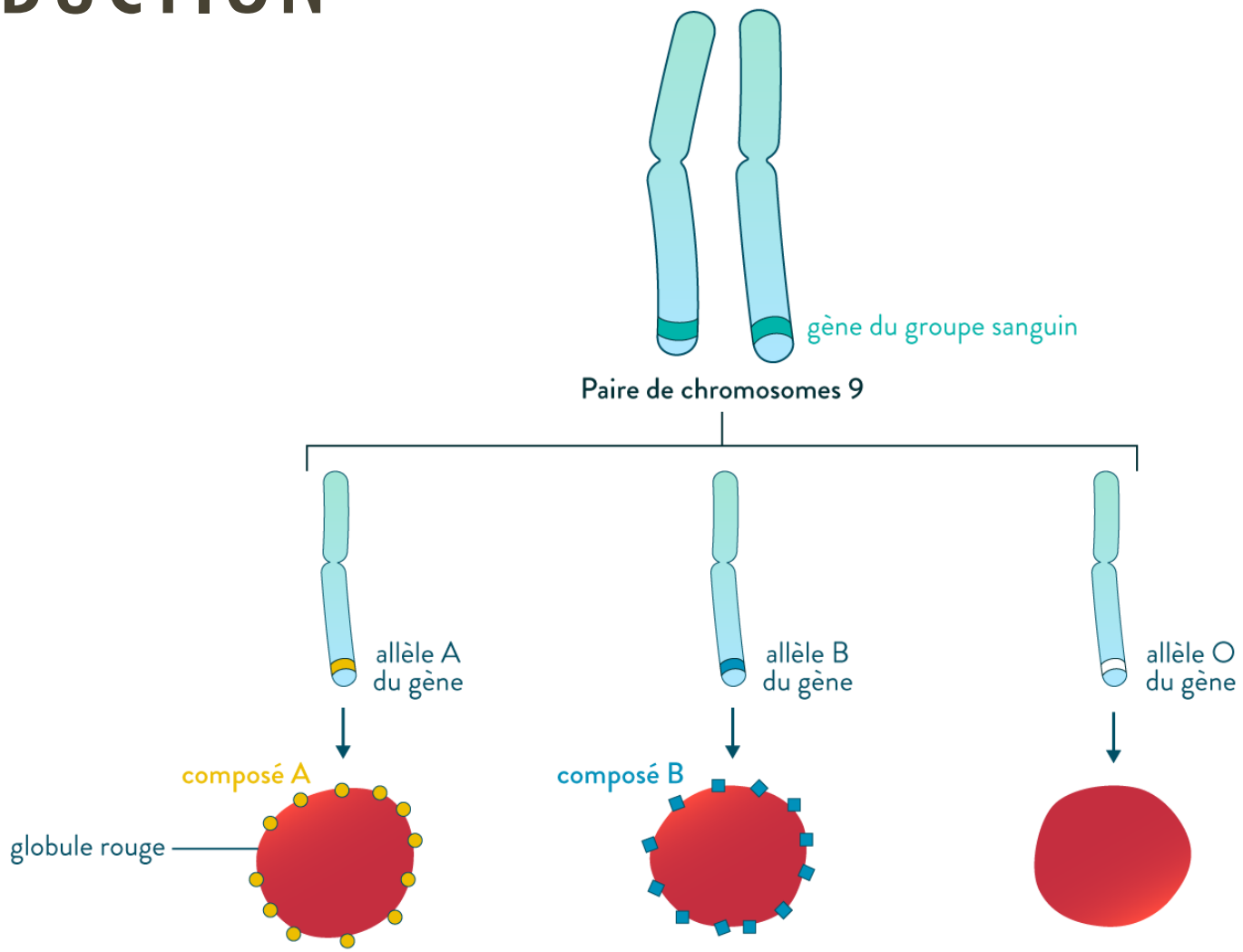
APPORT DU CALCUL HAUTE-PERFORMANCE DANS LA GÉNÉTIQUE DES MALADIES RARES

Verdez Simon,
doctorant, GAD Inserm UMR 1231

INTRODUCTION

INTRODUCTION



INTRODUCTION

Des maladies rares mais connues de tous

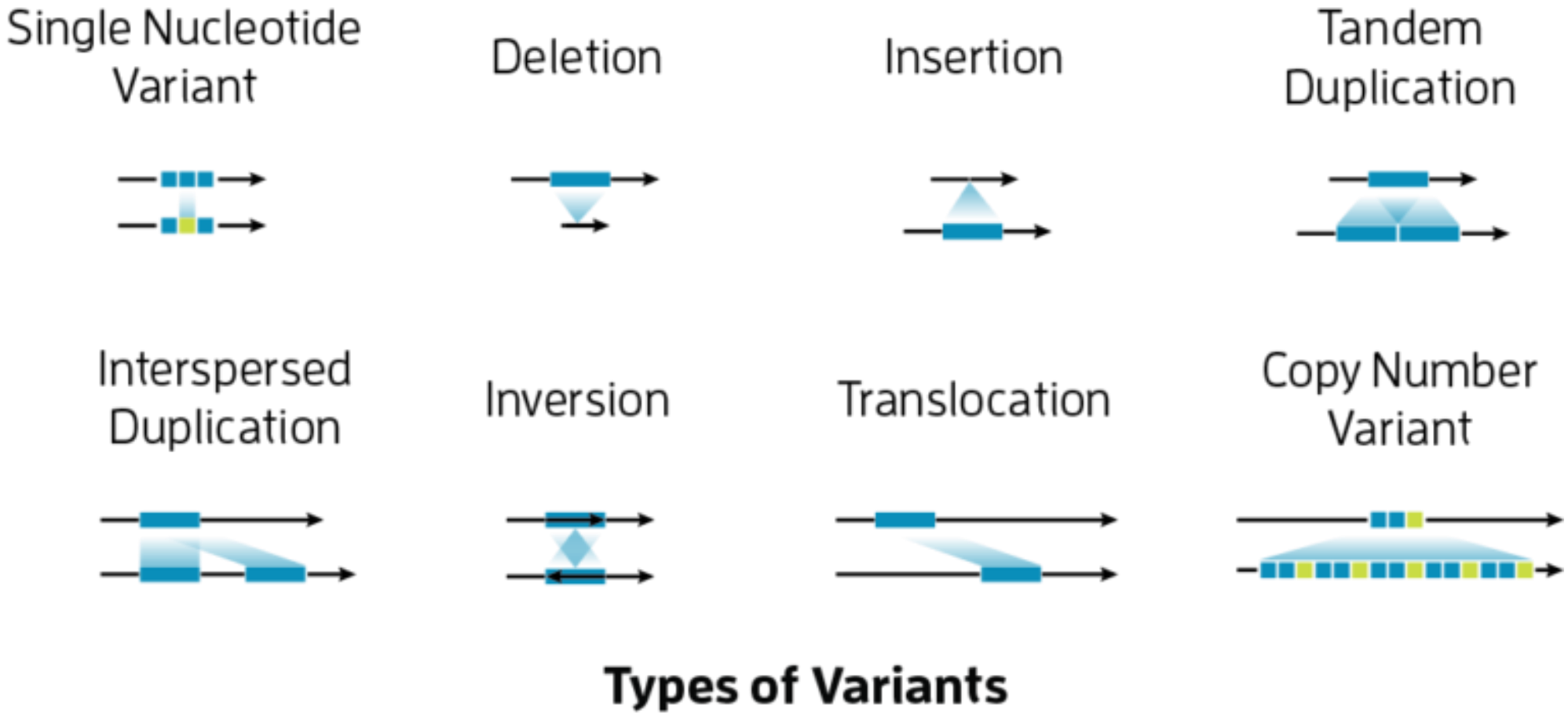
INTRODUCTION

Consultation de génétique

Consentement du patient

Analyse génomique

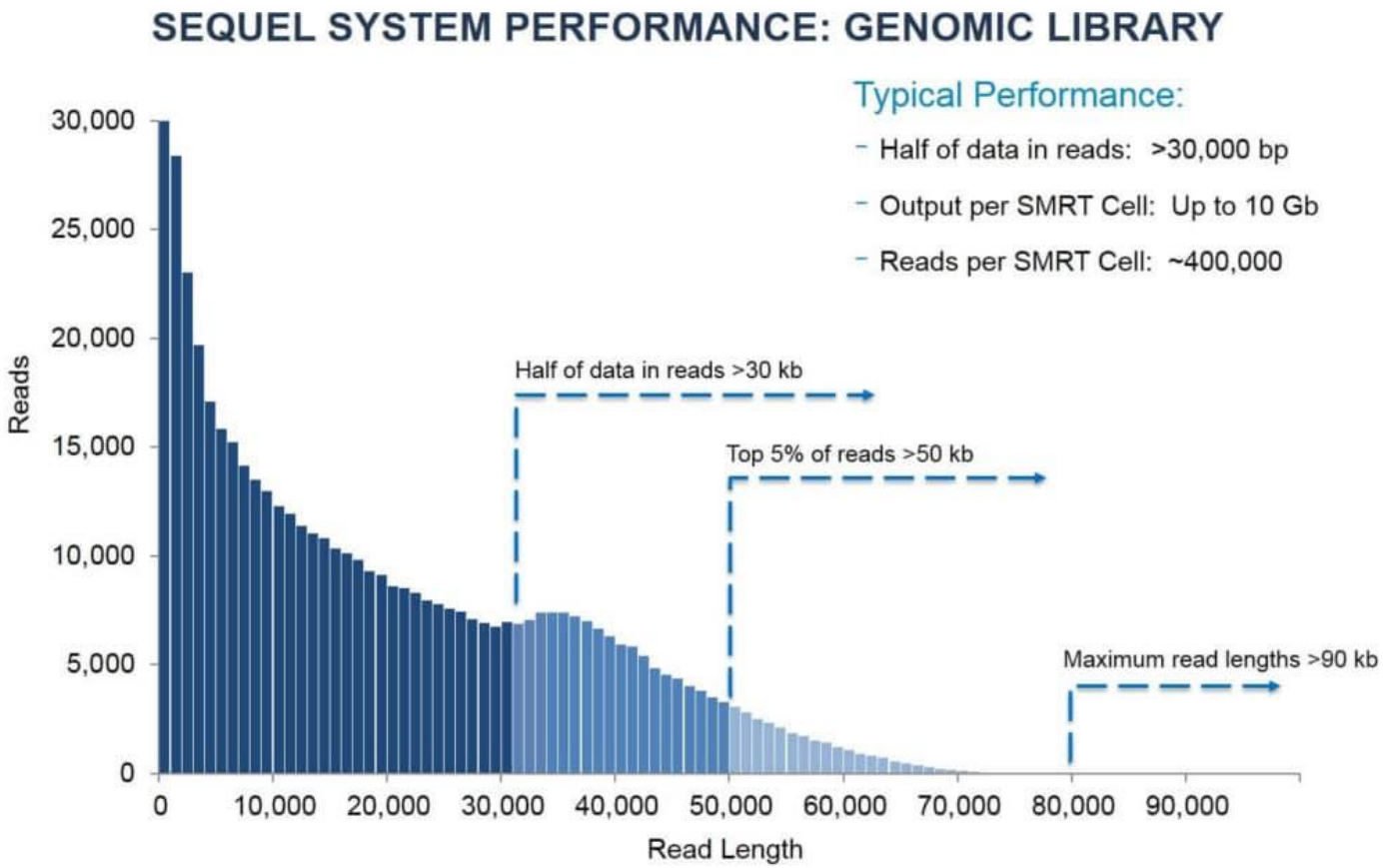
INTRODUCTION



Source : https://www.pacb.com/applications/whole-genome-sequencing/variant-detection/

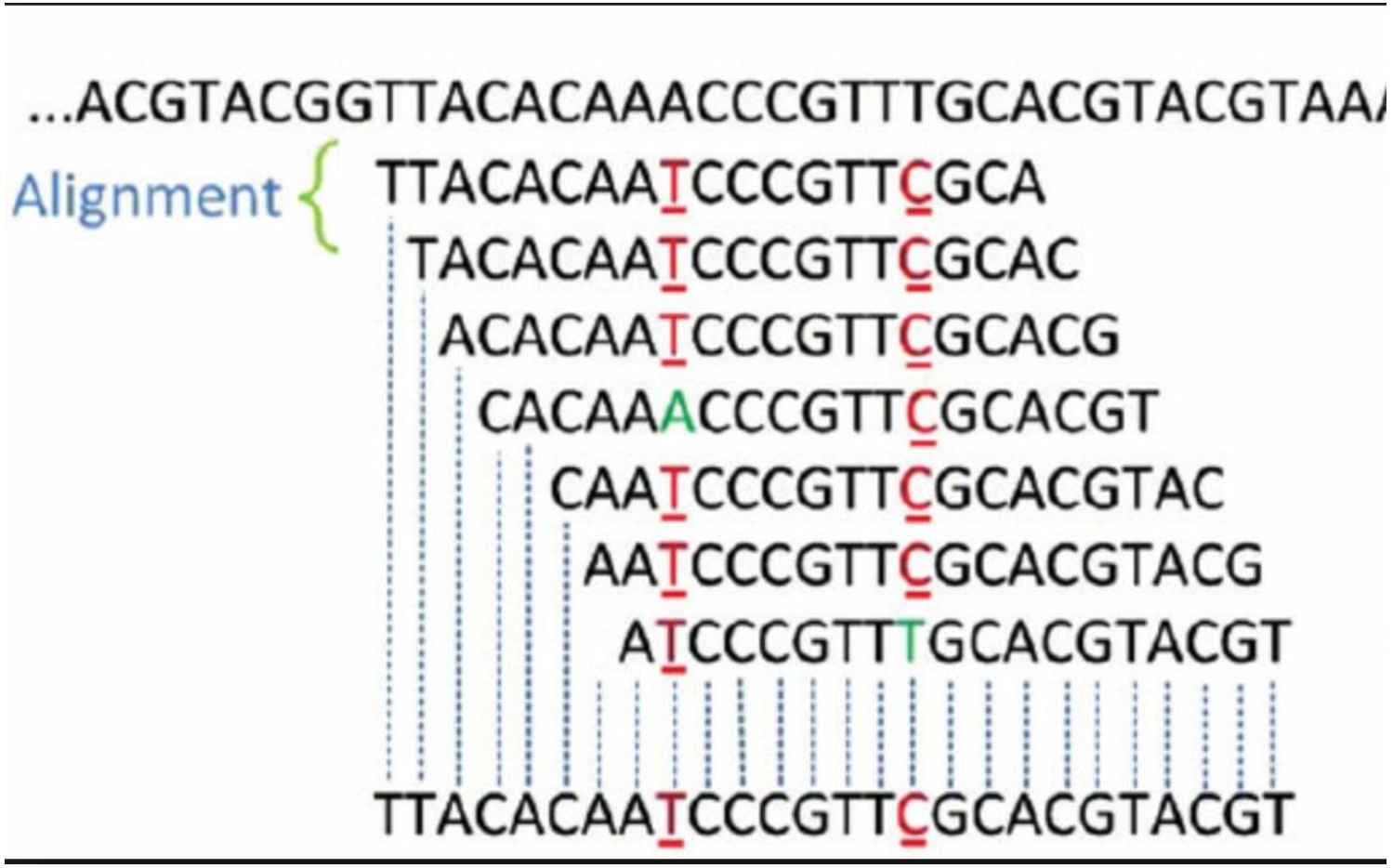
INTRODUCTION

INTRODUCTION



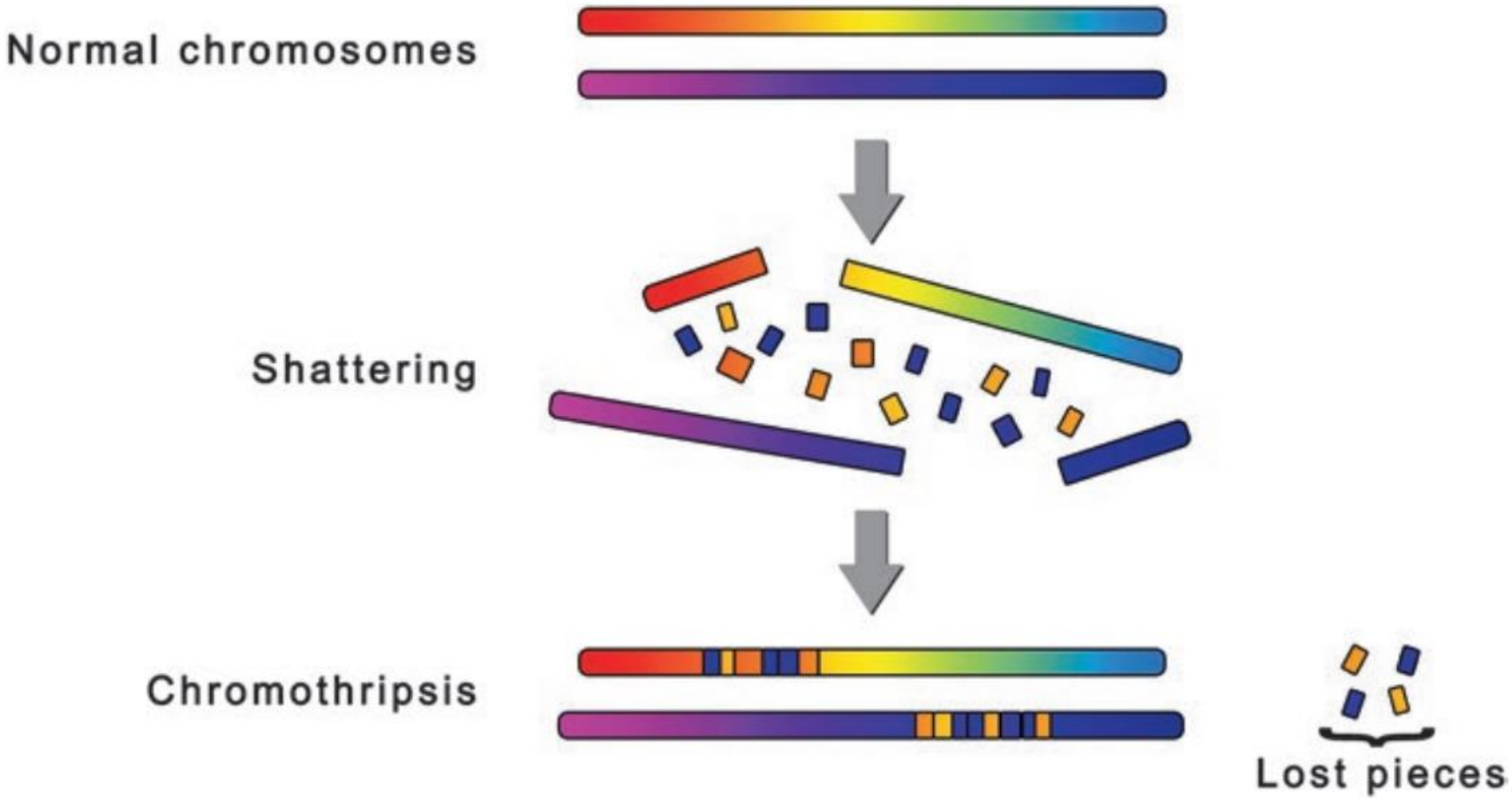
Source : pacb.com

INTRODUCTION



Séquence de référence

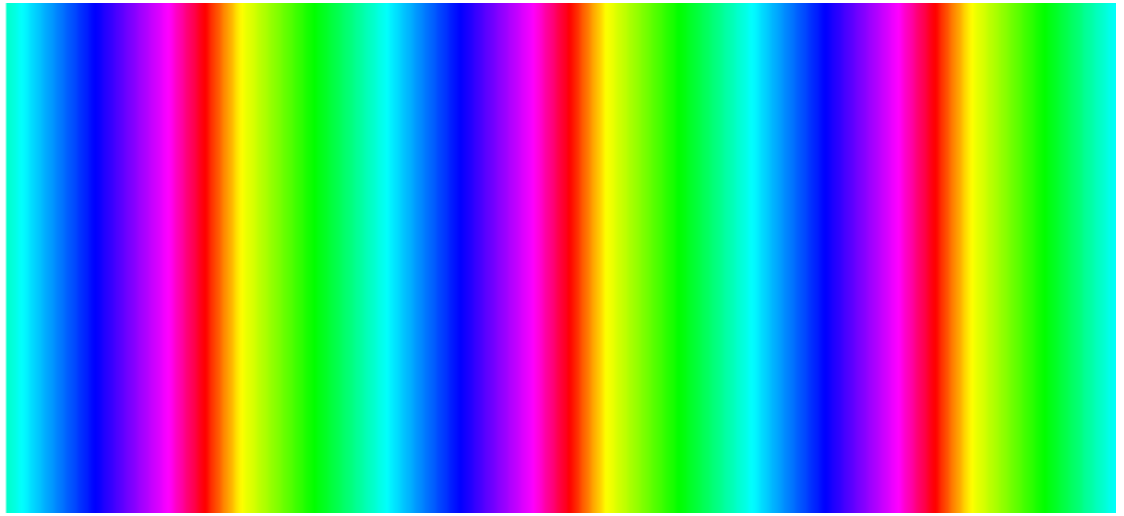
INTRODUCTION

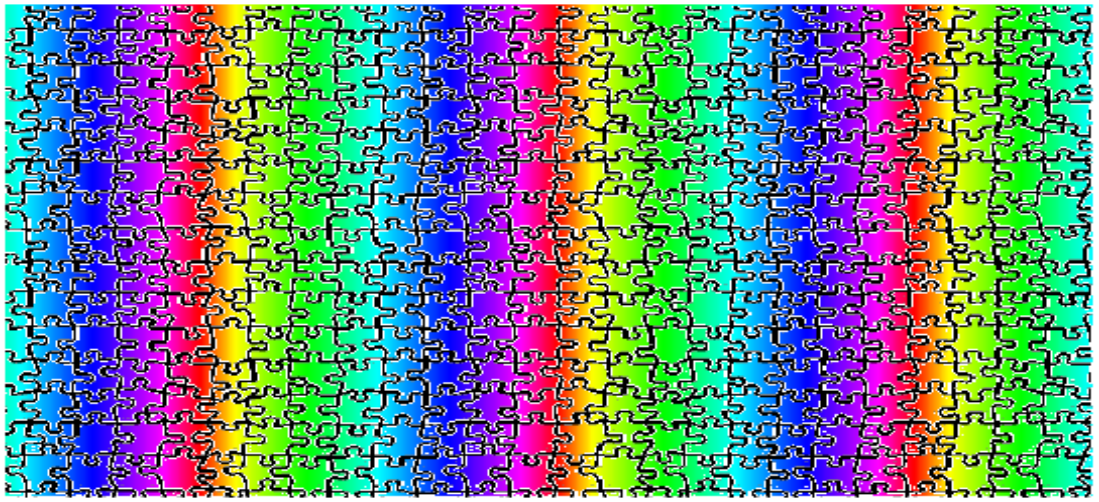


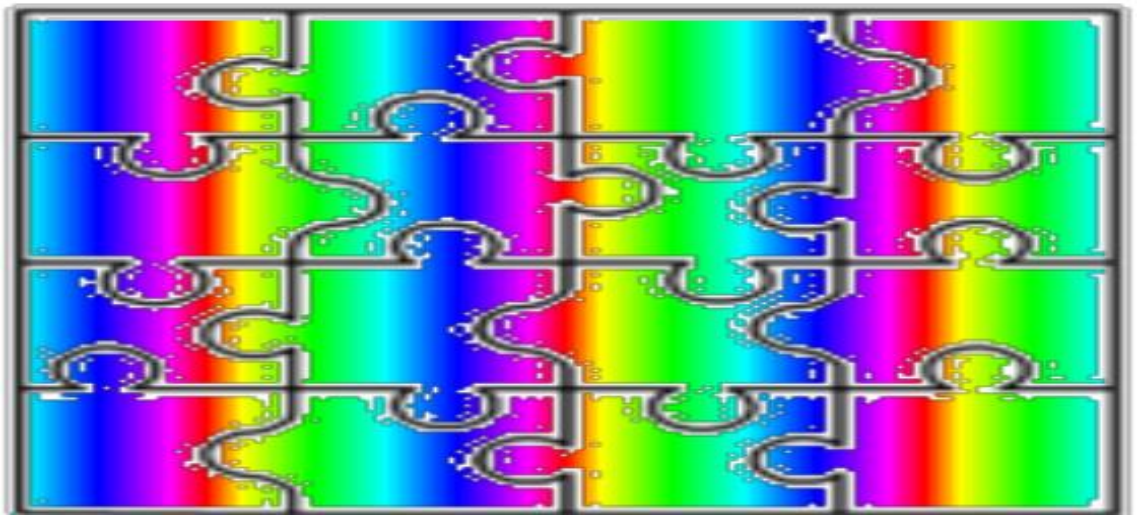
Source : The Genomic Characteristics and Origin of Chromothripsis
Alessio Marcozzi, Franck Pellestor, and Wigard P. Kloosterman

INTRODUCTION

The Human Genome as a Puzzle







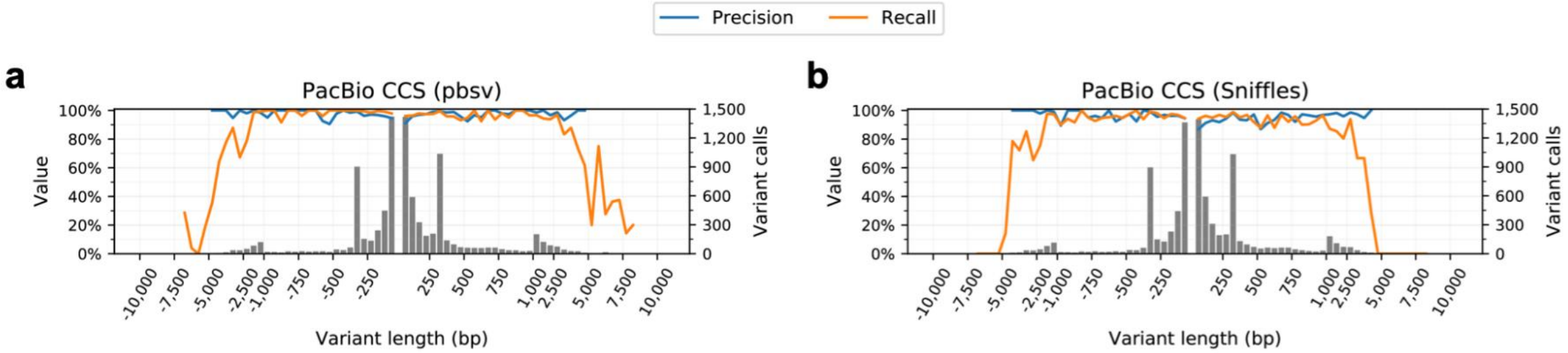
INTRODUCTION

Short-reads : 2 X 150 pb

Long-reads : 14 kb

Bloc de chromothripsis : 300 kb

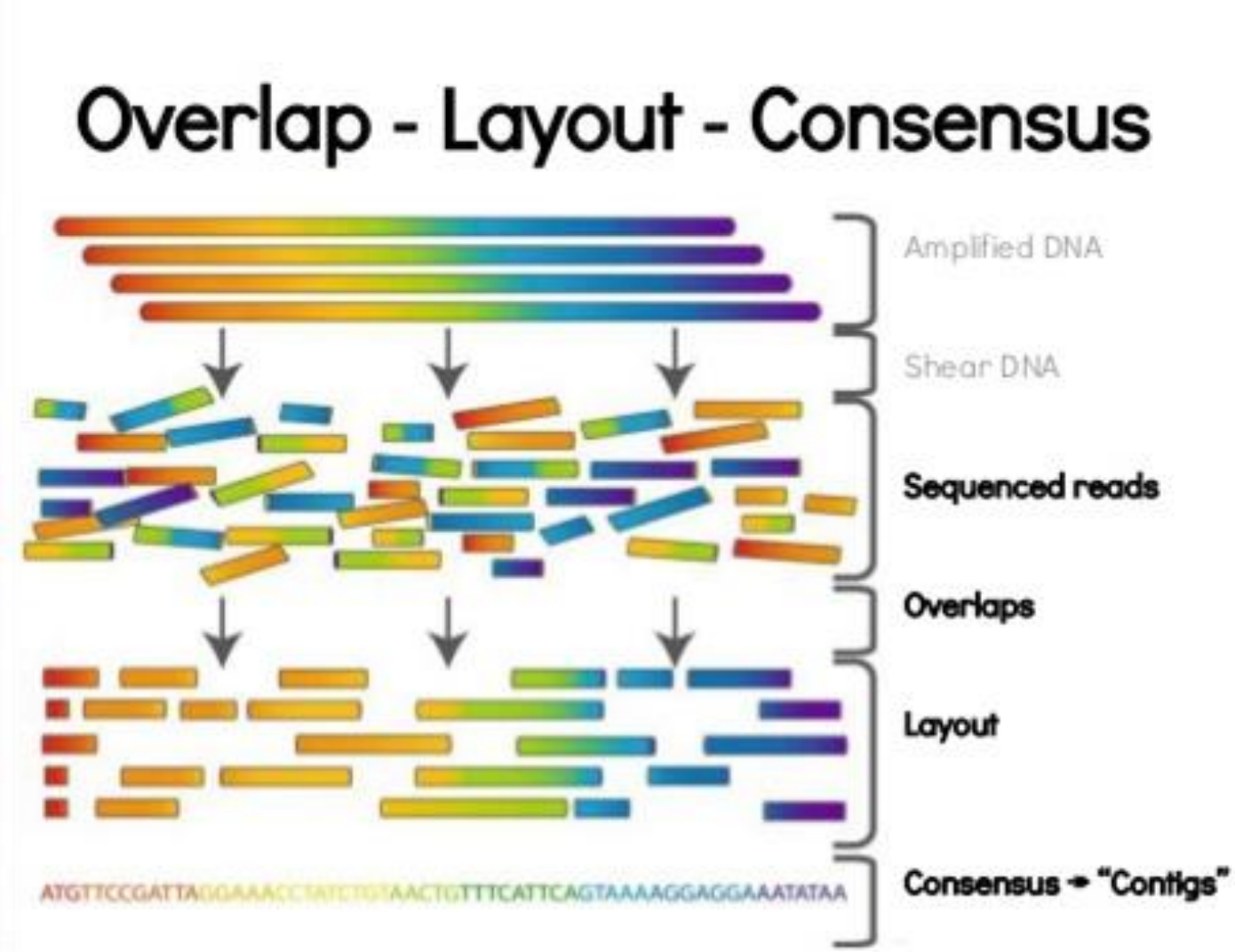
INTRODUCTION



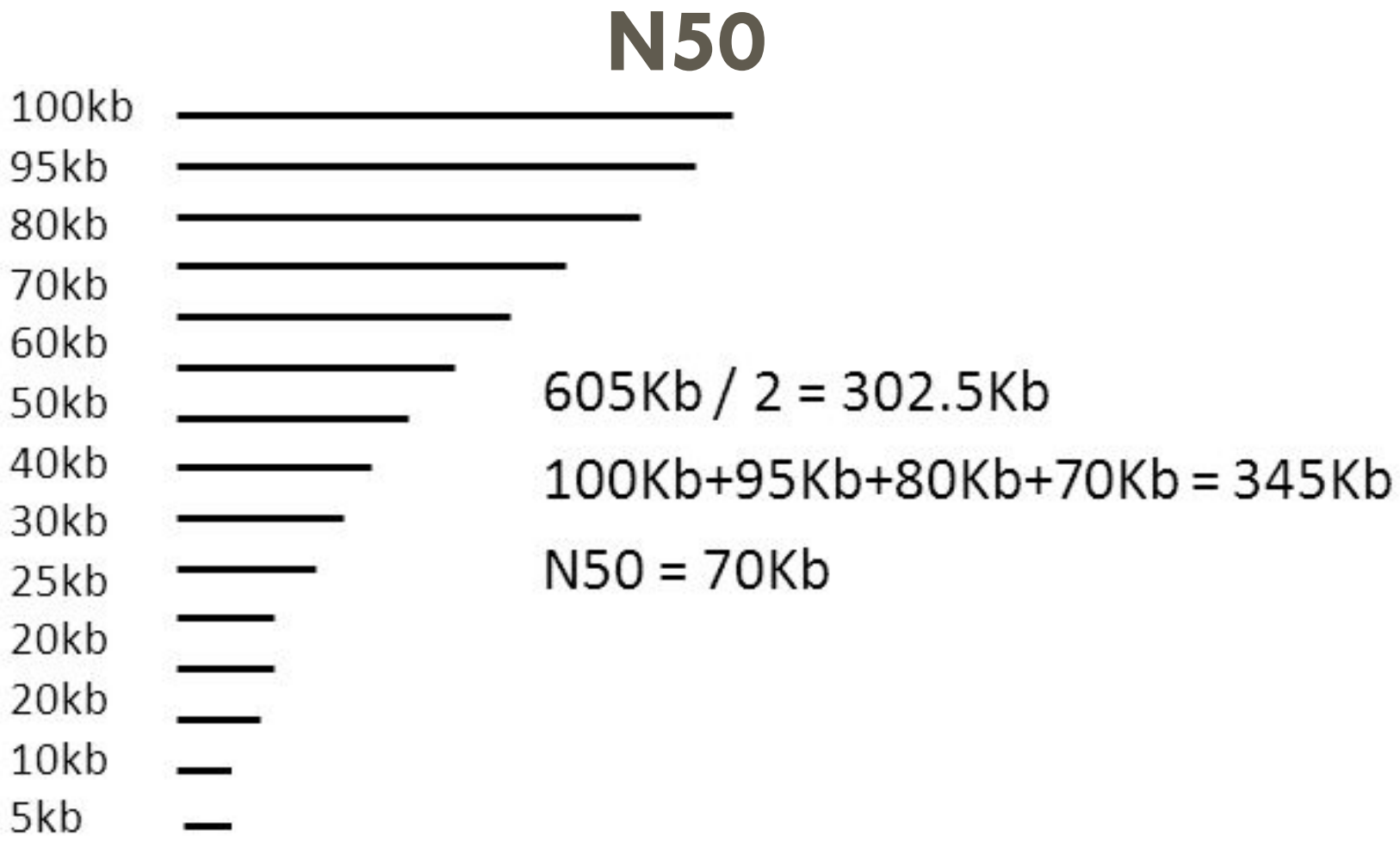
`pbsv` calls insertions, deletions, inversions, duplications, and translocations. Both single-sample calling and joint (multi-sample) calling are provided. `pbsv` is most effective for:

- insertions 20 bp to 10 kb
- deletions 20 bp to 100 kb
- inversions 200 bp to 10 kb
- duplications 20 bp to 10 kb
- translocations between different chromosomes or further than 100kb apart on a single chromosome

INTRODUCTION

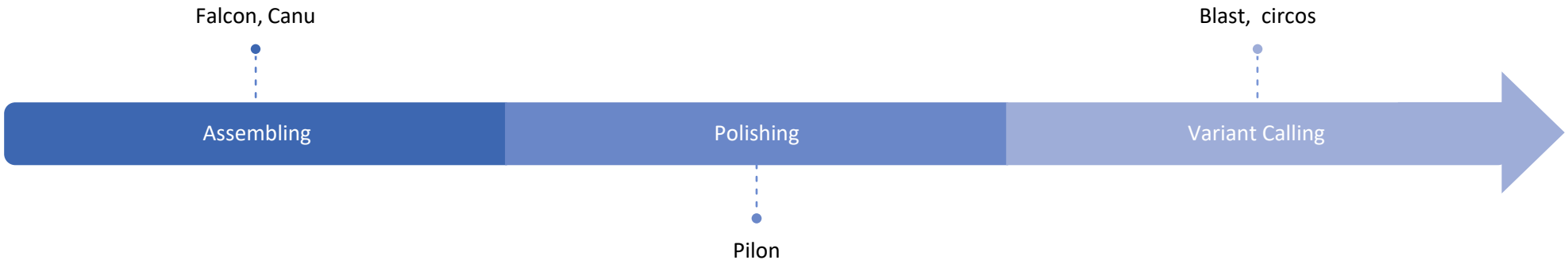


INTRODUCTION



MATERIAL & METHOD

- 2 patients atteints de troubles malformatifs et ayant un chromothripsis
- Pour chaque patient : SMRT Pacific Bioscience Long read genome Sequencing et Illumina Short read genome Sequencing



MATERIAL & METHOD

Patient 1

- Pacbio :

Longueur moyenne = 13 kb

Nombre de base = 69G

- Illumina :

Longueur moyenne = 100 pb

Nombre de base = 186G

Patient 2

- Pacbio :

Longueur moyenne = 14 kb

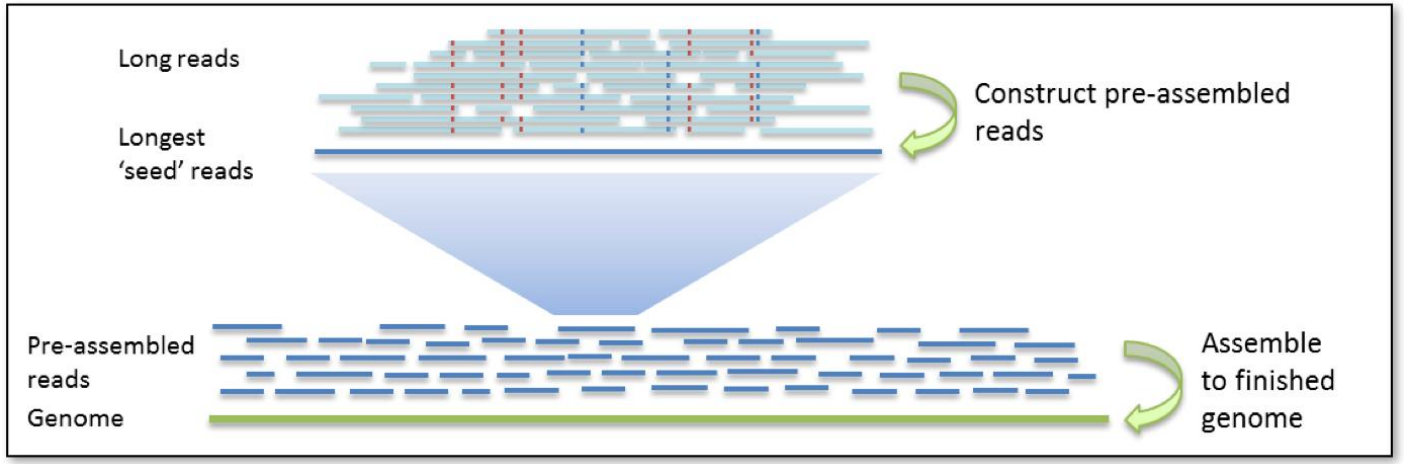
Nombre de base = 61G

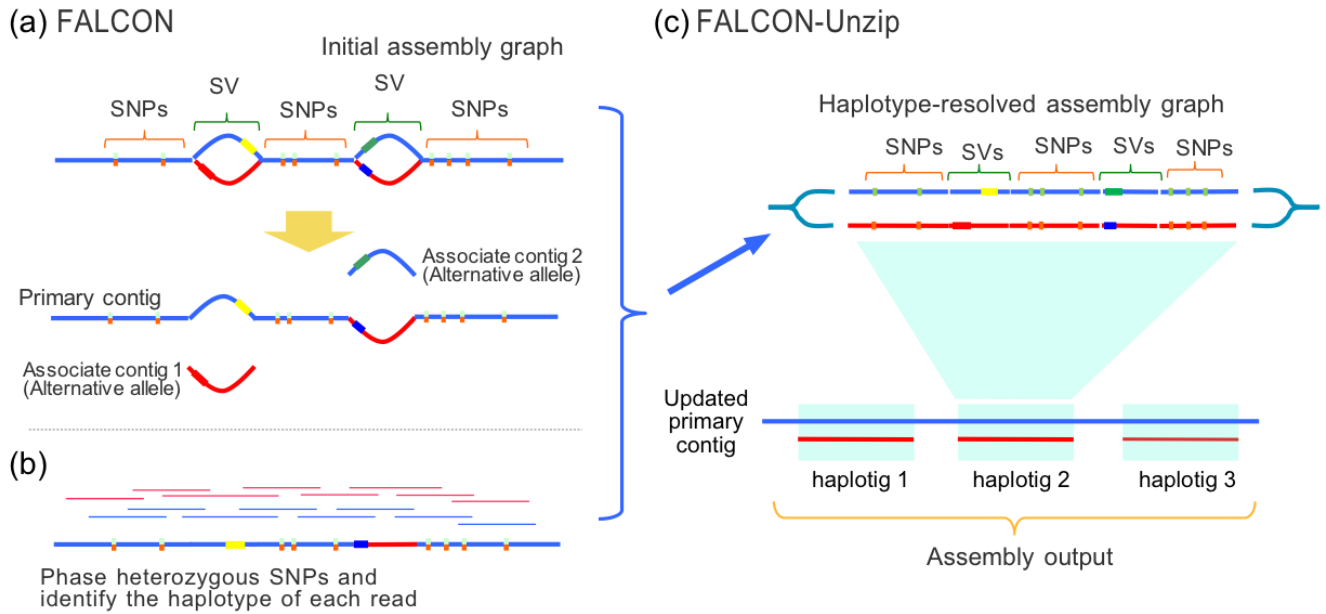
- Illumina :

Longueur moyenne = 100 pb

Nombre de base = 79G

MATERIAL & METHOD



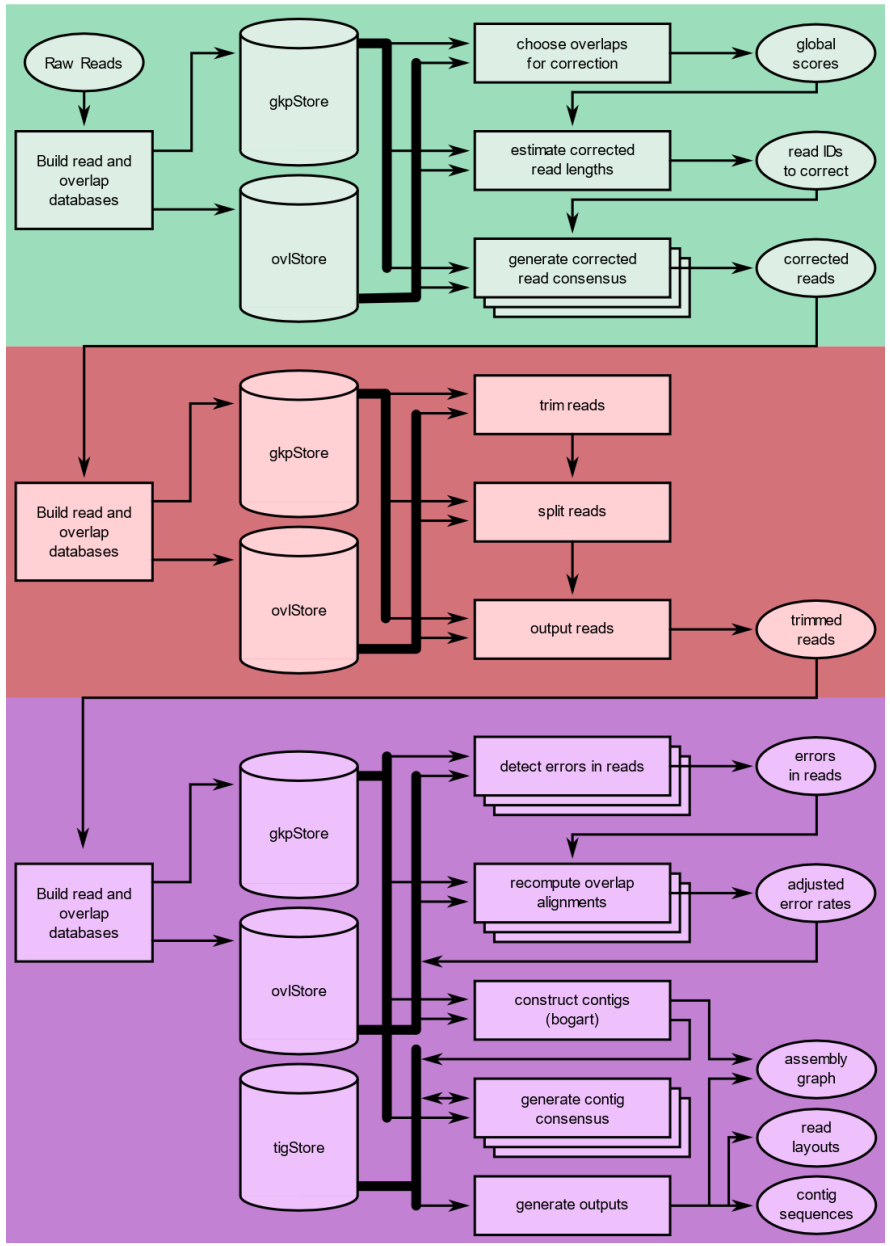


Source : Chin, et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature Methods. 10(6), 563.

Assembling | Polishing | Variant Calling

MATERIAL & METHOD

Maryland Bioinformatics Labs

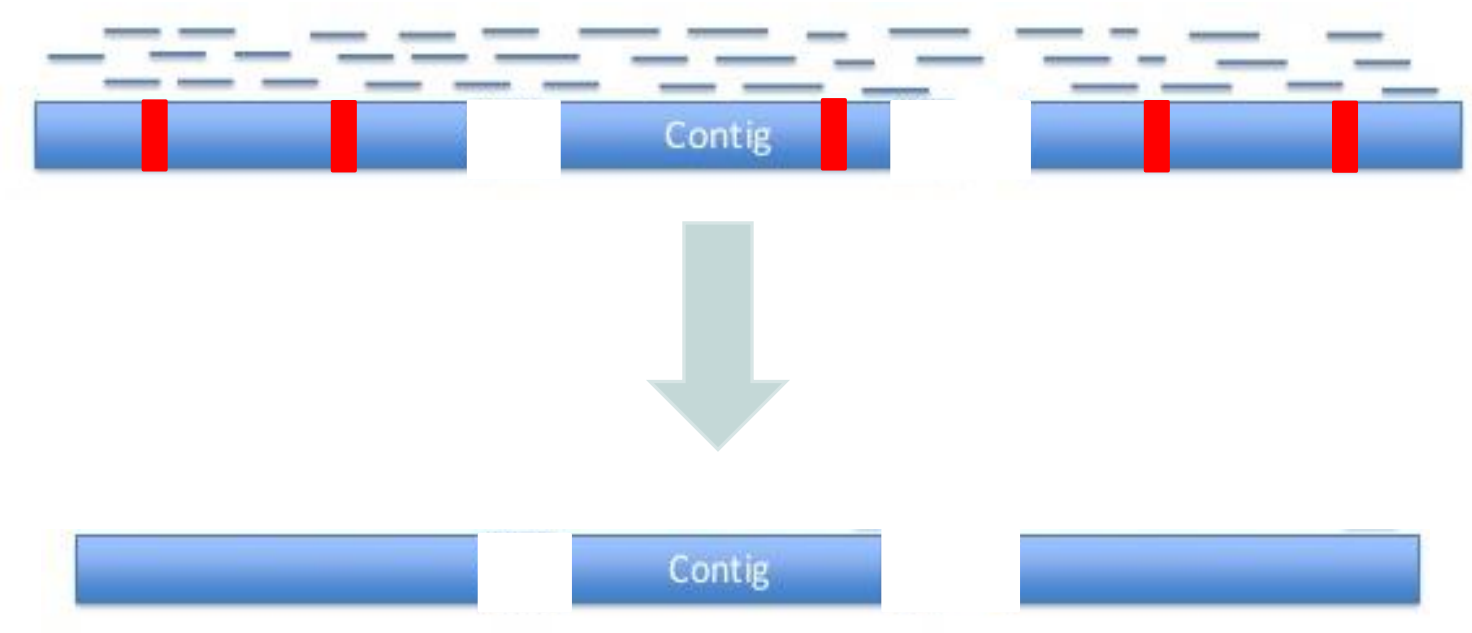


Source : Koren S, Walenz BP, Berlin K, Miller JR, Phillippy AM. **Canu**: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. bioRxiv. (2016).

Assembling | Polishing | Variant Calling

MATERIAL & METHOD

Pilon



Source : Bruce J. Walker, Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha Sakthikumar, Christina A. Cuomo, Qiandong Zeng, Jennifer Wortman, Sarah K. Young, Ashlee M. Earl (2014) **Pilon**: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS ONE 9(11): e112963. doi:10.1371/journal.pone.0112963

MATERIAL & METHOD

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009 Dec 15;10:421.

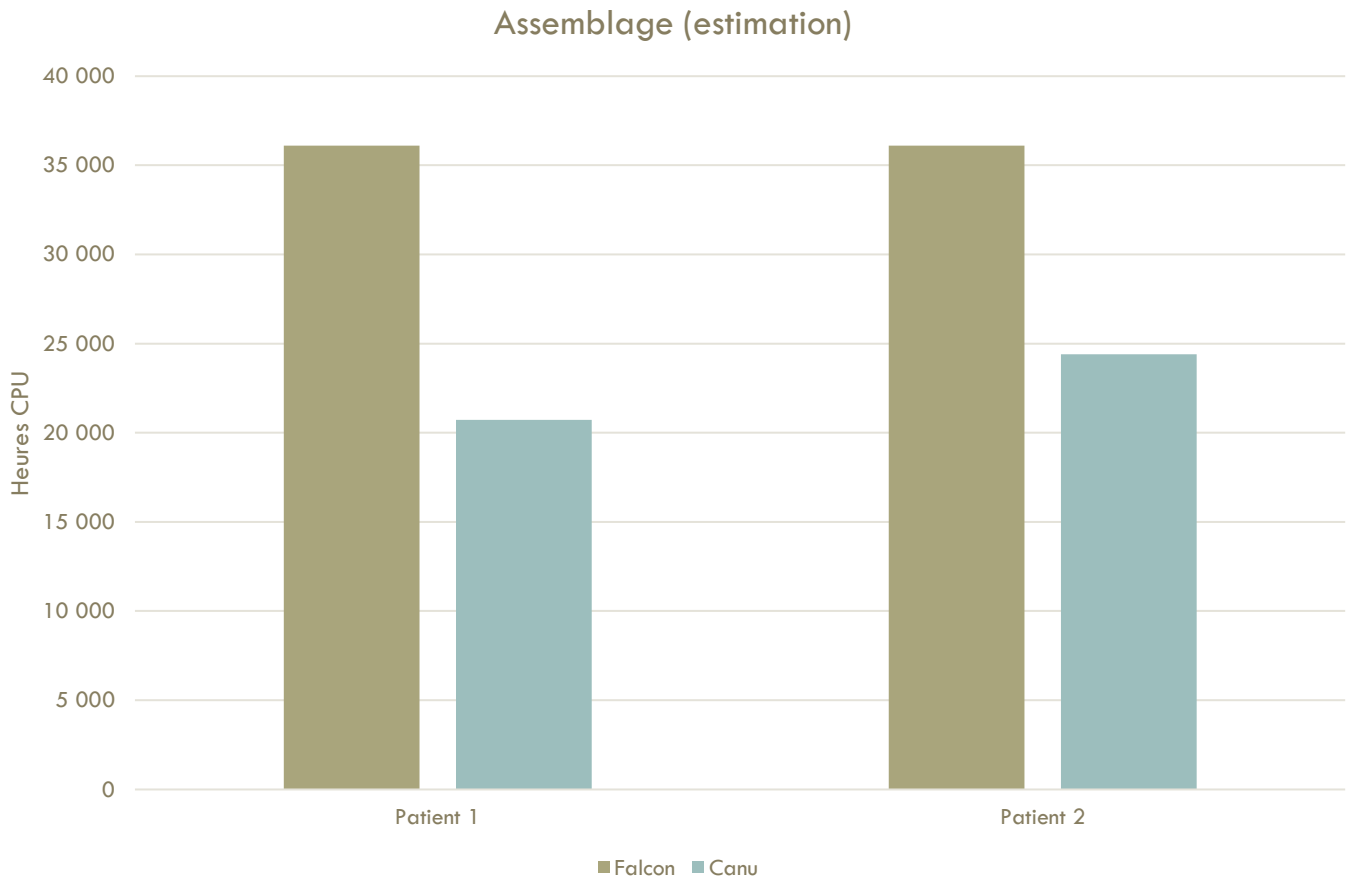
KRZYWINSKI, Martin, SCHEIN, Jacqueline, BIROL, Inanc, et al. Circos: an information aesthetic for comparative genomics. Genome research, 2009, vol. 19, no 9, p. 1639-1645.

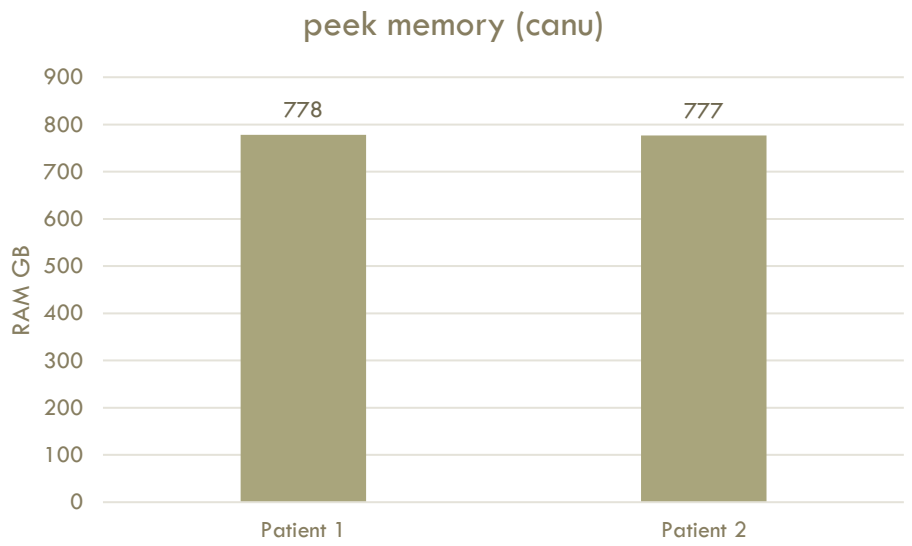
Assembling | Polishing | Variant Calling

RÉSULTATS

Falcon			Canu	
Patient 1	Patient 2		Patient 1	Patient 2
2,8Gb	2,6Gb	Nombre de bases	2,4Gb	2,4Gb
5693	9968	Nombre de contigs	3423	6507
132 kb	545 kb	N50	979 kb	584 kb
99,7%	99,7%	Pourcentage d'identité moyen	99,5%	99,7%

RÉSULTATS

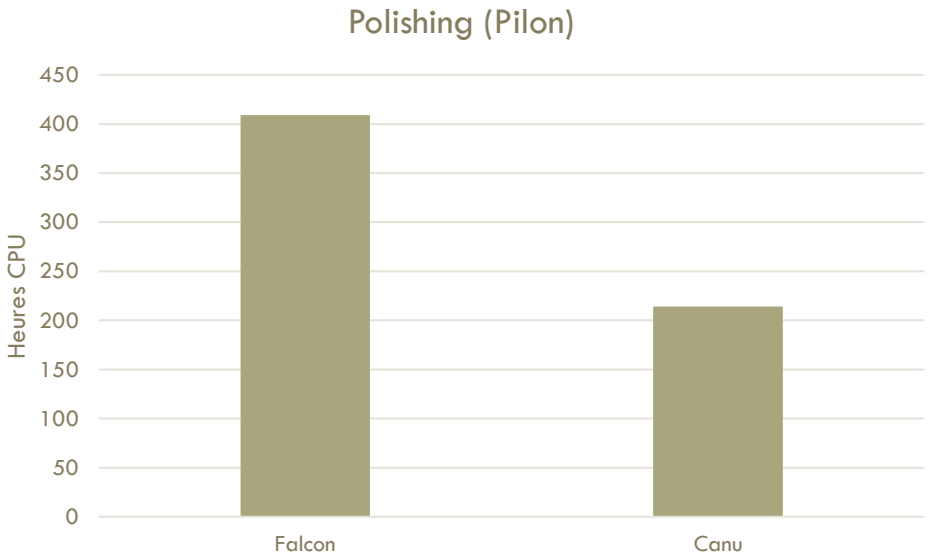

Assemblage (estimation)
40 000
35 000
30 000
25 000
20 000
15 000
10 000
5 000
0
Heures CPU
Patient 1
Patient 2
Falcon
Canu

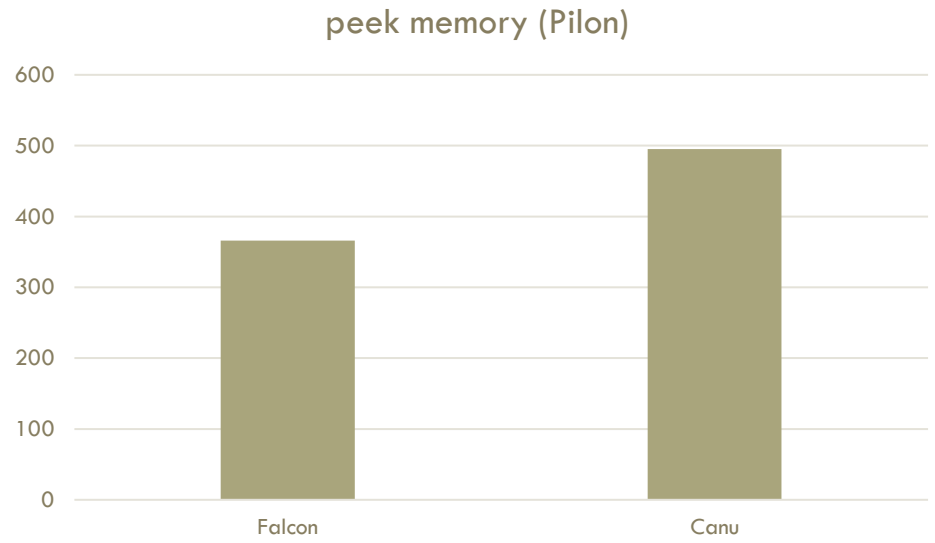

peek memory (canu)
900
800
700
600
500
400
300
200
100
0
RAM GB
778
777
Patient 1
Patient 2


Assembling
Polishing
Variant Calling

RÉSULTATS

Falcon		Canu
2,6Gb	Nombre de bases	2,4Gb
9968	Nombre de contigs	6507
545 kb	N50	584 kb

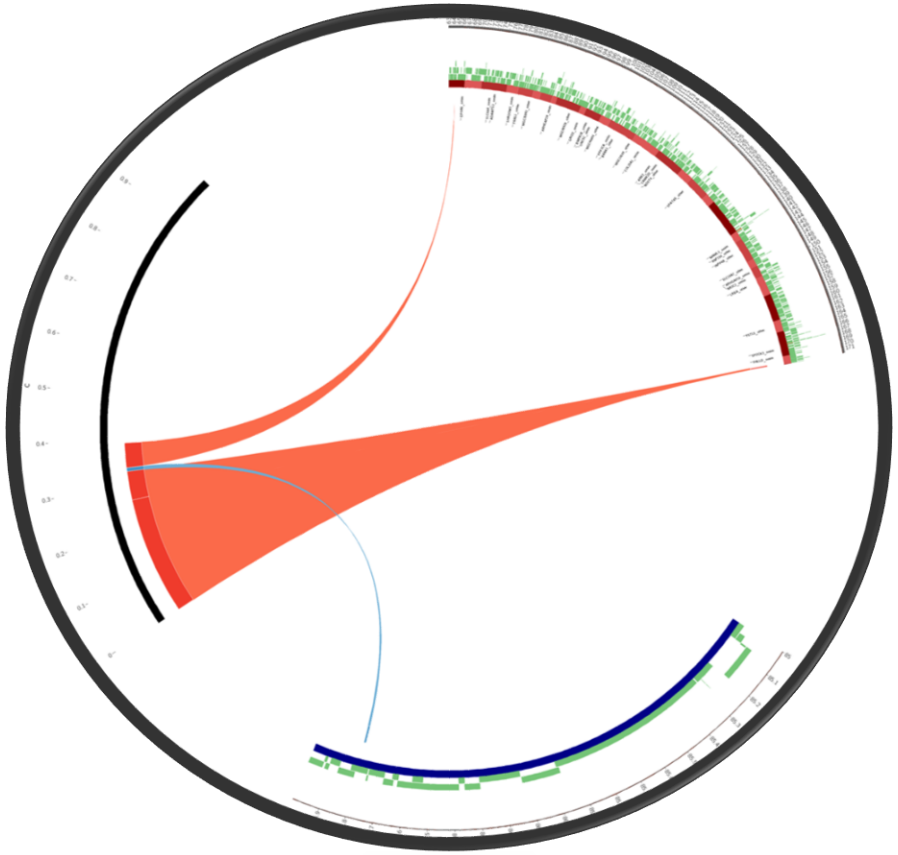


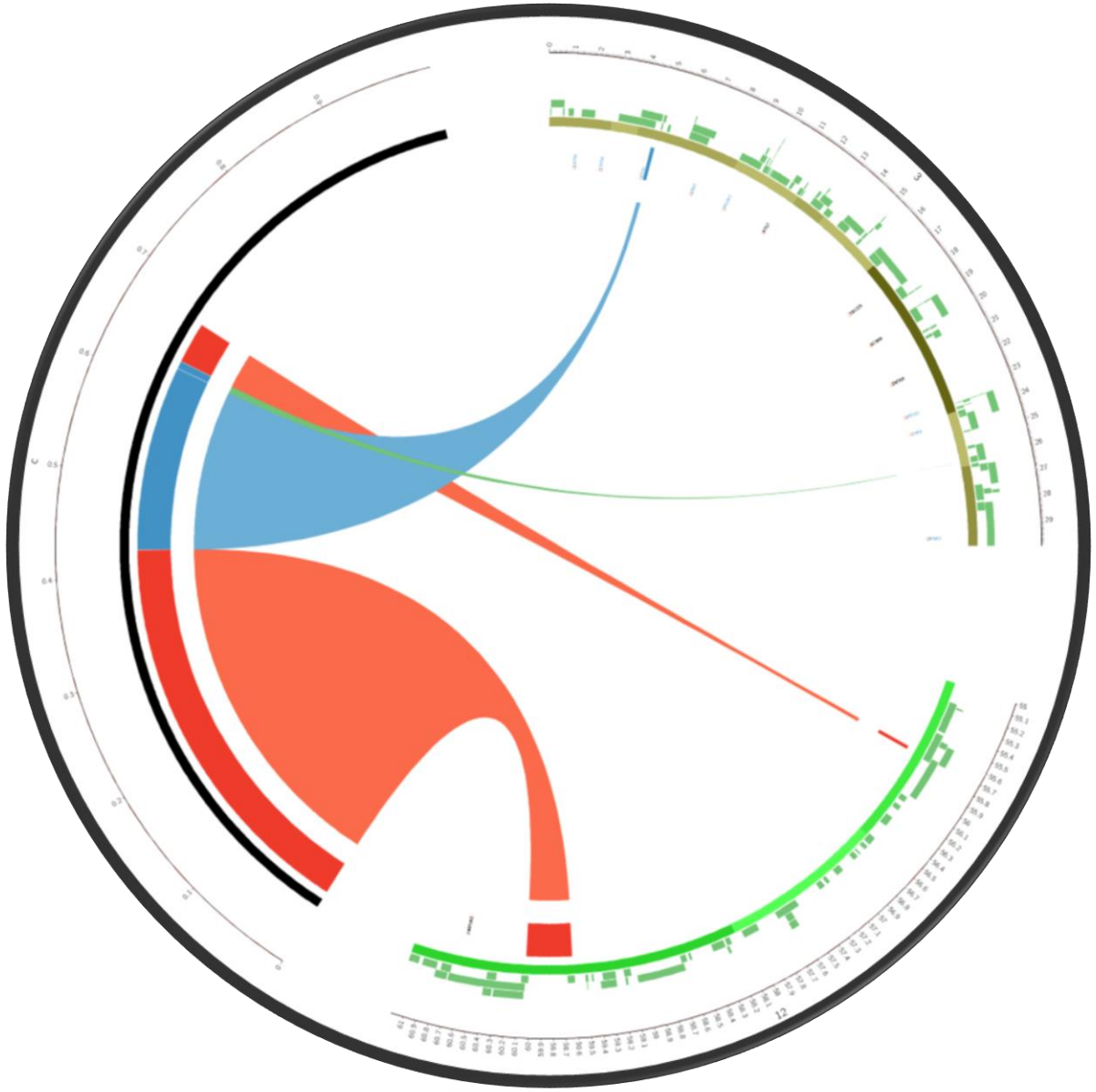


RÉSULTATS

Statistiques des contigs avec au moins un bloc de chromothripsis	Falcon		Canu	
	Patient 1	Patient 2	Patient 1	Patient 2
Nombre de contigs	107	114	84	78
Taille totale en base des contigs	91 Mb	40 Mb	41 Mb	35 Mb
Taille moyenne des blocs de chromosome dans les contigs	279 kb	132 kb	193 kb	177 kb
Pourcentages moyens d'identité des contigs (hg19)	98,89	98,79	98,91	98,92

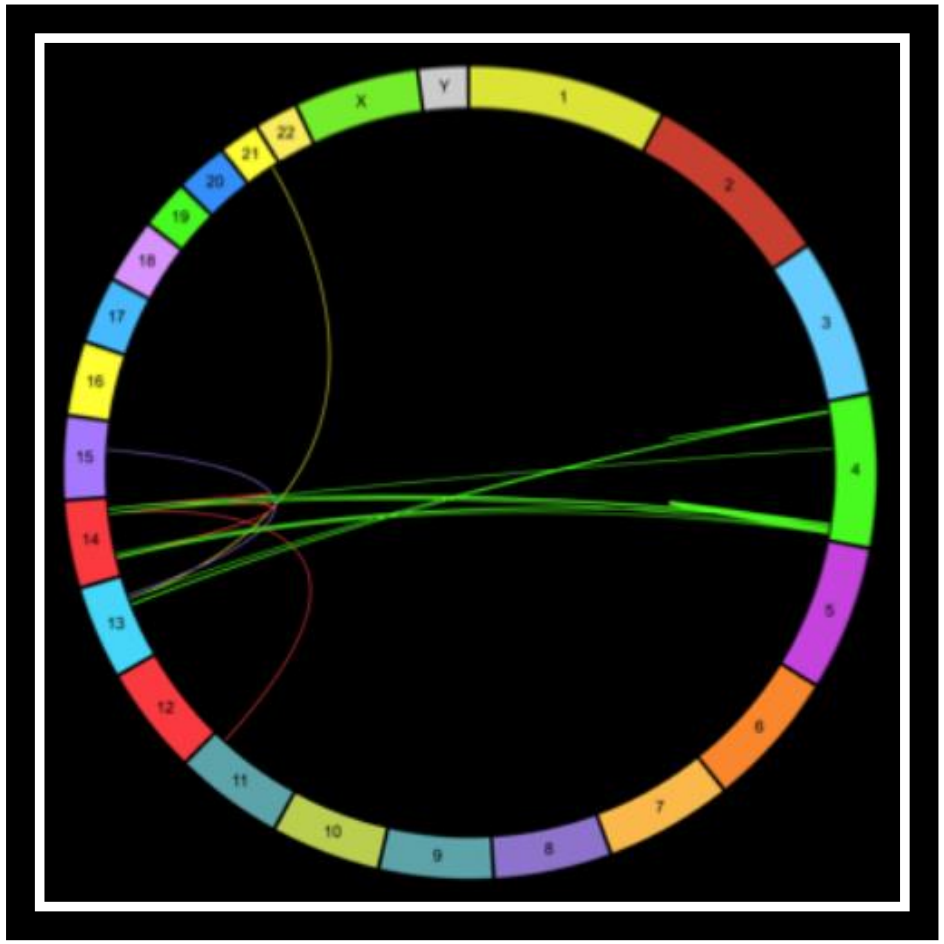
RÉSULTATS

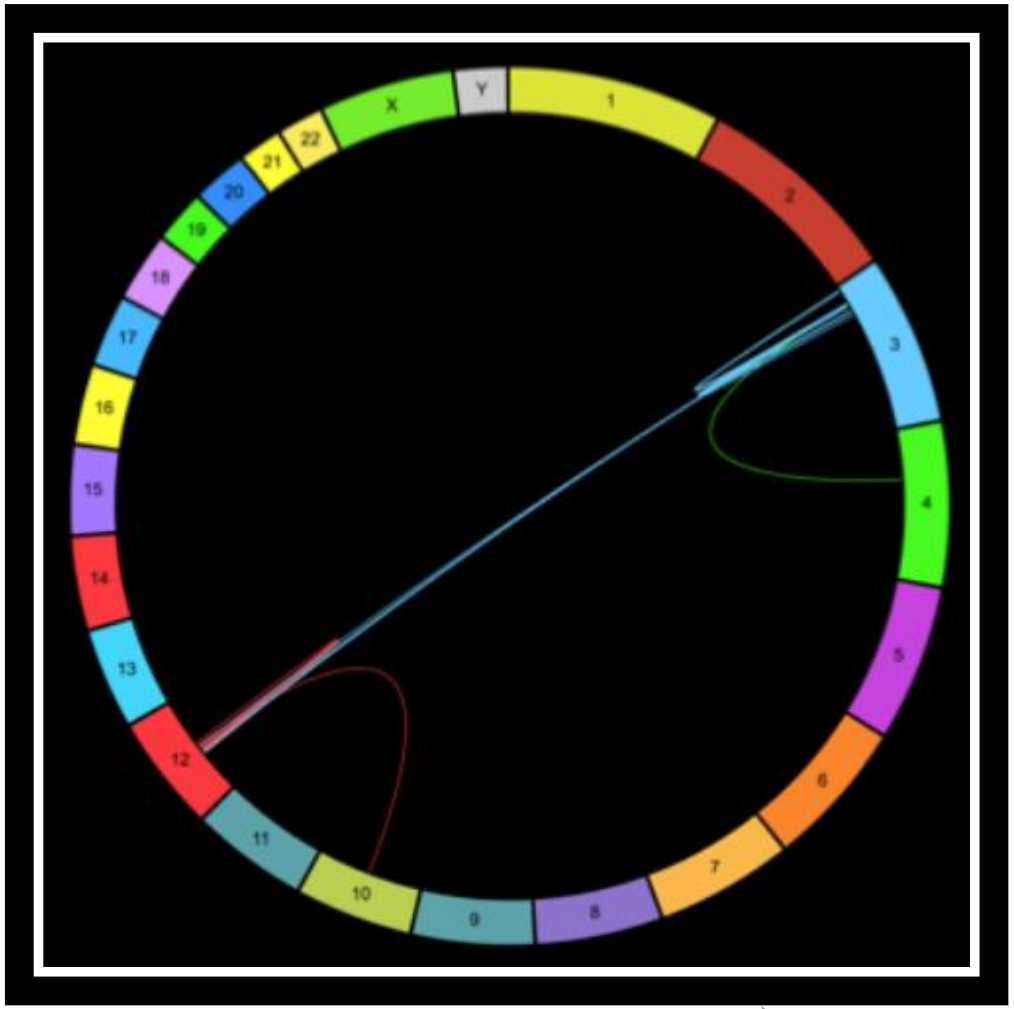




Assembling | Polishing | Variant Calling

RÉSULTATS



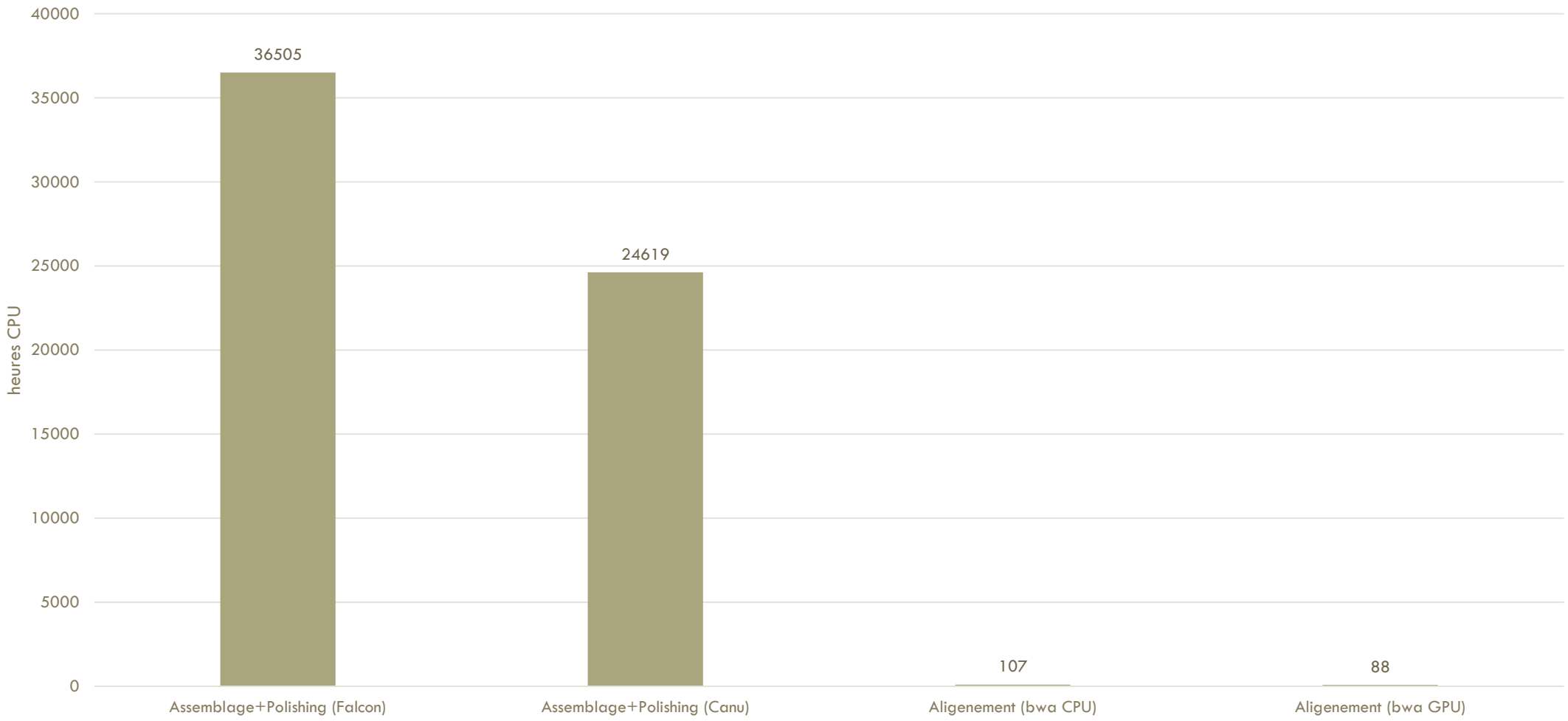


Assembling | Polishing | Variant Calling

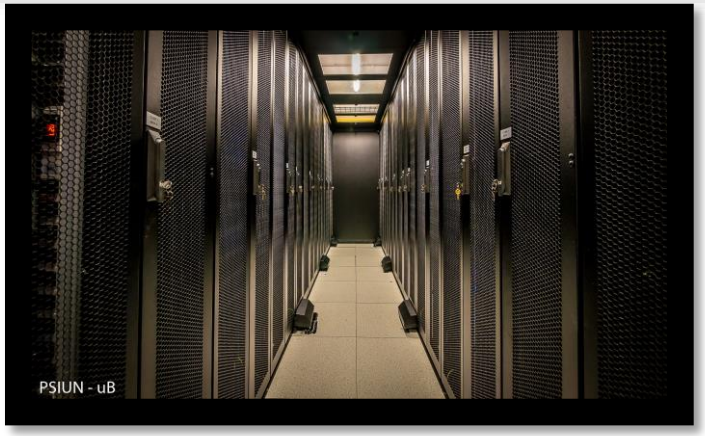
RÉSULTATS

Patient 1		Patient 2
35	Nombre de gènes	50
4	Nombre de gènes reliés avec une pathologie humaine	3

COMPARAISON



CENTRE DE CALCUL DE L'UNIVERSITÉ DE BOURGOGNE (CCUB)



cluster de calcul du CCUB

- 388 serveurs
- 6590 cœurs
- 28 824 Go RAM

Bioinformaticien senior:
Yannis Duffourd

Génétique et Anomalie du développement (GAD)

Membres du laboratoire:

- Professeurs des universités et praticien hospitalier

- Docteurs
- ingénieurs
- techniciens
- doctorants

MERCI POUR VOTRE ATTENTION

Équipe GAD :

Laurence Faivre
Christel Thauvin-Robinet

Christophe Philippe
Frédéric Tran-Mau Them
Pierre Vabres
Yannis Duffourd
Emilie Tisserant
Virginie Quéré
Laurence Jego
Ange-Line Bruel
Antonio Vitobello
Mirna Assoum
Romain Da Costa
Philippine Garret
Paul Kuentz
Nada Houcinat

Sébastien Moutton
Nolwenn Jean-Marçais
Patrick Callier
Anne-Laure Mosca-Boidron
Nathalie Marle
Thibaud Jouan
Charlotte Poë
Martin Chevarin
Justine Lavoyer
Mathilde Lefèbvre
Jean-Charles Crepin
Nicolas Bourgon
Sophie Nambot
Julian Delanne