

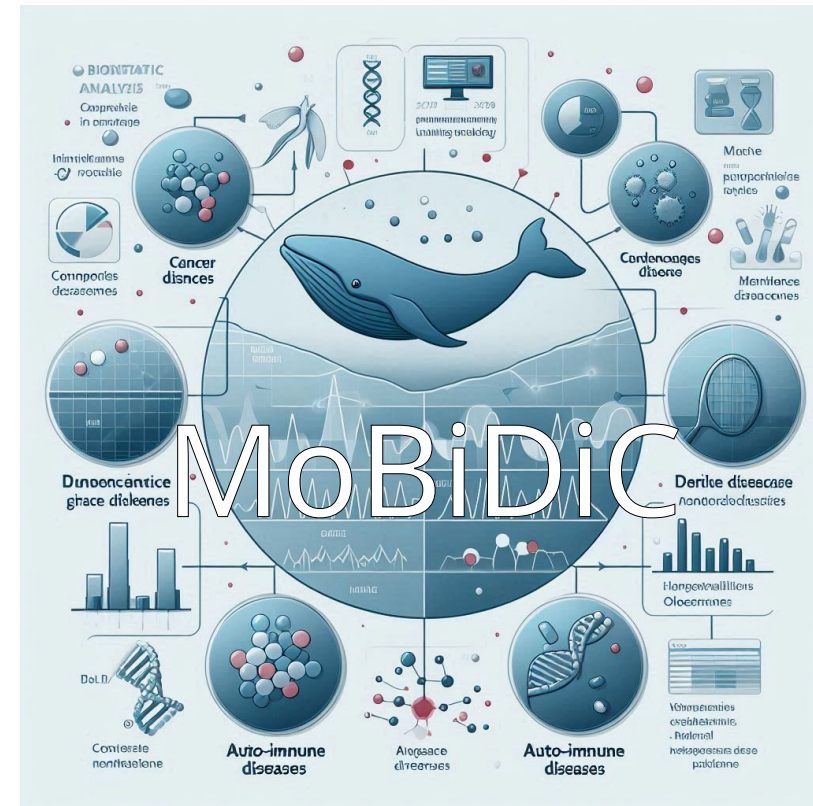
MobiCT: un pipeline open-source pour l'analyse de l'ADN tumoral circulant

SÉMINAIRE BIOINFODIAG 2025

SIMON CABELLO-AGUILAR, PHD
LABORATOIRE DE BIOLOGIE DES TUMEURS SOLIDES
CHU MONTPELLIER

Sommaire

- Introduction
 - Rationnel
 - L'ADN libre circulant
 - L'ADN tumoral circulant
- Le pipeline: MobiCT
 - Deduplication par UMI
 - Variant calling
 - Interface d'interprétation
- Performances
- Conclusions & perspectives



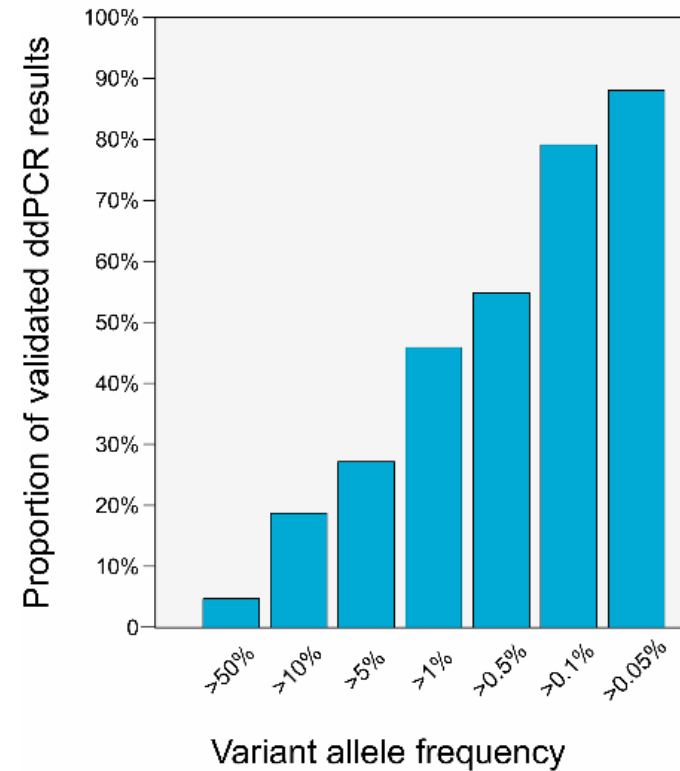
Introduction

Pourquoi?

- Détection SNV / petits indels
- Activité effectuée au laboratoire en ddPCR
- Augmentation activité
- Nouvelles indications
- Passage en NGS

Comment?

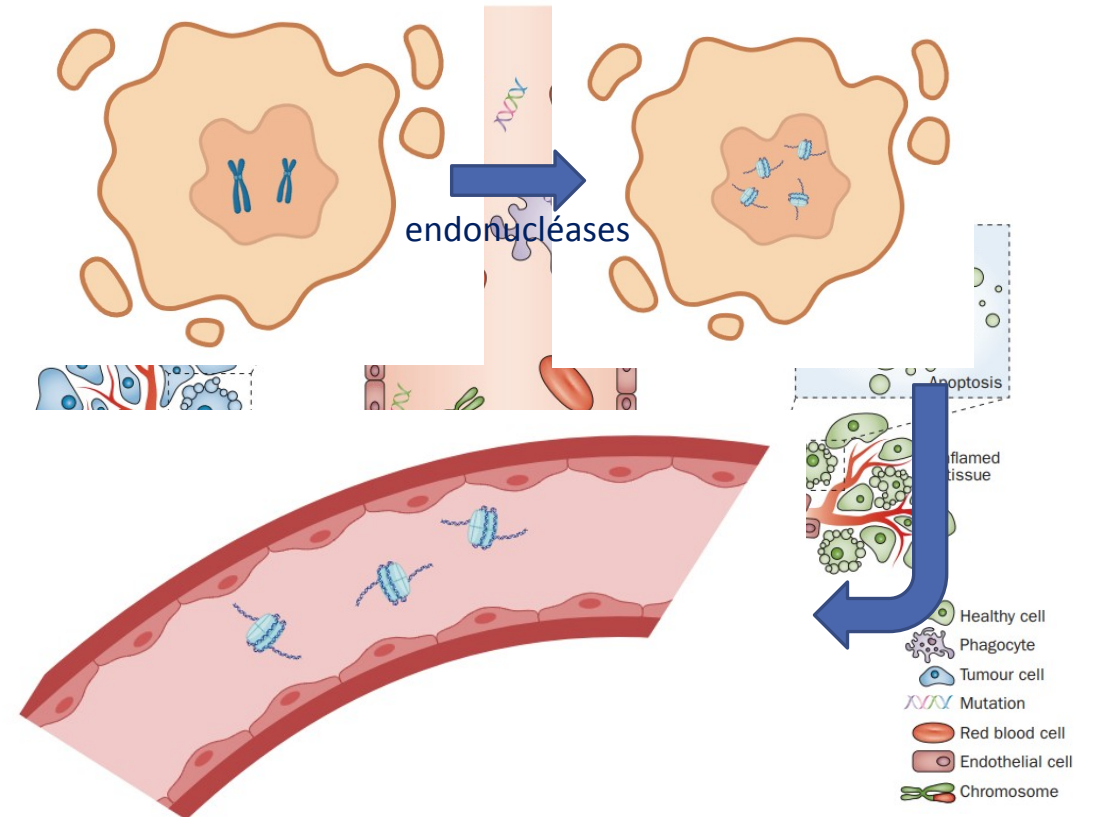
- Panel Twist avec UMI
- Très faibles VAF ($>0,1\%$)
- Très grande profondeur



L'ADN libre circulant

ADN libre circulant (cfDNA):

- 100 milliards cellules nucléées / jour
- Peu de quantité (1000 genome equivalents / mL plasma)
- Demi-vie 15-100min
- Variable selon les types cellulaires
- Nucleosomes

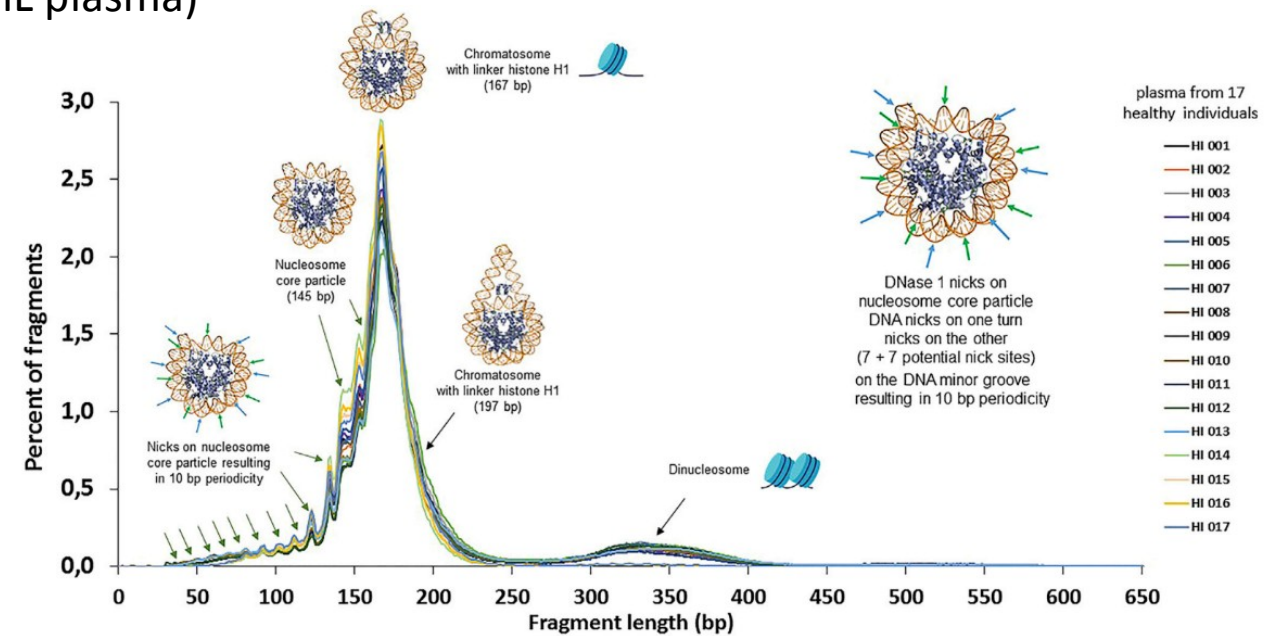


Crowley et al., *Nat Rev Clin Onc*, 2013
Sender and Milo, *Nature Medicine*, 2021
Che et al., *EVCNA*, 2022
Loyfer et al., *Nature*, 2023
Sender et al. *eLife*, 2023

L'ADN libre circulant

ADN libre circulant (cfDNA):

- 300 milliards cellules / jour
- Peu de quantité (1000 genome equivalents / mL plasma)
- Demi-vie 15-100min
- Variable selon les types cellulaires
- Nucleosomes
- Taille des fragments

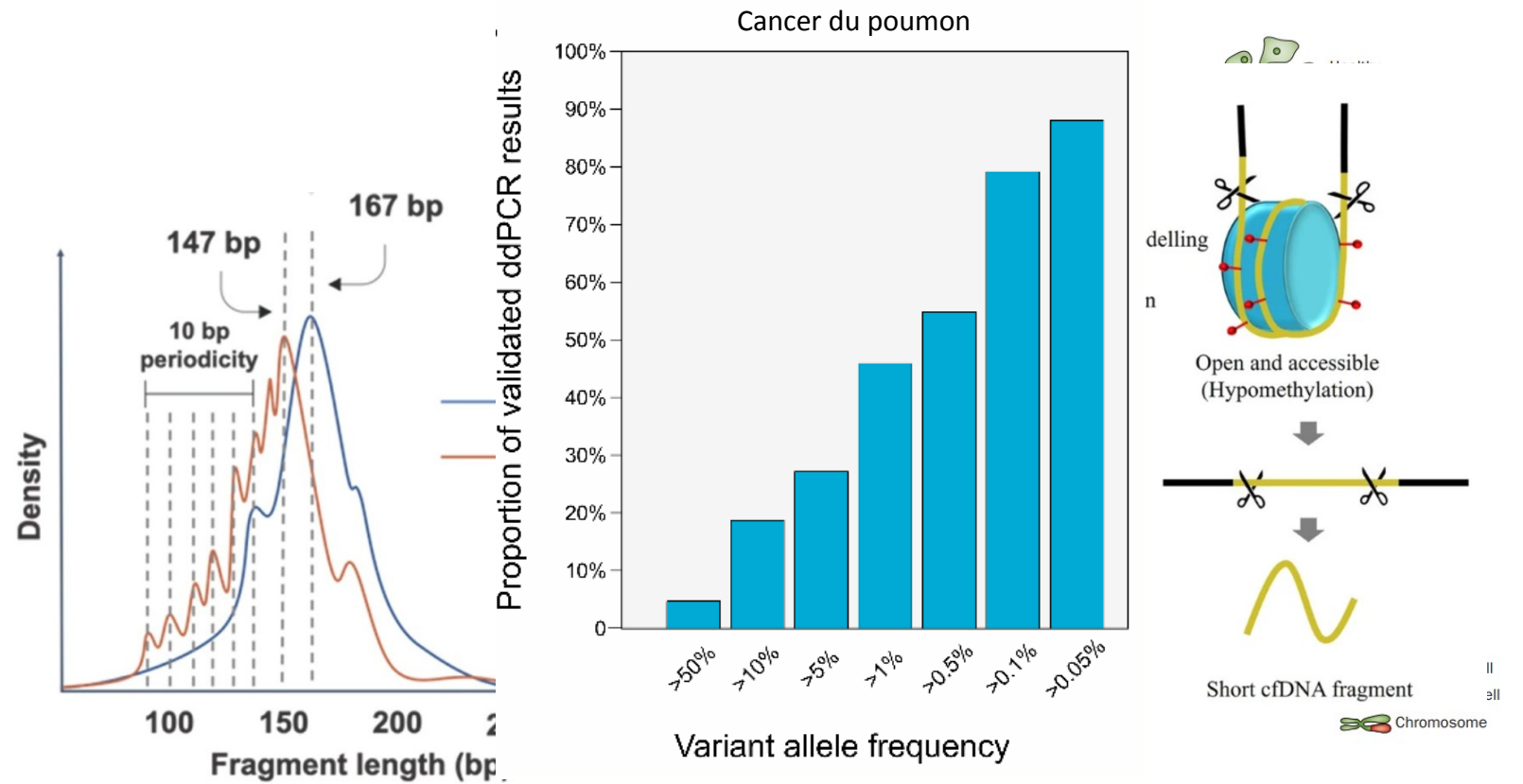


Crowley et al., *Nat Rev Clin Onc*, 2013
Sender and Milo, *Nature Medicine*, 2021
Che et al., *EVCNA*, 2022
Loyfer et al., *Nature*, 2023
Sender et al. *eLife*, 2023

L'ADN tumoral circulant

ADN tumoral circulant (ctDNA):

- Une fraction du cfDNA
- Hétérogénéité clonale
- Dépend du stade
- Dépend du type de tumeur
- Taille des fragments
- Paradoxes (origine, proportion ~50x « trop »)

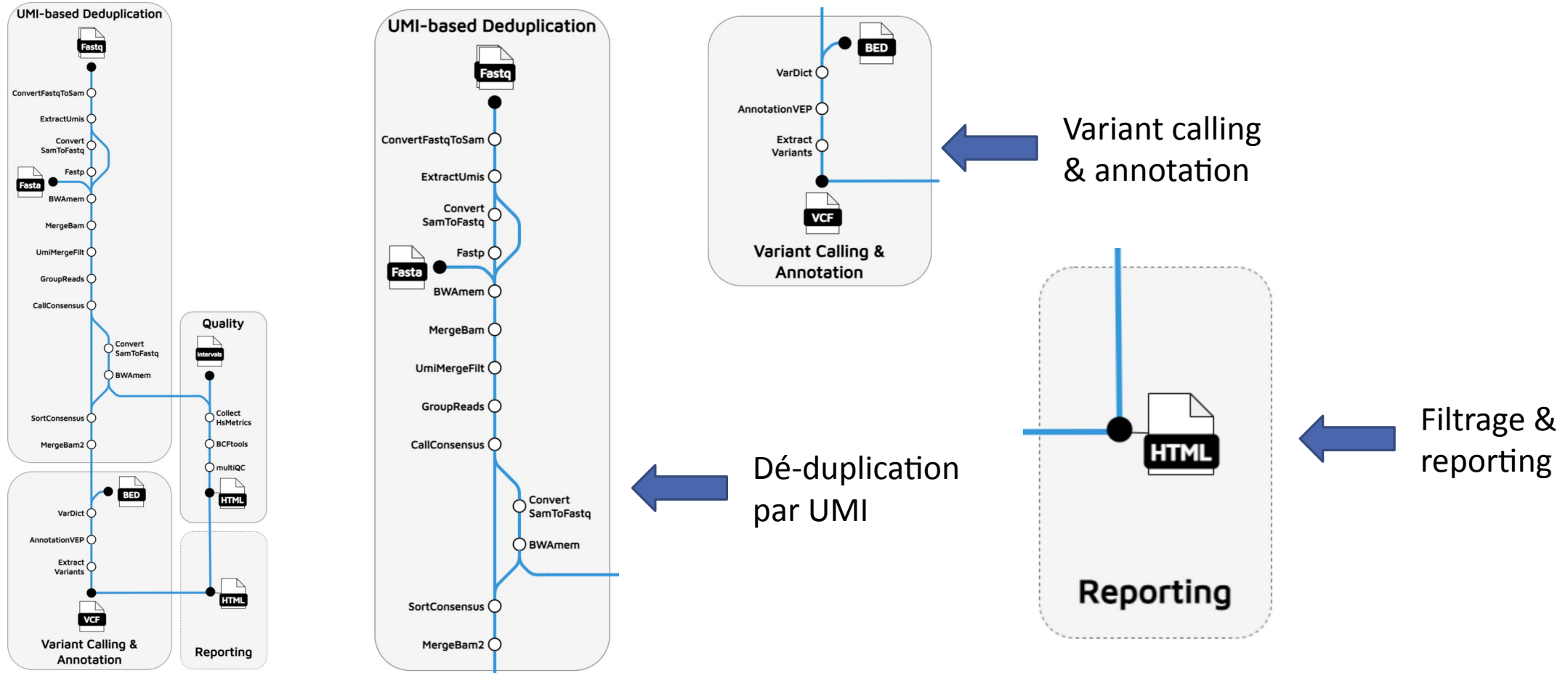


Crowley et al., *Nat Rev Clin Onc*, 2013

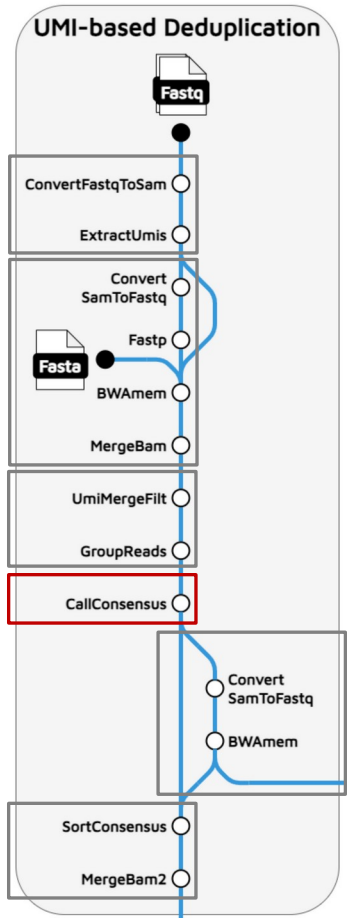
Udomruk et al., *Crit Rev Onc/Hem*, 2021

Wang et al., *Epi & Chrom*, 2023

Le pipeline: MobiCT



Dé-duplication par UMI



Meilleurs résultats:

- Le plus de « reads »
- La VAF la plus proche de la référence

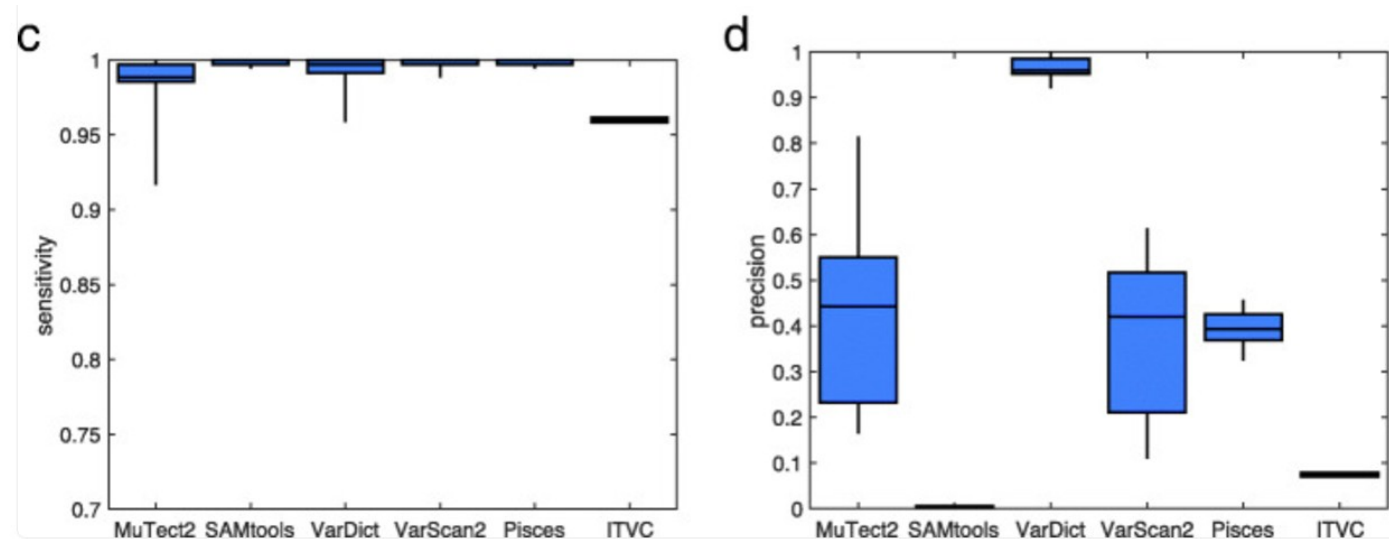
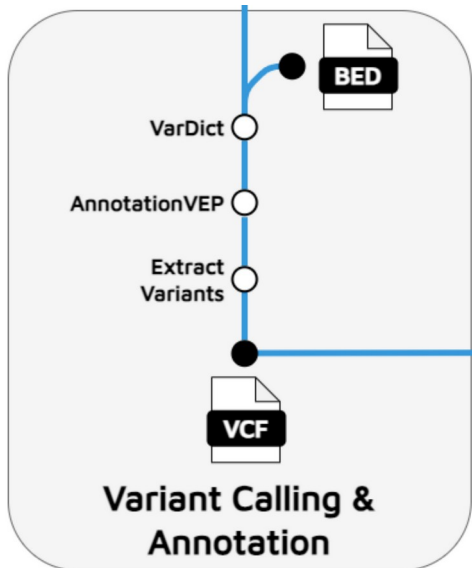
`fgbio::CallMolecularConsensusReads` <https://fulcrumgenomics.github.io/fgbio/tools/latest/CallMolecularConsensusReads.html>

Appelle des séquences consensuelles à partir des reads partageant le même UMI.

Les reads ayant le même UMI sont analysés base par base afin d'évaluer la probabilité de chaque base dans la molécule source. Le modèle de probabilité est le suivant :

2. Ensuite, la qualité de la base consensus a été ajustée en intégrant la probabilité d'une erreur. On anticipe que la base consensus est correcte, mais il est possible que la base consensus soit erronée. La probabilité de la base consensus est donc ajustée en fonction de la probabilité d'une erreur. Les probabilités calculées sont ensuite normalisées en les divisant par la somme des quatre probabilités, afin d'obtenir une probabilité a posteriori, qui représente la probabilité que la molécule source ait été un A, C, G ou T juste après l'intégration de l'UMI et jusqu'au séquençage, en tenant compte des observations. La base présentant la probabilité a posteriori maximale est choisie comme base consensuelle, et cette probabilité est utilisée comme la qualité de cette base consensuelle.

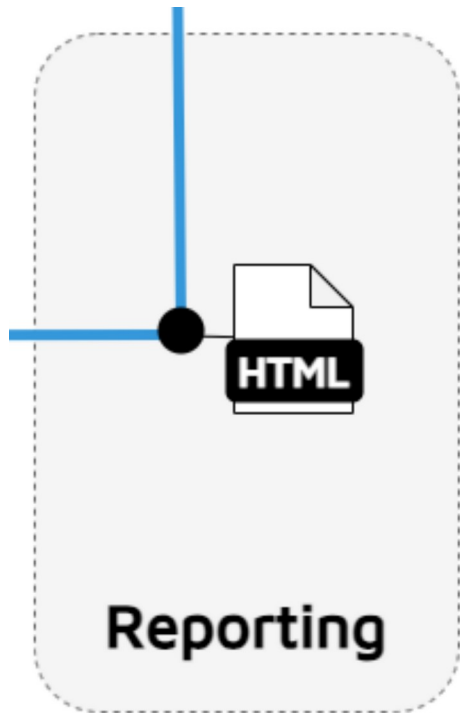
Variant calling & annotation



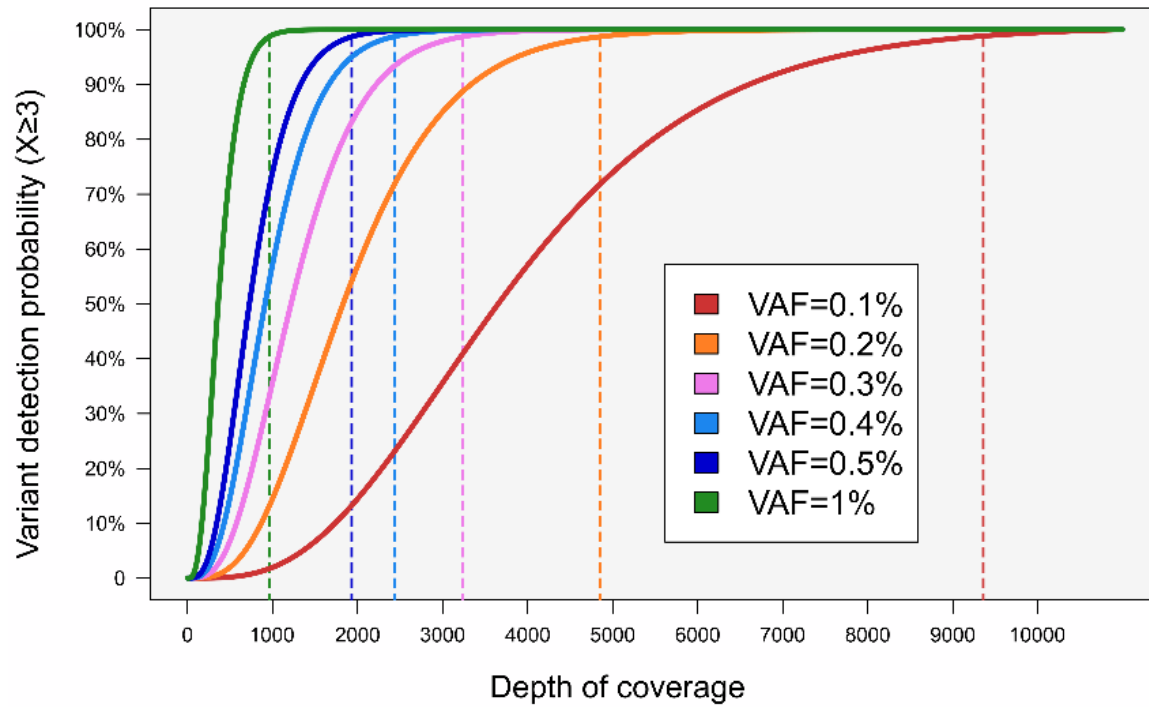
→ Utilisation de VarDict et VEP pour l'annotation

Karimnezhad et al., *BMC Med Gen*, 2020.

Interface d'interprétation



→ Comment rendre les résultats négatifs ?



Limite de détection:
VAF théorique minimale pour laquelle la probabilité de détecter au moins 3 reads mutés est > 0.99 (LoD $p < 0.01$)

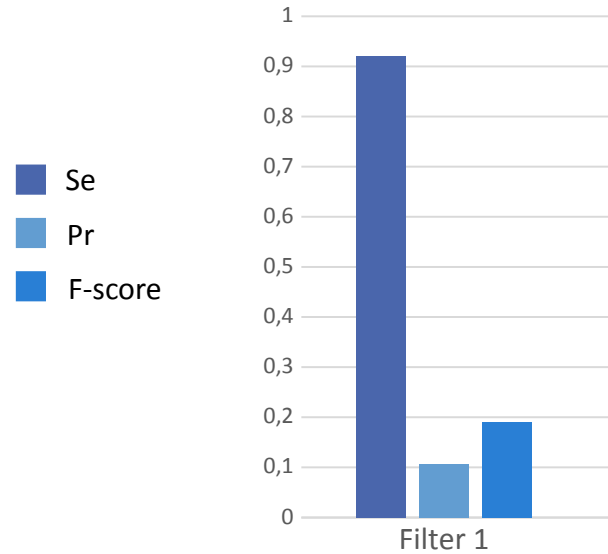
/hitelist	VD	DP	chrom	pos	ref
	89	6907	chr7	55174771	AGGAATTAAGAG
	59	6389	chr2	29222410	T
	46	6109	chr17	7675119	G
	8	1417	chr9	136513019	A
	26	5878	chr9	136496492	TGTG

Performances

5 stratégies de filtrage:

- No filter
- >3 reads
- Whitelist
- LoD
- Combinaison

19 échantillons patients
5 contrôles commerciaux



Total	1539
TP	164
FP	1375
FN	14
Se	0,921
Pr	0,107
F-score	0,191

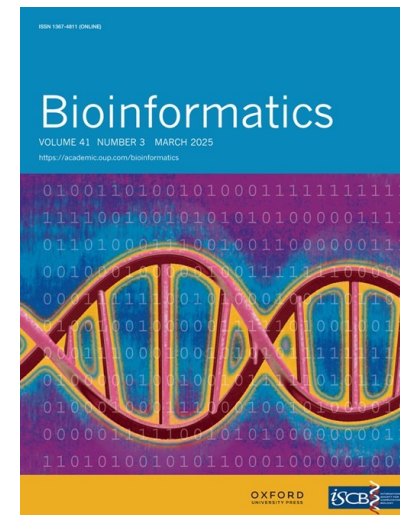
Conclusions & perspectives

Conclusions:

- Problème compliqué
- Optimisation de la déduplication
- LoD dynamique + « whitelist »
- Se et Pr $\sim 0,9$

Perspectives:

- Calcul de la fraction tumorale circulante
- ctDNA dans d'autres fluides (urine, salive, LCR, ...)



MobiCT: a UMI-based circulating tumor DNA analysis pipeline

Simon Cabello-Aguilar^{1,4}, Charles Van Goethem^{2,4,+}, Jean-Charles Delmas^{3,4,+}, Oussama Bourbia^{1,4}, Romain Senal¹, Mireille Cossée^{2,5}, Olivier Ardouin⁴, Jérôme Solassol^{1,6,*}, Julie A. Vendrell¹

→ *Under review*

Remerciements

Laboratoire de BioTS:

J. Solassol
J. Vendrell
O. bourbia
R. Senal

MERCI DE VOTRE ATTENTION

Questions?

MoBiDiC:

O. Ardouin
D. Baux
J. C. Delmas
T. Guignard
F. Vandermeeren
C. Van Goethem



s-cabelloaguiar@chu-montpellier.fr



<https://github.com/SimCab-CHU/MobiCT>