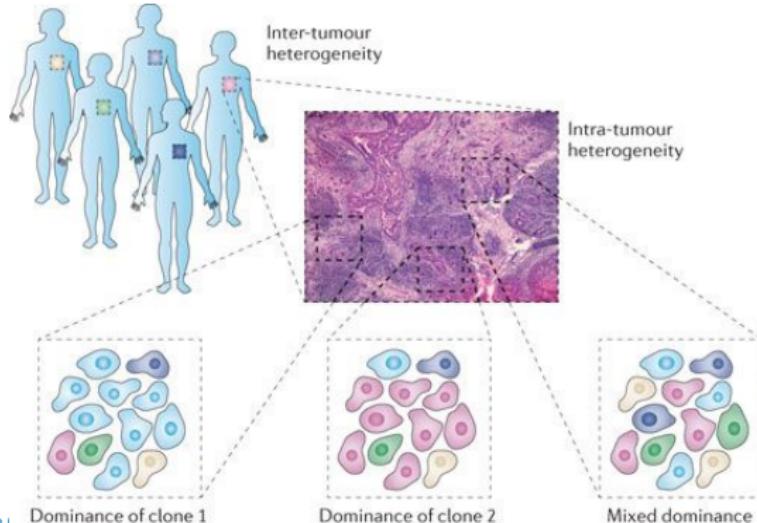
- cancer evolutionary trajectories
- cancer resistance to therapies
- cancer molecular classifications

ITH pipeline example

Intra-tumor heterogeneity leads to multiple subclones

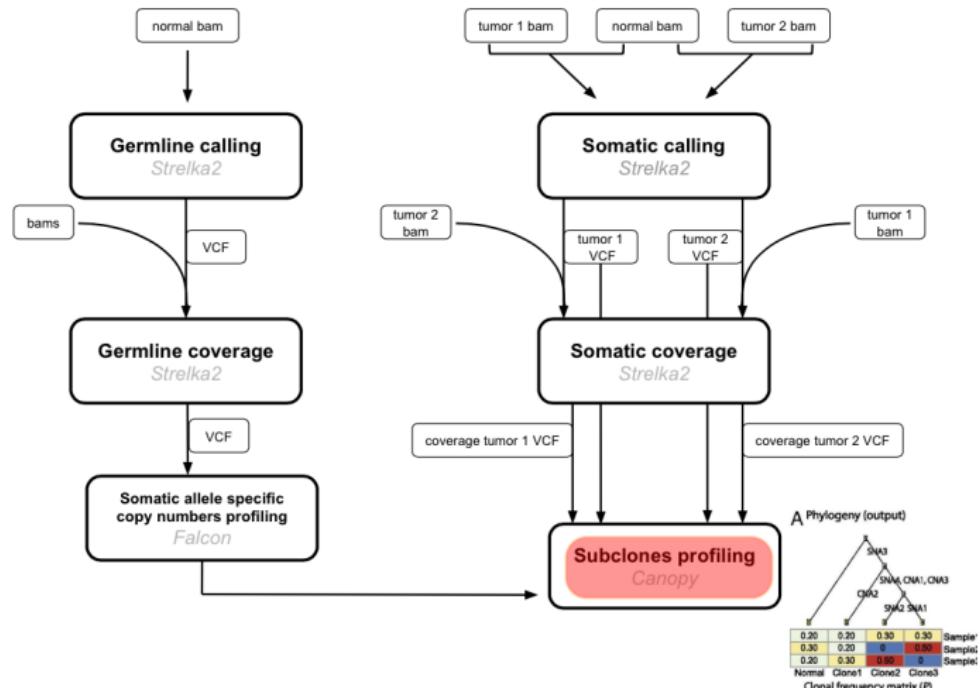
Genetic variation within individual tumors (ITH) provides insights into:

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



ITH pipeline example

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)



delhommet@students.iarc.fr



@tm_delhomme



tdelhomme