

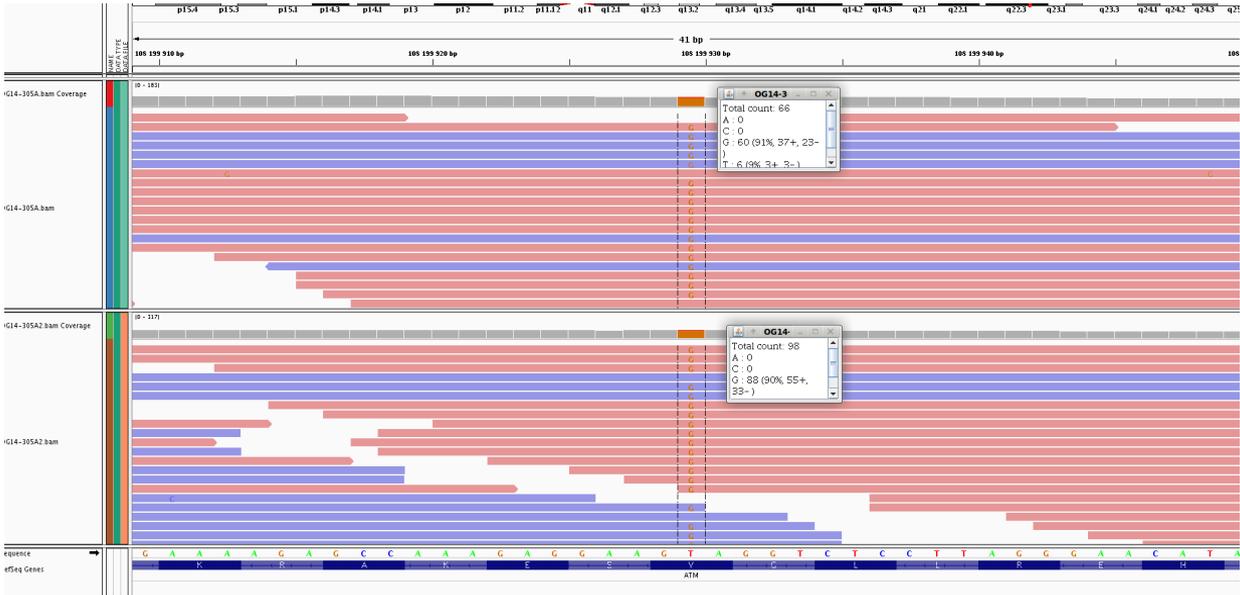
Les technologies de SHD de 3ème génération

M. Duffourd Yannis
Bioinformaticien – CHU Dijon

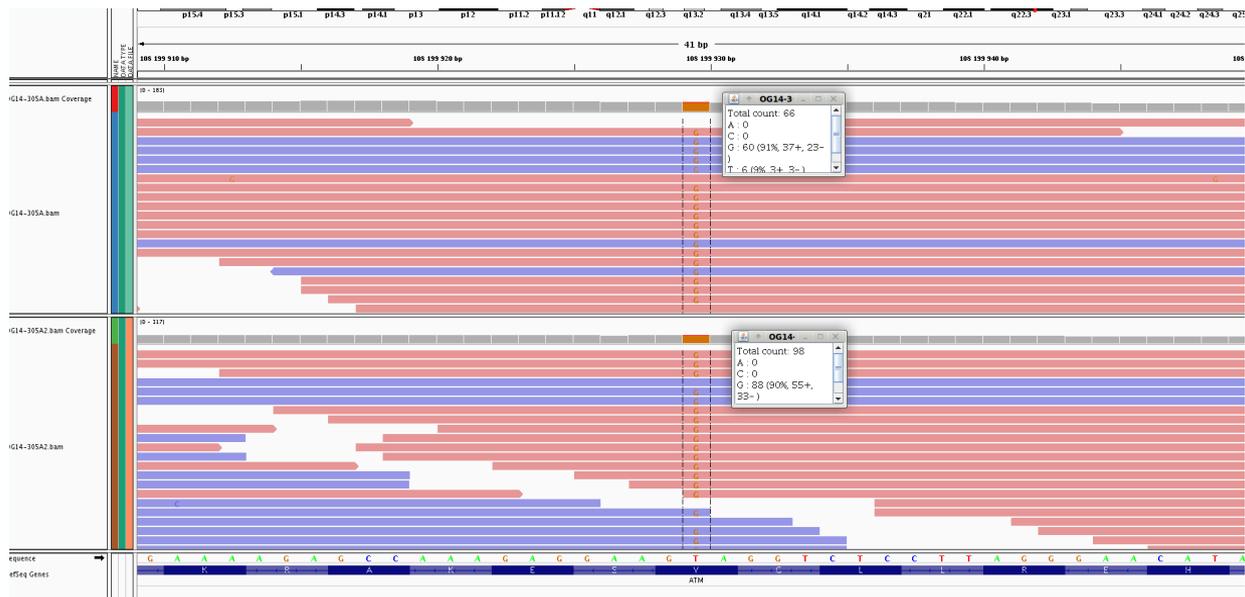
Séminaire BioinfoDiag
Le 19 Septembre 2019



Le combo classique *bwa* + *GATK* fait l'unanimité



Le combo classique *bwa* + *GATK* fait l'unanimité



Les limites

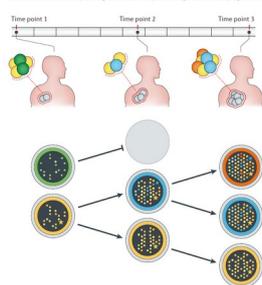
Phasing

Répétitions

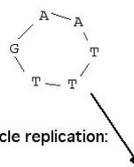
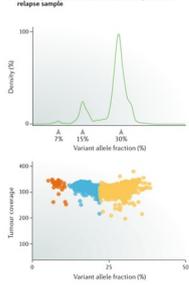
Séquences spéciales
(mitochondrie, EMI, haplotypes)

Parafine, Mosaïque, Somatique,
etc ...

a Clonal evolution: sequencing of tumour samples throughout disease progression

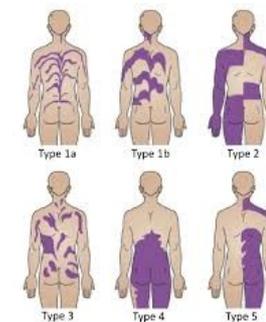
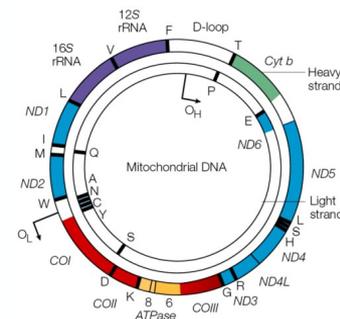


b Clonality analysis: a visualization of analysis of the relapse sample



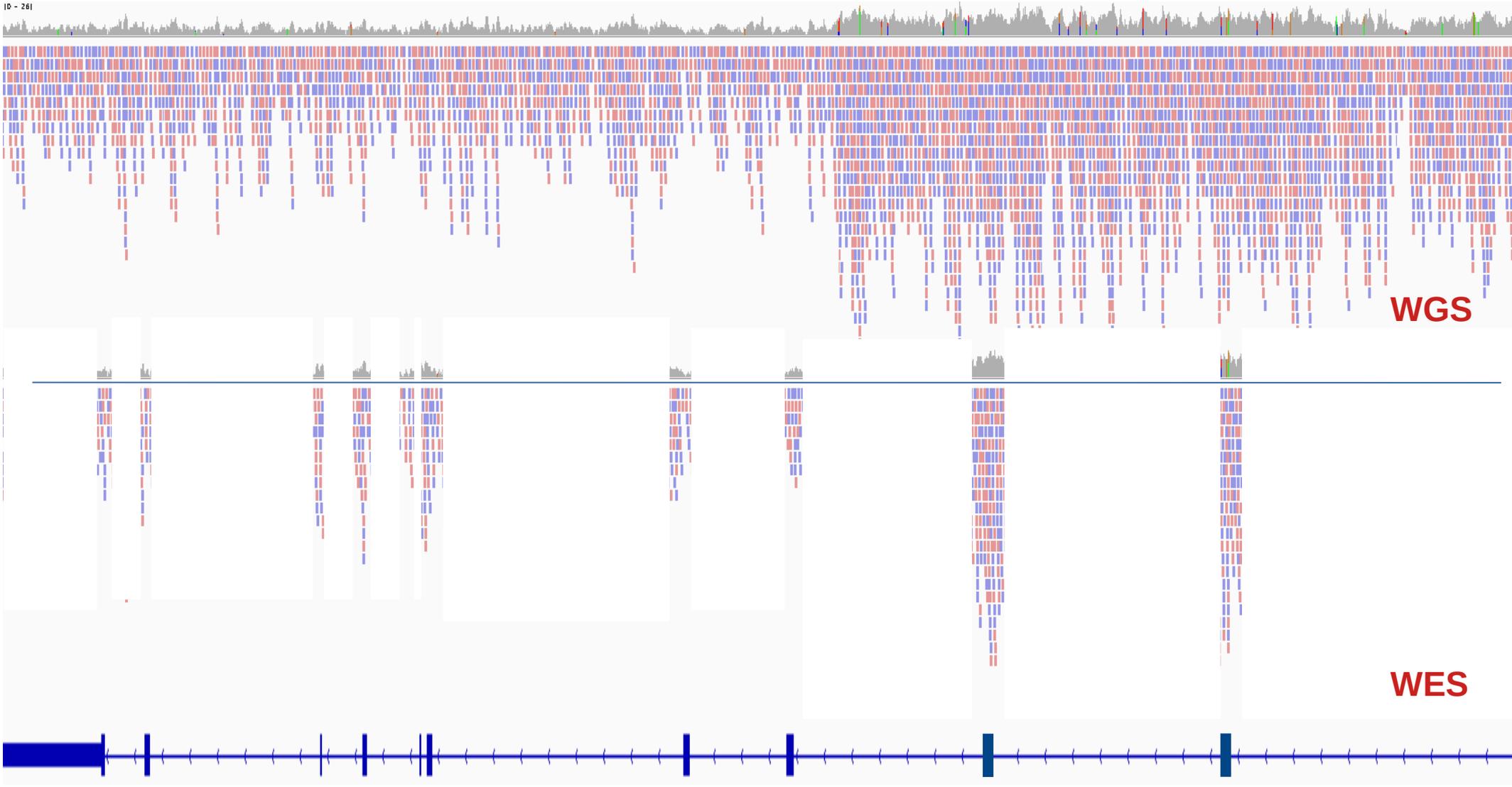
Rolling circle replication:

A A T T T G A A T T T G A A T T T G A A T T T G
A A T T T G A A T T T G A A T T T G A A T T T G



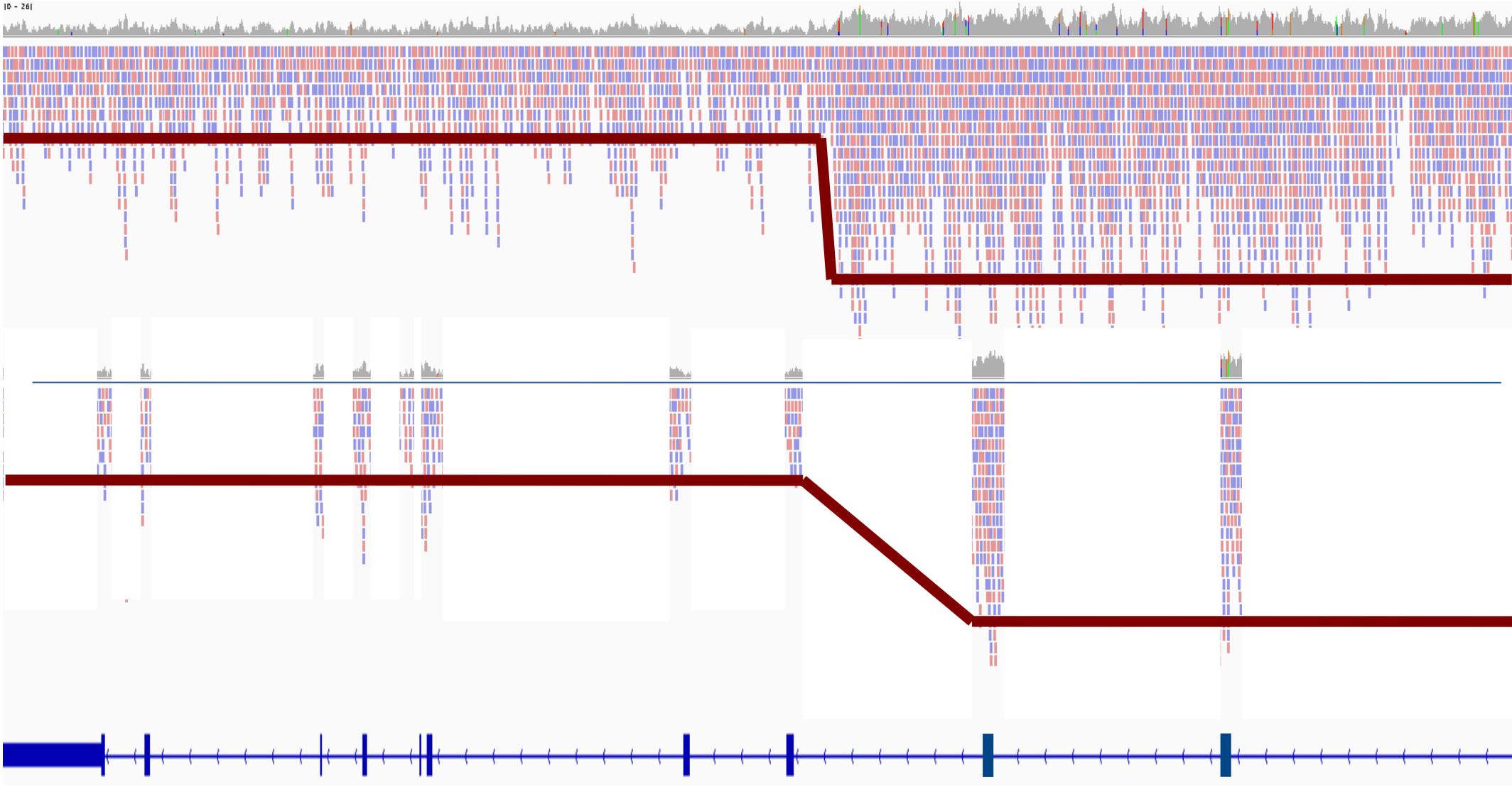
Les variations de structure en SHD

ID - 261

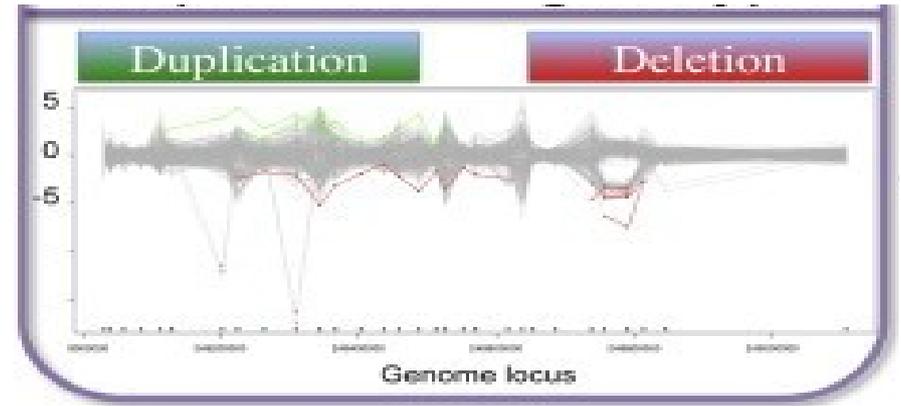
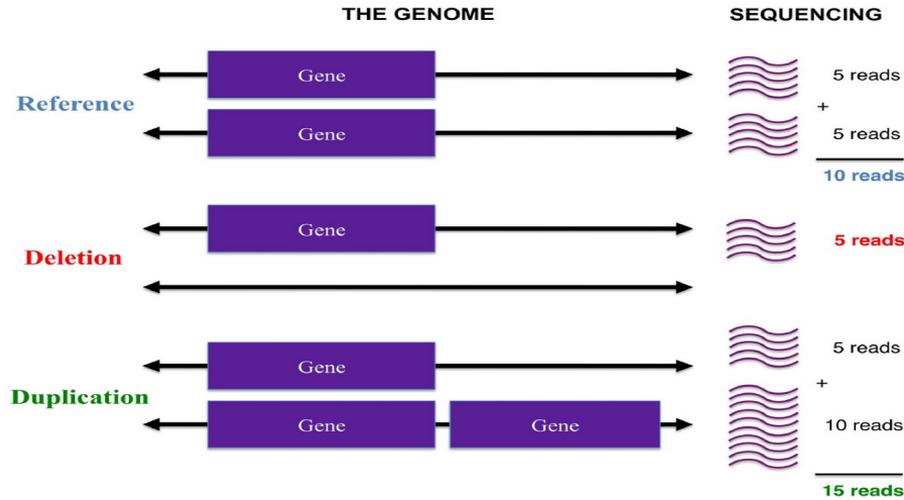


Les variations de structure en SHD

ID - 261

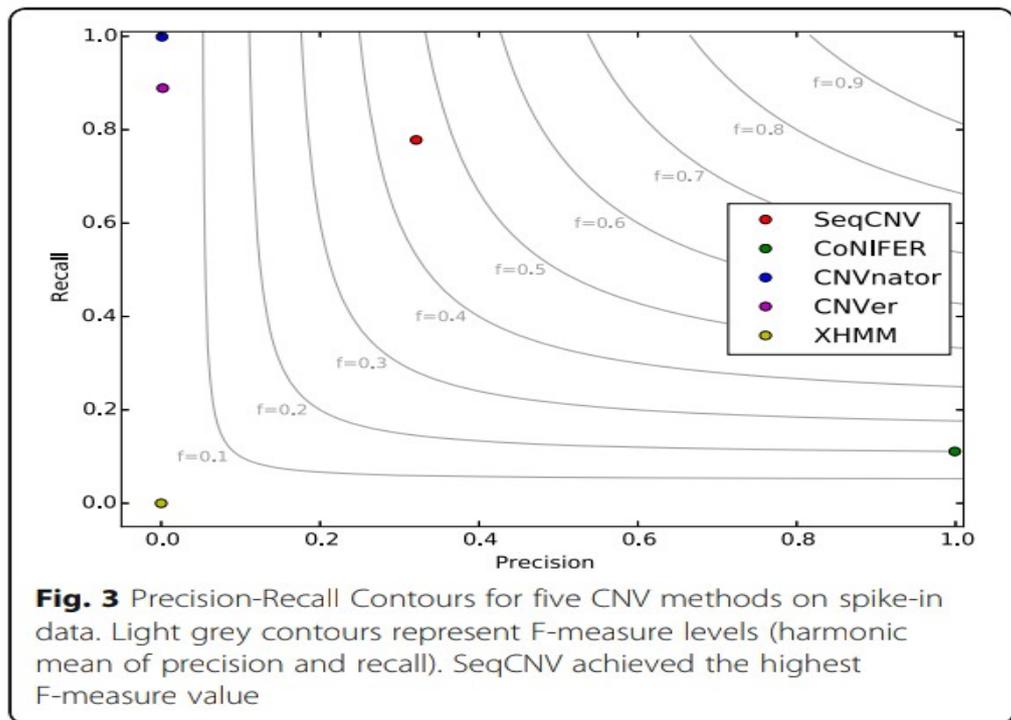


XHMM :



- CNV : Gain ou perte de matériel chromosomique
- Profondeur = nombre de lecture
- Normalisation des données (Bruit, biais de capture)
- Transformation en z-score de la profondeur moyenne de chaque exon

Sensibilités et Spécificités variables



SeqCNV : Chen & al. 2017

Faible recouvrement des approches

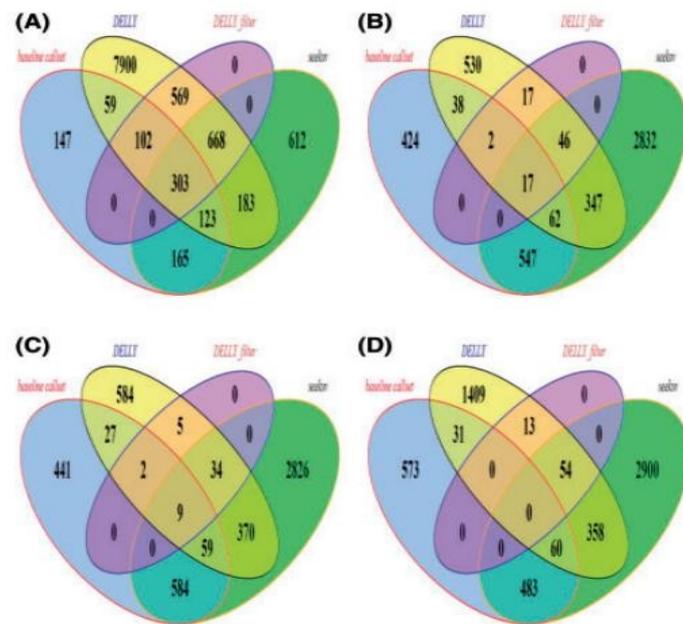


Fig. 3 Private/common calls of four samples. (A) NA12878. (B) NA19238. (C) NA19239. (D) NA19240. The cornflower blue area is SV result of baseline, green area is seeksv, yellow area is DELLY and dark orchid is DELLY after filtered

Seeksv : Liang & al. 2017

Multiplicité des approches de CNV calling

Curr Protoc Bioinformatics, 2014;45:15.6.1-11. doi: 10.1002/0471250953.bi1506s45.

BreakDancer: Identification of Genomic Structural Variation from Paired-End Read Mapping.

Fan X¹, Abbott TE², Larson D², Chen K¹.

Genome Biol. 2014 Jun 26;15(6):R84. doi: 10.1186/gb-2014-15-6-r84.

LUMPY: a probabilistic framework for structural variant discovery.

Layer RM, Chiang C, Quinlan AR, Hall IM.

Curr Protoc Hum Genet, 2014 Apr 24;81:7.23.1-21. doi: 10.1002/0471142905.hg0723s81.

Using XHMM Software to Detect Copy Number Variation in Whole-Exome Sequencing Data.

Fromer M¹, Purcell SM.

PeerJ, 2015 Mar 17;3:e836. doi: 10.7717/peerj.836. eCollection 2015.

GROM-RD: resolving genomic biases to improve read depth detection of copy number variants.

Smith SD¹, Kawash JK¹, Grigoriev A¹.

PLoS One, 2011 Jan 31;6(1):e16327. doi: 10.1371/journal.pone.0016327.

ReadDepth: a parallel R package for detecting copy number alterations from short sequencing reads.

Miller CA¹, Hampton O, Coarfa C, Milosavljevic A.

Bioinformatics, 2012 Feb 1;28(3):423-5. doi: 10.1093/bioinformatics/btr670. Epub 2011 Dec 6.

Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data.

Boeva V¹, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, Janoueix-Lerosey J, Delattre O, Barillot E.

Front Genet, 2014 Aug 1;5:248. doi: 10.3389/fgene.2014.00248. eCollection 2014.

The CNVrd2 package: measurement of copy number at complex loci using high-throughput sequencing data.

Nguyen HT¹, Merriman TR², Black MA².

Nucleic Acids Res, 2012 May;40(9):e69. doi: 10.1093/nar/gks003. Epub 2012 Feb 1.

cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate.

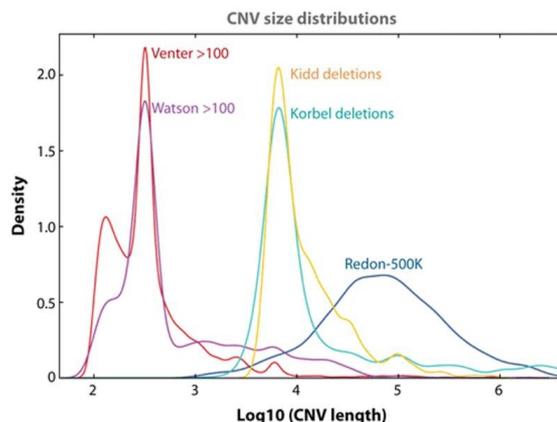
Klambauer G¹, Schwarzbauer K, Mayr A, Clevert DA, Mitterecker A, Bodenhofer U, Hochreiter S.

Bioinformatics, 2016 Aug 1;32(15):2375-7. doi: 10.1093/bioinformatics/btw163. Epub 2016 Mar 24.

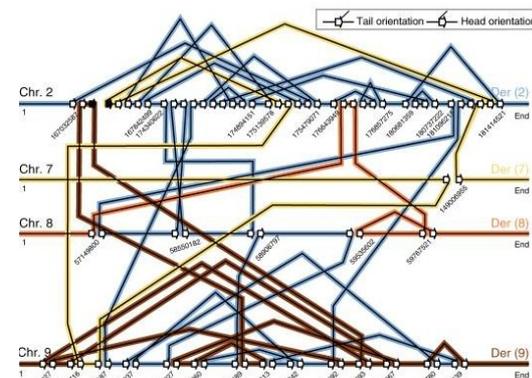
Canvas: versatile and scalable detection of copy number variants.

Roller E¹, Ivakhno S², Lee S¹, Royce T³, Tanner S¹.

« Trou » de détection dans la taille des CNVs



Difficultés à reconstituer des évènements complexes (chromothripsis)

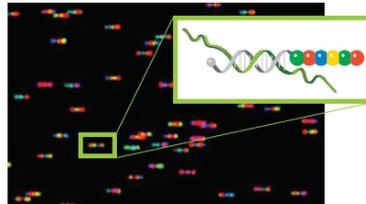
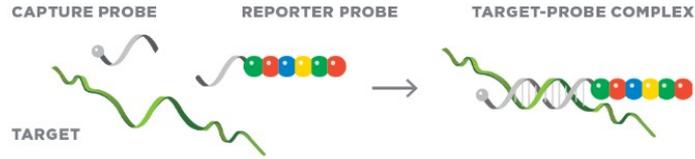
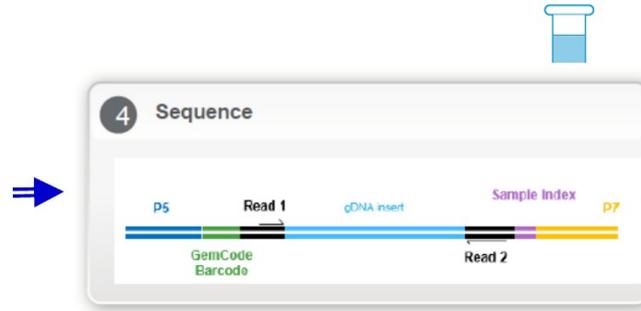
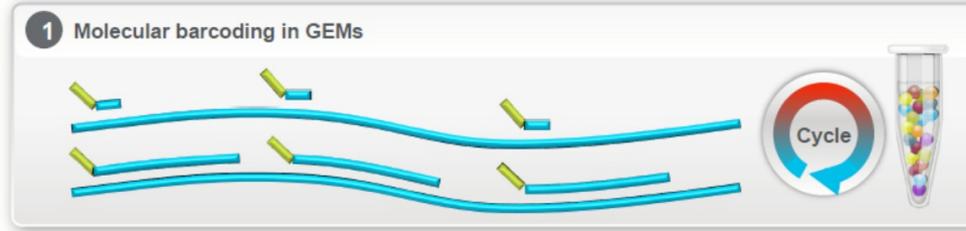


Les technologies de 2,5 ème génération ?

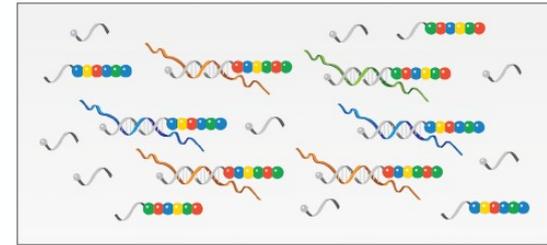
Low-input molecular barcoding in GEMs

10X
GENOMICS

Chromium

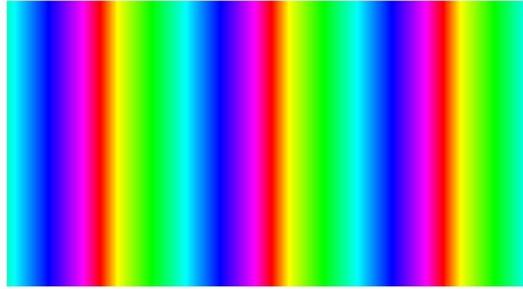
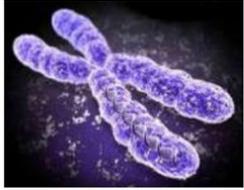


BARCODES	COUNTS	IDENTITY
	1	INSLN
	2	FOXO
	1	XLSA

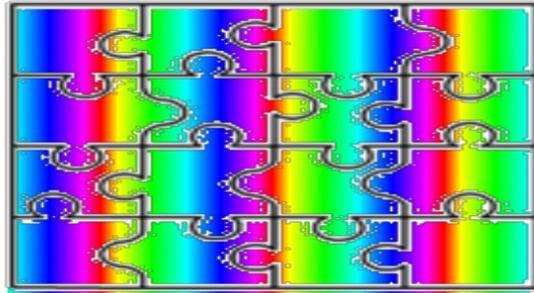
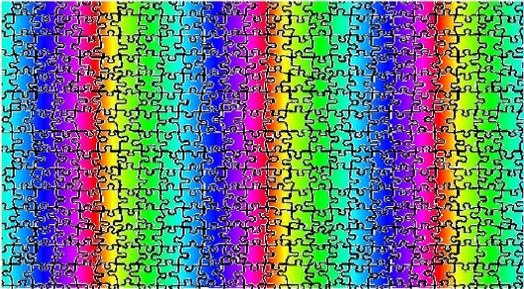


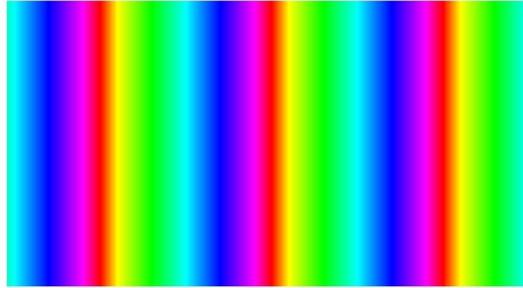
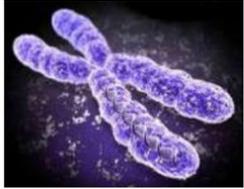
NanoString

Les technologies de 3^{ème} génération

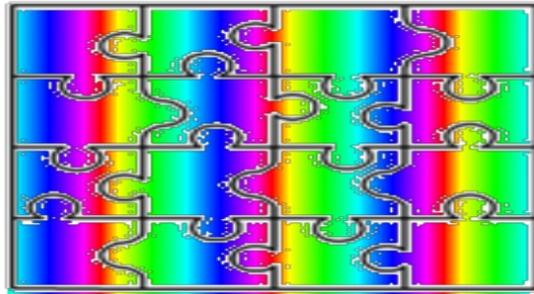
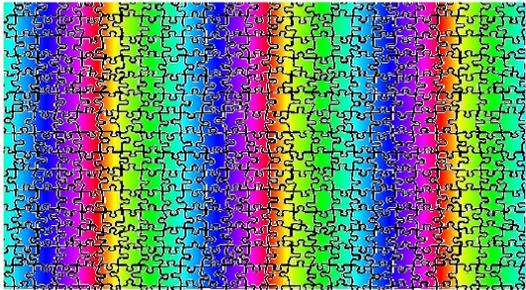


Métaphore du puzzle





Métaphore du puzzle



- Pas de Wash & Scan

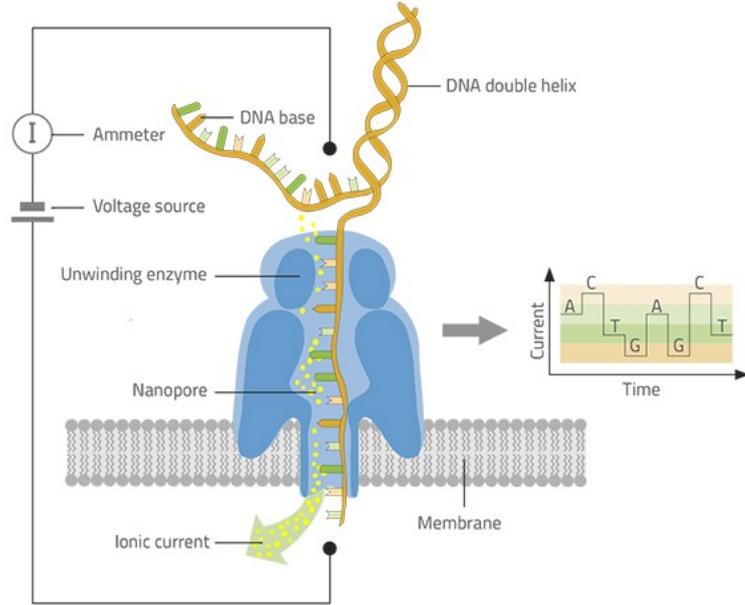
- Pas de PCR

- « Temps réel »

- Longs fragments !!!!



CNV & SV ??





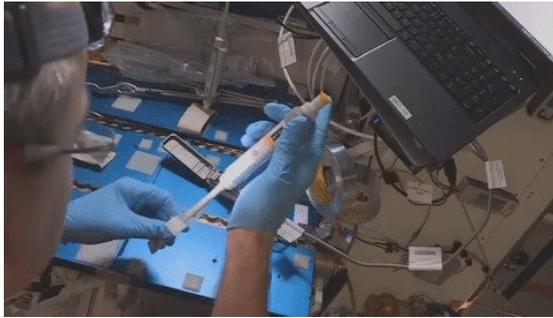
Dans l'espace ...



Dans le desert ...



En antarctique ...



Dans l'espace ...



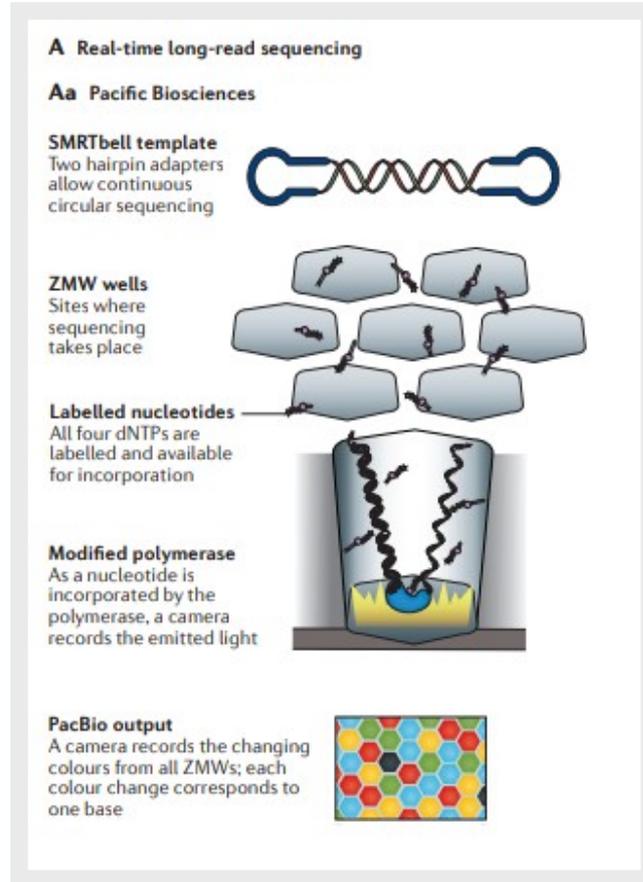
Dans le desert ...



En antarctique ...



Mais aussi dans un laboratoire ...

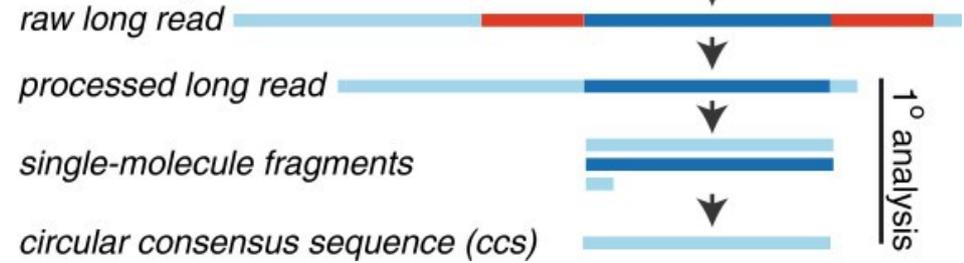


1. generate amplicon

2. ligate adaptors

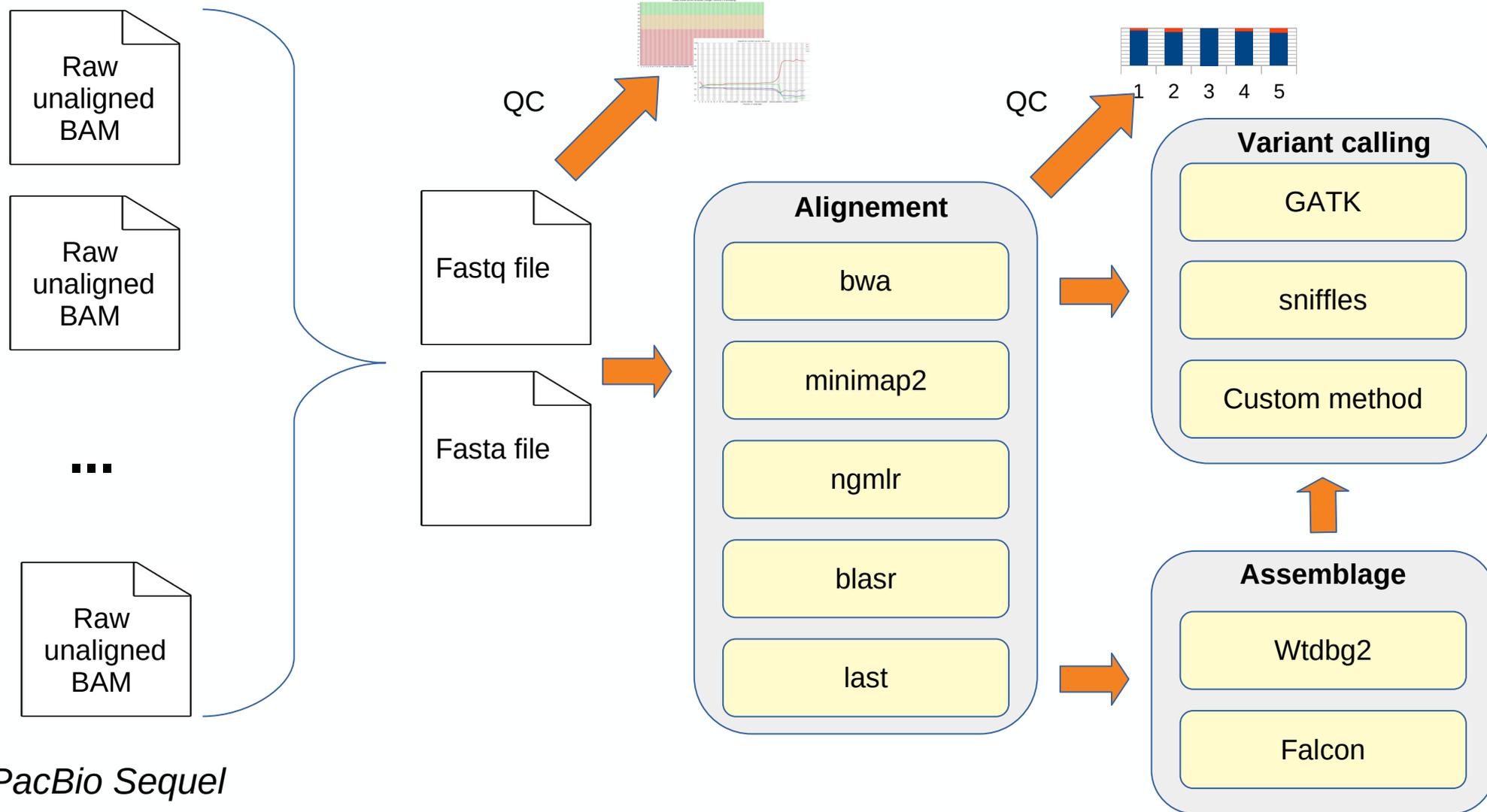
3. sequence

4. data analysis

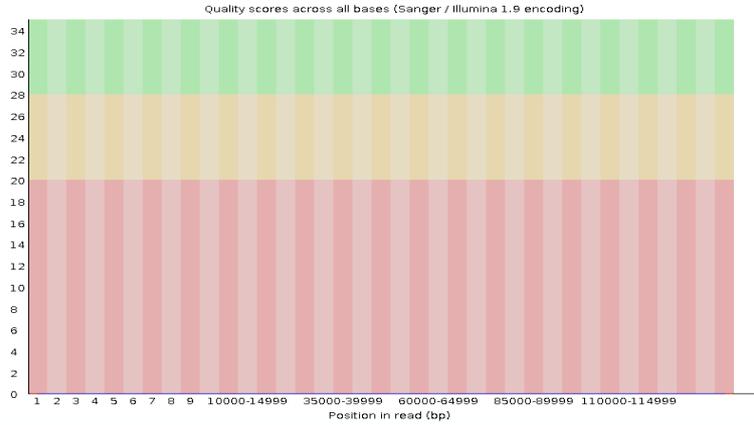


	Patient 1	Patient 2
Nombre de bases séquencées	73 milliards	64 milliards
Couverture	24X	21X
Nombre de lecture	4 801 843	4 401 045
Taille moyenne des lectures	15 212	14 685
N50	24 406	23 143
Taille maximum	90 344	138 732

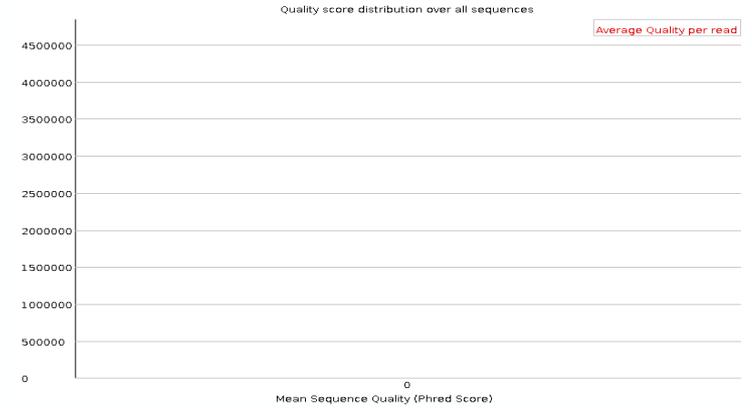
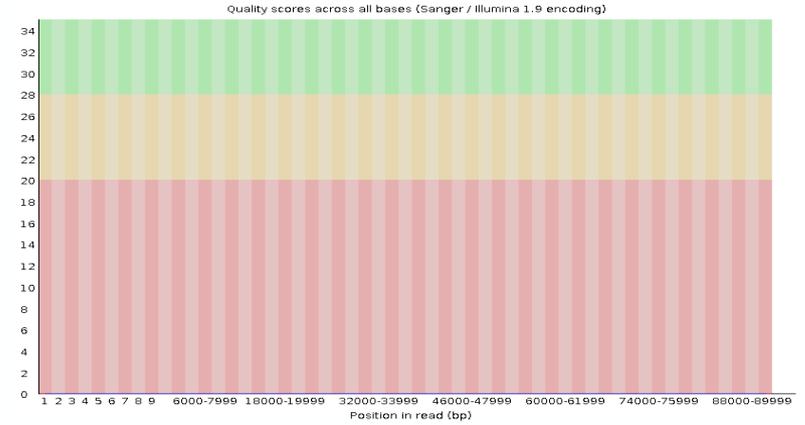
Plan expérimental



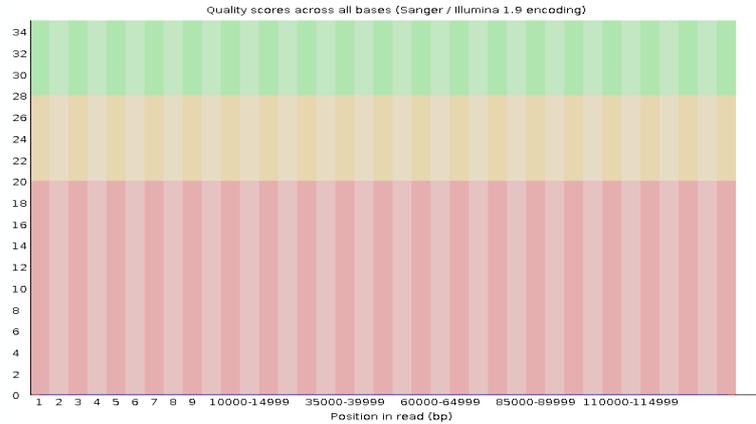
18A3182



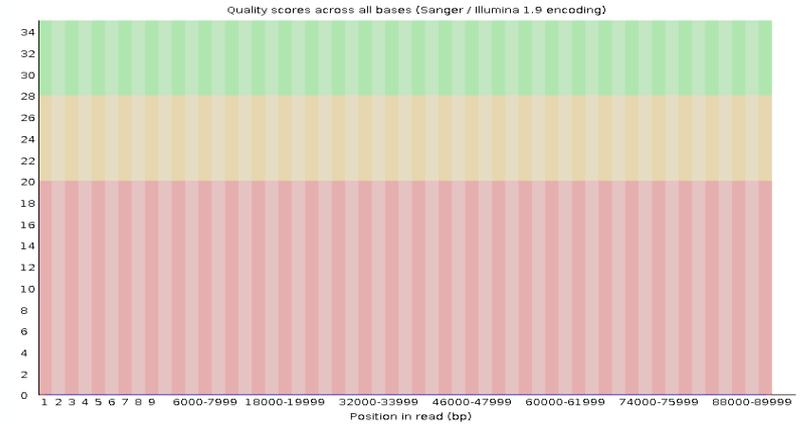
18A3183



18A3182



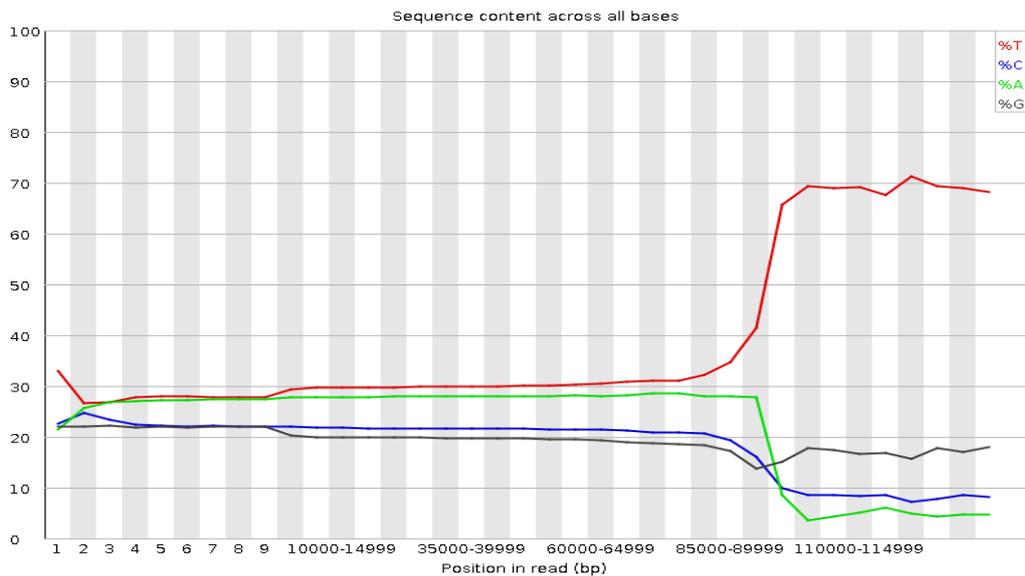
18A3183



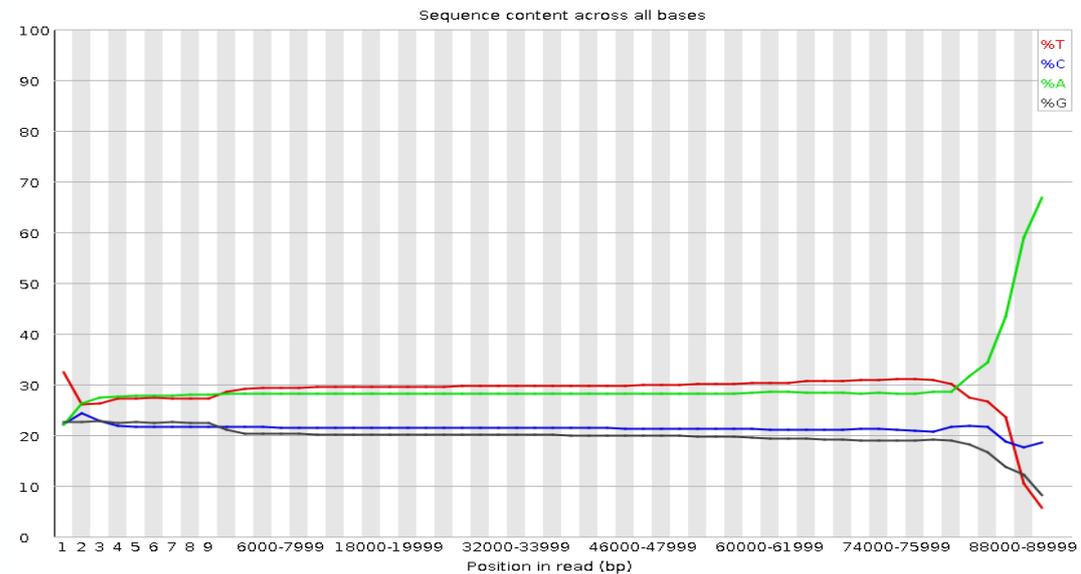
Score de qualité des bases arbitrairement fixé à 0 par PacBio car « **Trop de calcul pour le séquenceur** »....

QC des échantillons

18A3182

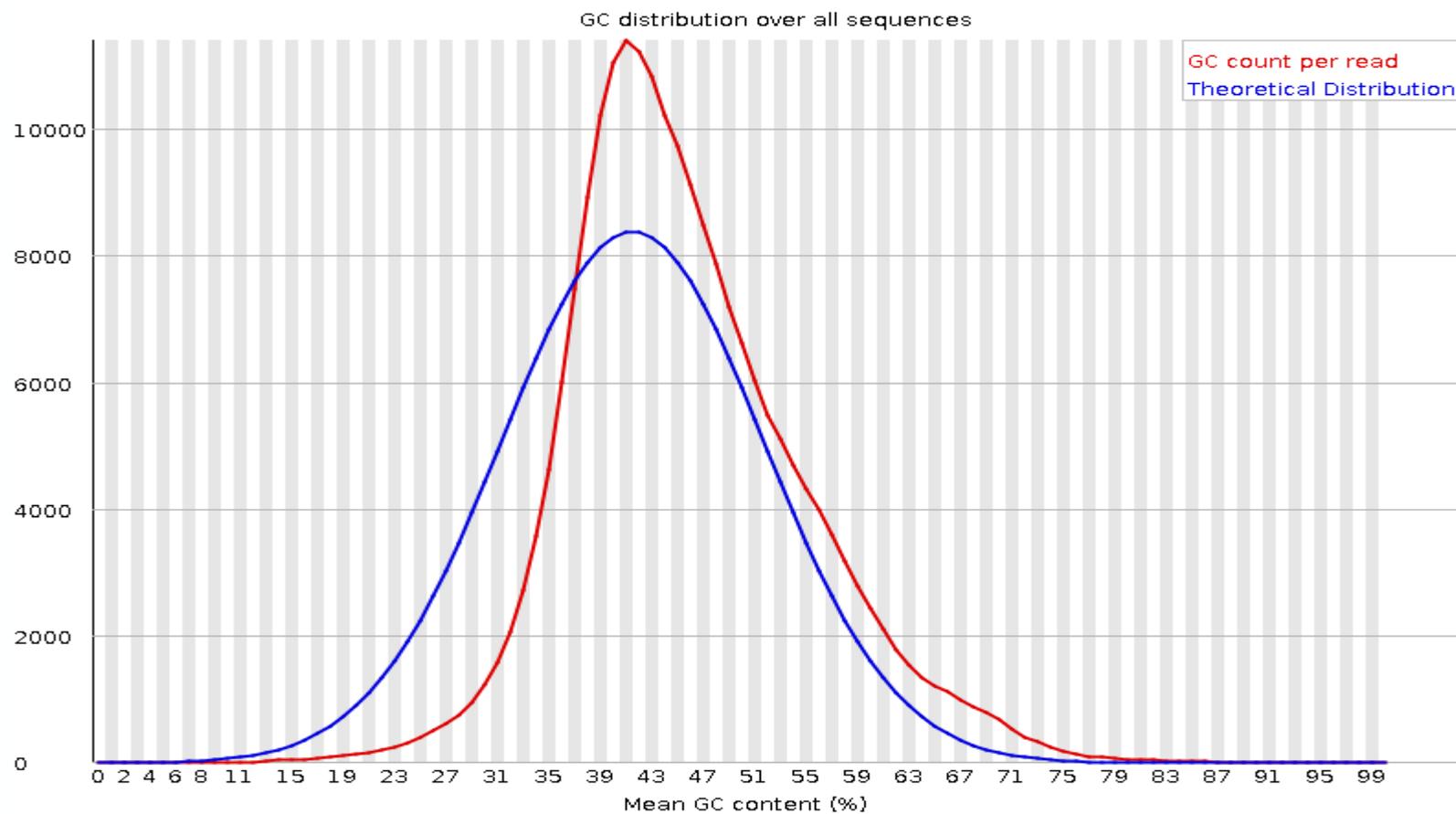


18A3183



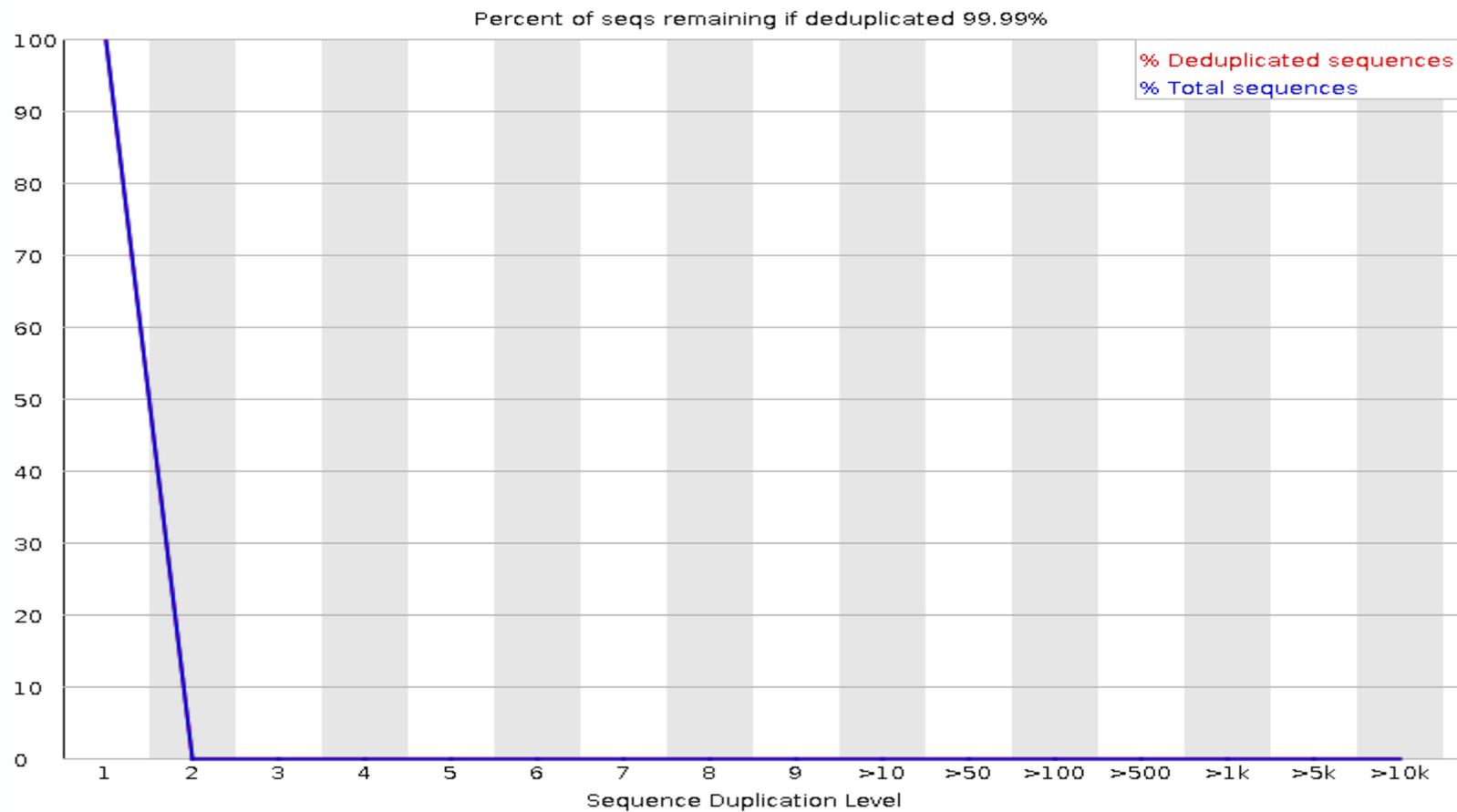
18A3182

18A3183



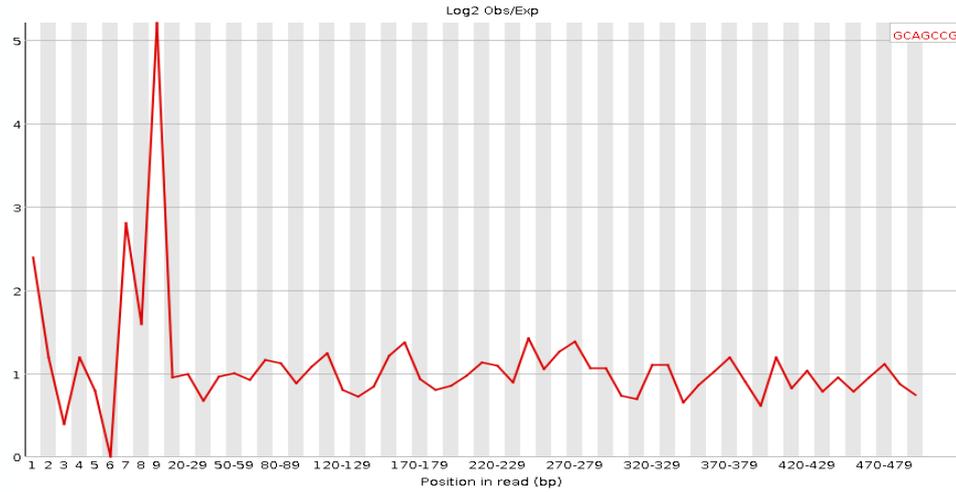
18A3182

18A3183



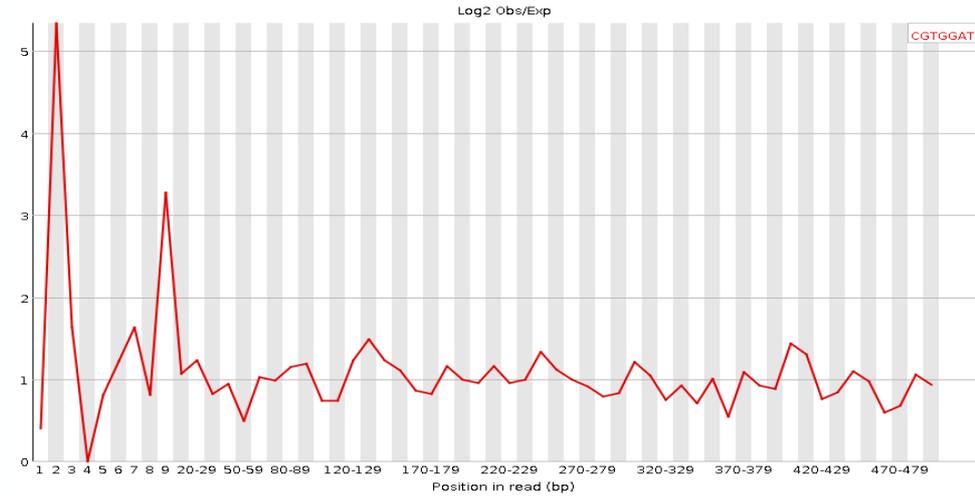
QC des échantillons

18A3182



Sequence	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
GCAGCCG	6060	0.006769838	5.206869	9

18A3183

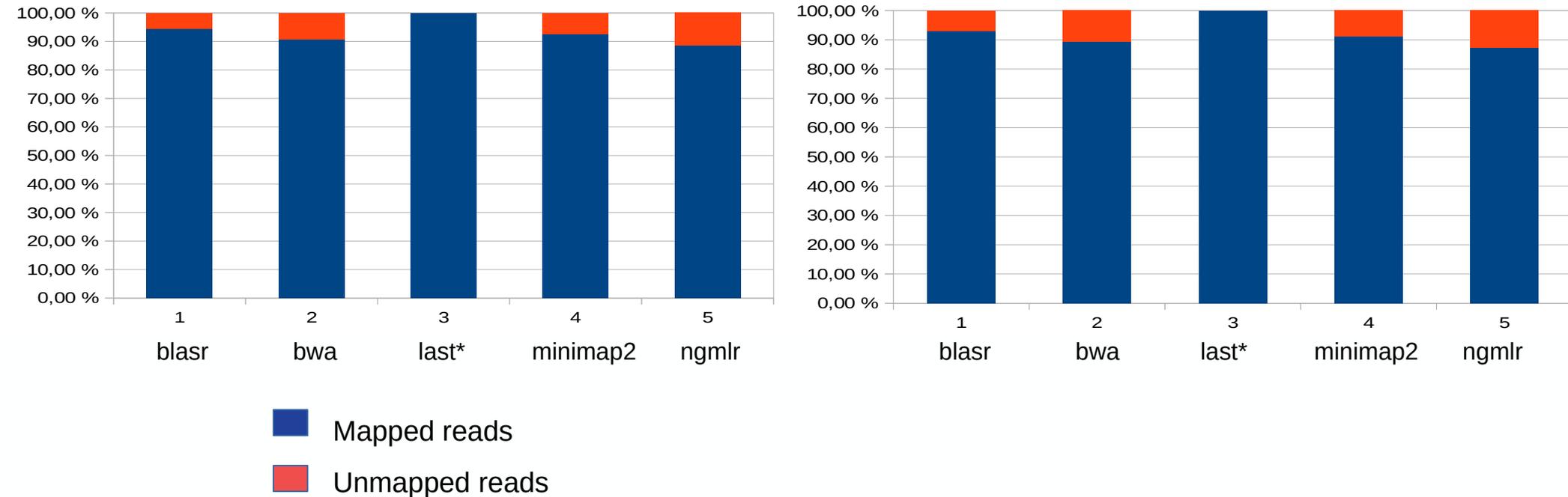


Sequence	Count	PValue	Obs/Exp Max	Max Obs/Exp Position
CGTGGAT	5915	0.0050614	5.337657	2

Alignments

18A3182

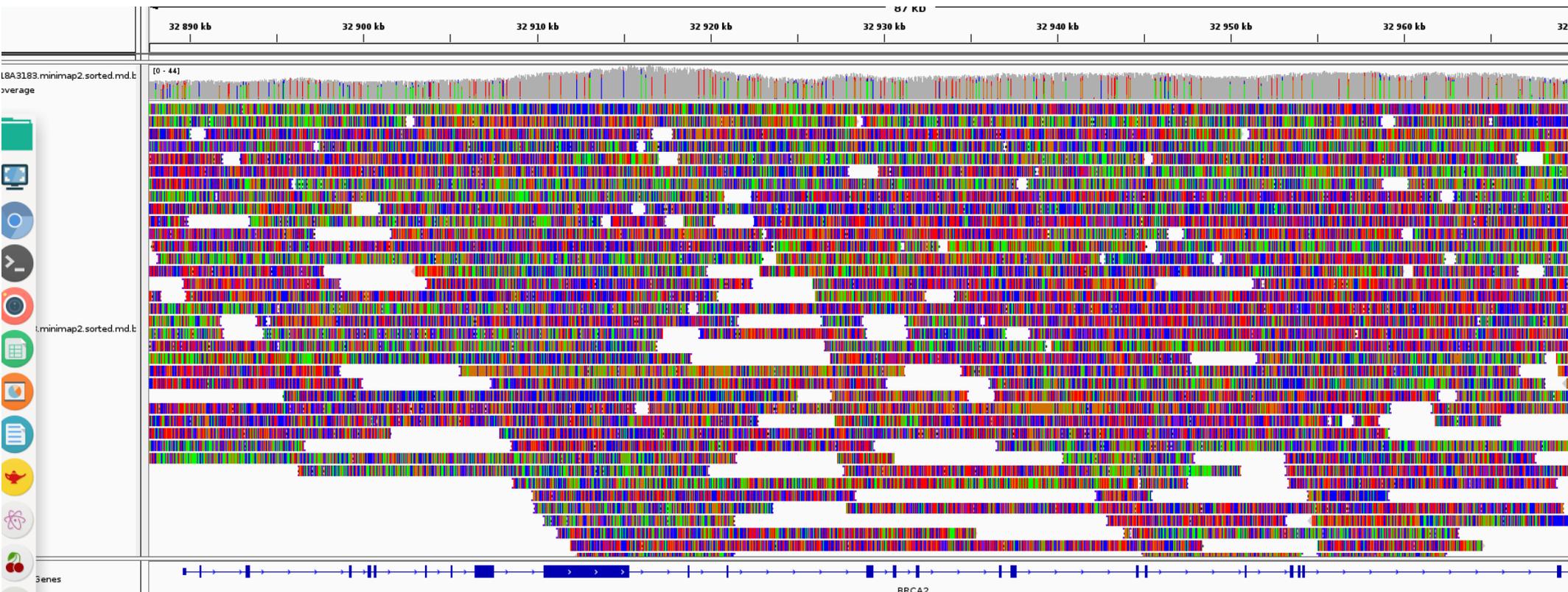
18A3183



Last produit 10 000 000 d'alignements / 4 500 000 reads !?!

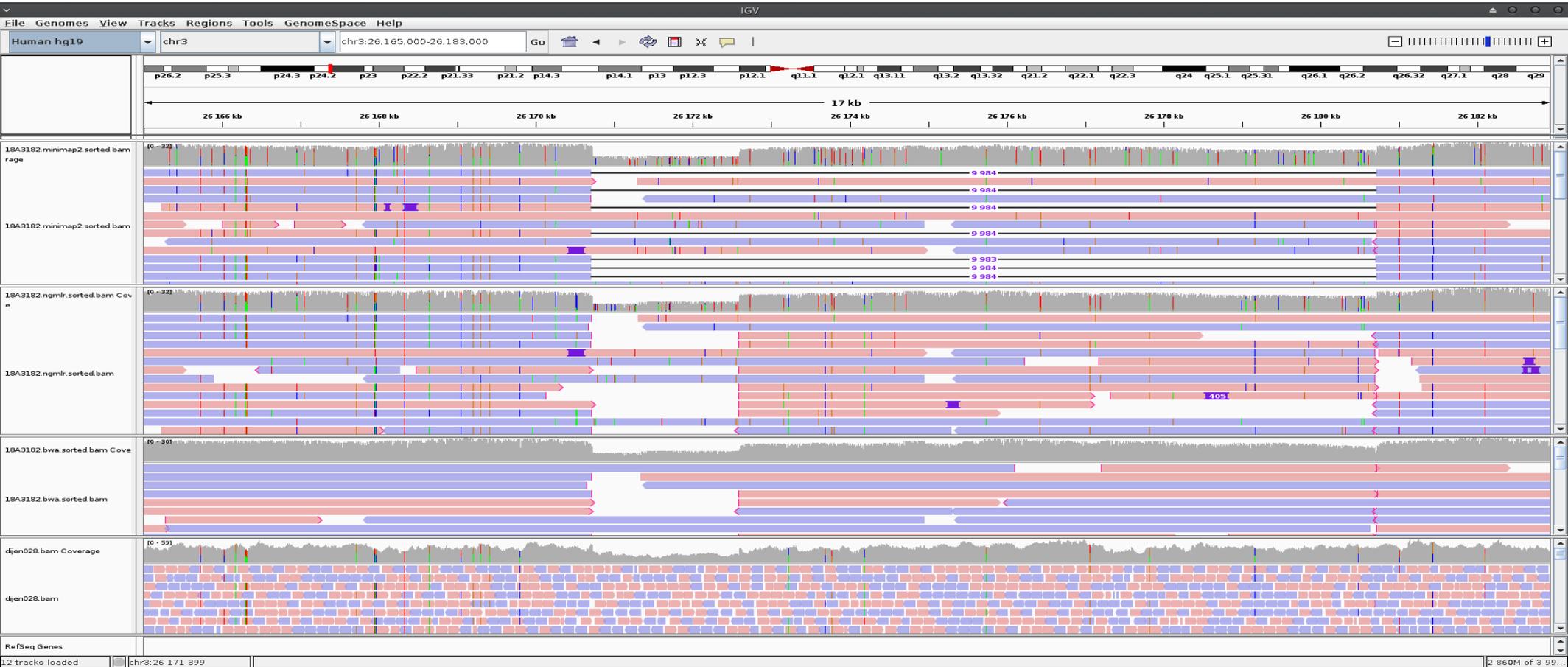
Pas de distinction primary / secondary / supplementary

Des SNVs ? Vraiment ?



Non ...

Alignments

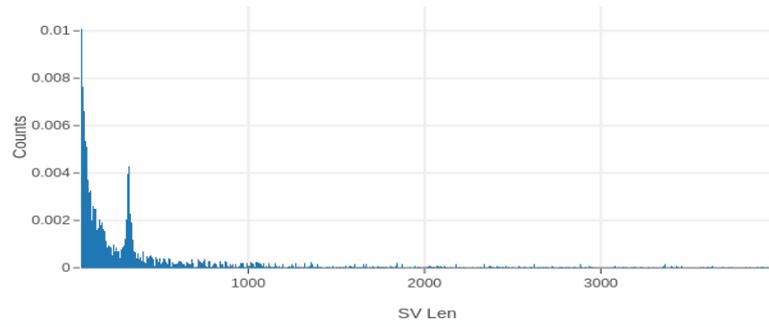


Sniffles v1.0.8

- Minimum number of reads that support a SV 3
- Maximum number of splits per read to be still taken into account 7
- Maximum distance to group SV together 1000
- Minimum length of SV to be reported 50
- Minimum Mapping Quality 20
- Discard read if non of its segment is larger than this 2000

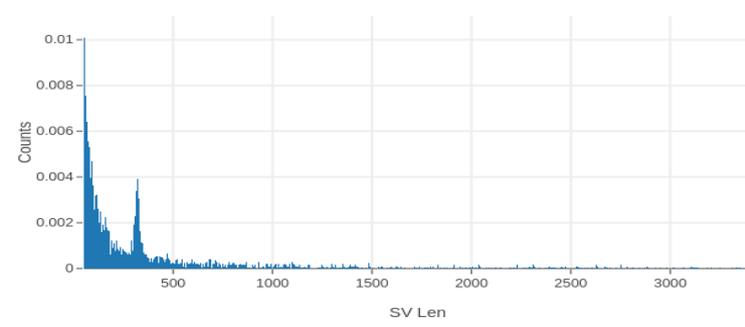
18A3182

Distribution



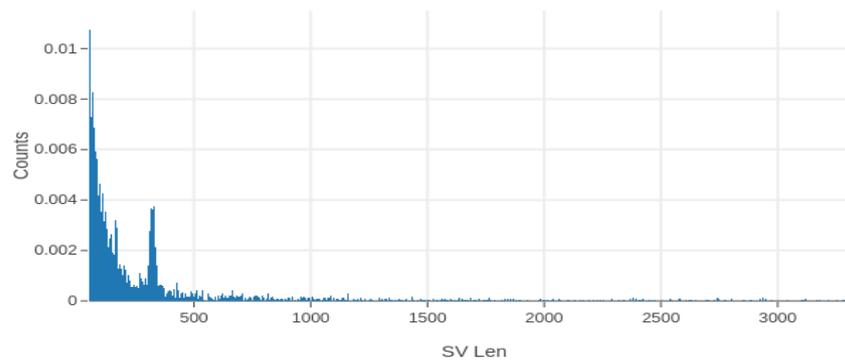
bwa

Distribution



ngmlr

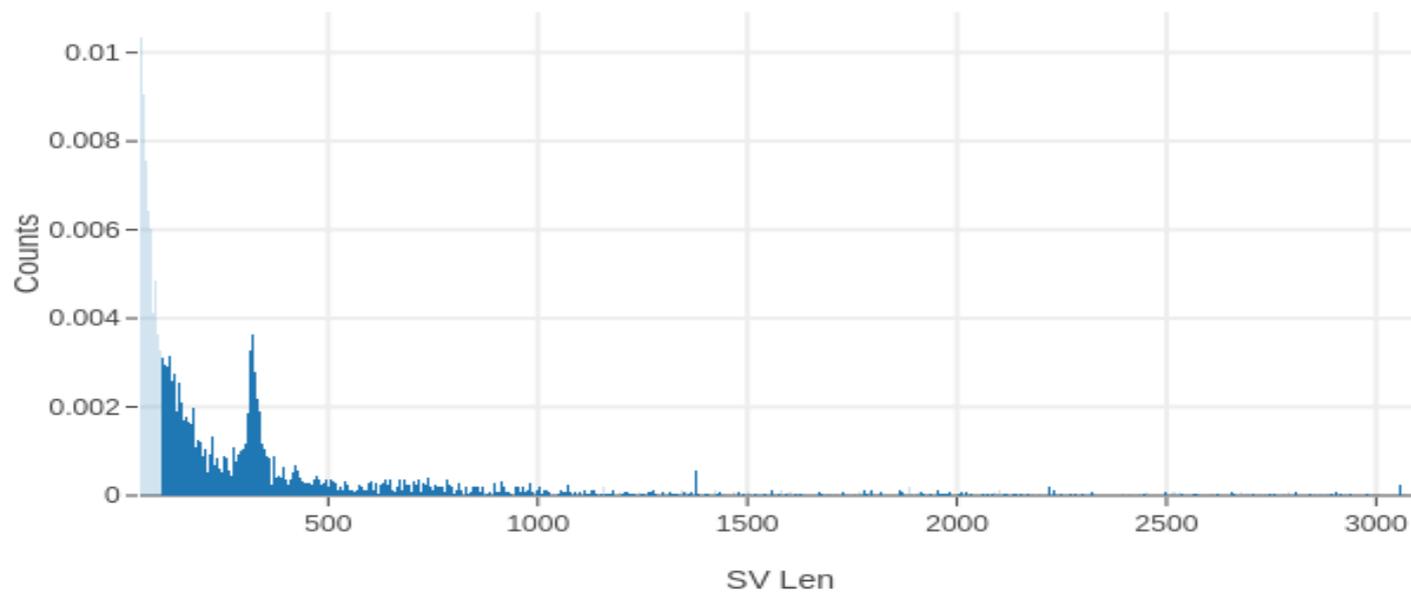
Distribution



minimap2

18A3183

Distribution

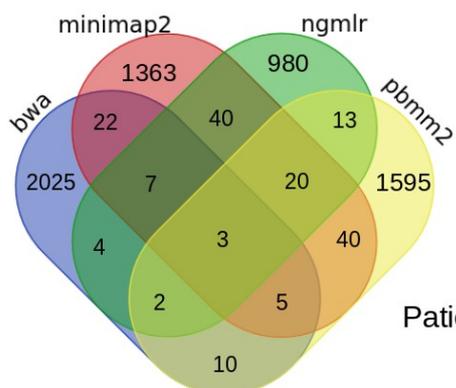


Recherche de variants de structure

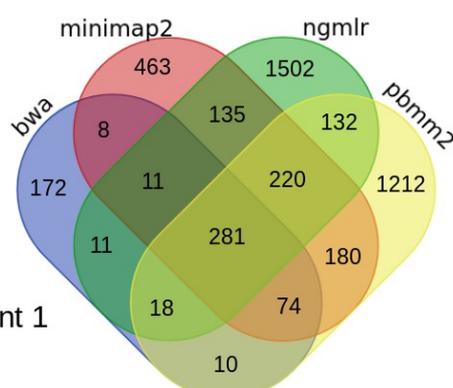
Number of breaking points with sniffles		
	Patient 1	Patient 2
Bwa/sniffles	2078	2197
Blasr/sniffles	0	0
Lordfast/sniffles	0	0
Ngmlr/sniffles	1102	1090
Minimap2/sniffles	1527	1576
pbmm2/sniffles	1685	1658
Number of breaking points with pbsv		
	Patient 1	Patient 2
Bwa/pbsv	585	545
Blasr/pbsv	0	0
Lordfast/pbsv	0	0
Ngmlr/pbsv	2310	2924
Minimap2/pbsv	1372	1732
pbmm2/pbsv	2127	1836

Recherche de variants de structure

Common Breakpoints with sniffles

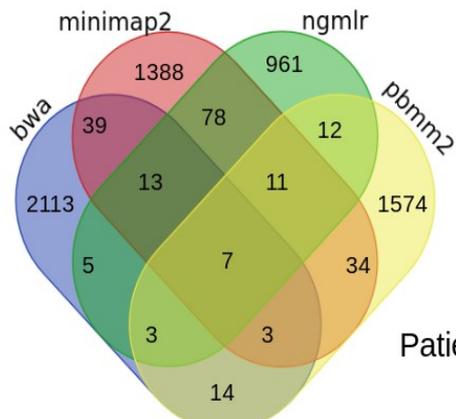


Common Breakpoints with pbsv

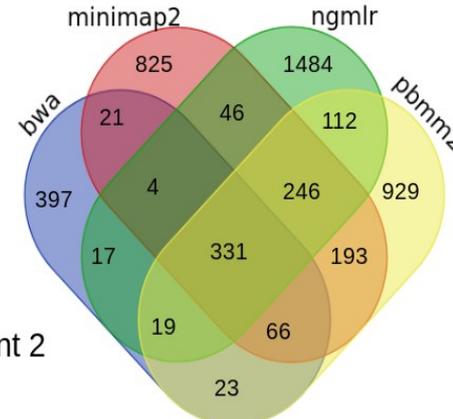


Patient 1

Common Breakpoints with sniffles



Common Breakpoints with pbsv



Patient 2

Number of breaking points with **sniffles**

	Patient 1	Patient 2
Bwa/sniffles	2078	2197
Blasr/sniffles	0	0
Lordfast/sniffles	0	0
Ngmlr/sniffles	1102	1090
Minimap2/sniffles	1527	1576
pbmm2/sniffles	1685	1658

Number of breaking points with **pbsv**

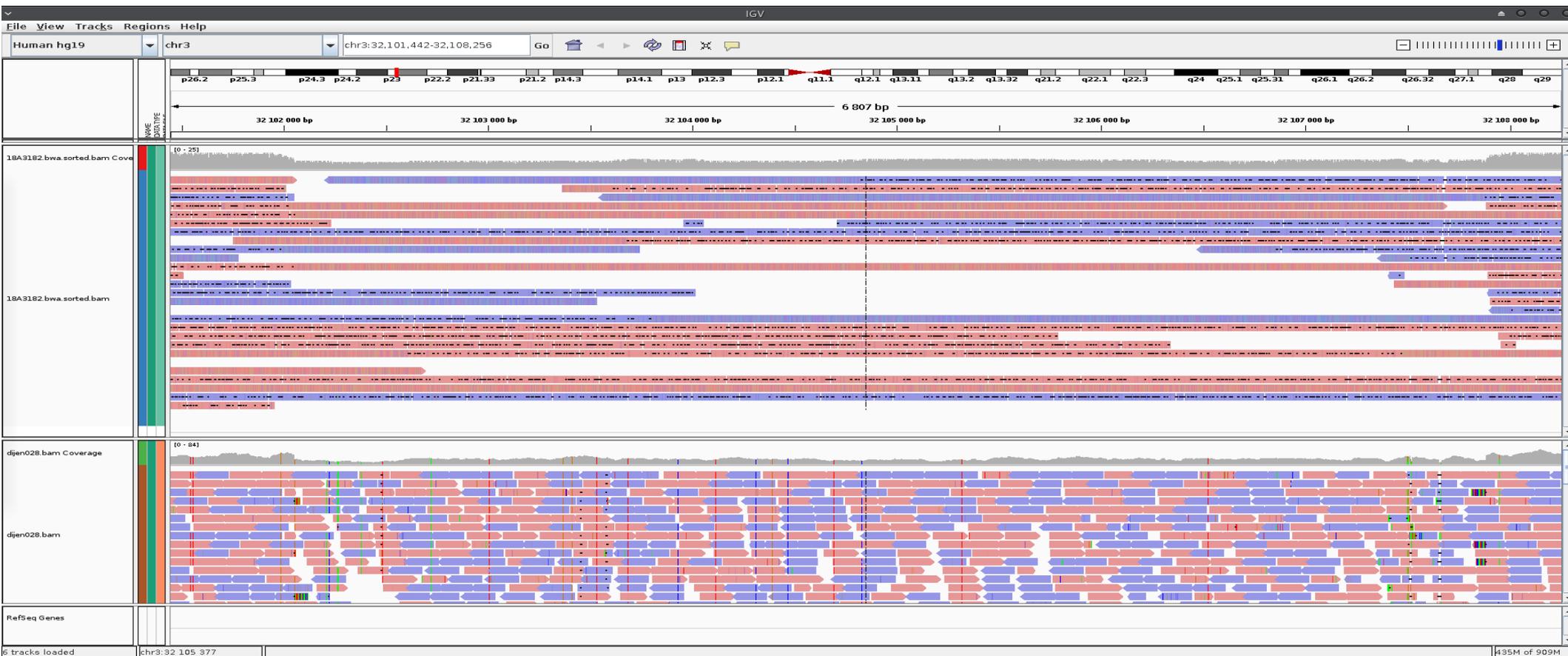
	Patient 1	Patient 2
Bwa/pbsv	585	545
Blasr/pbsv	0	0
Lordfast/pbsv	0	0
Ngmlr/pbsv	2310	2924
Minimap2/pbsv	1372	1732
pbmm2/pbsv	2127	1826

Recherche de variants

18A3182

chr3:32101442-32108256

VALIDÉ

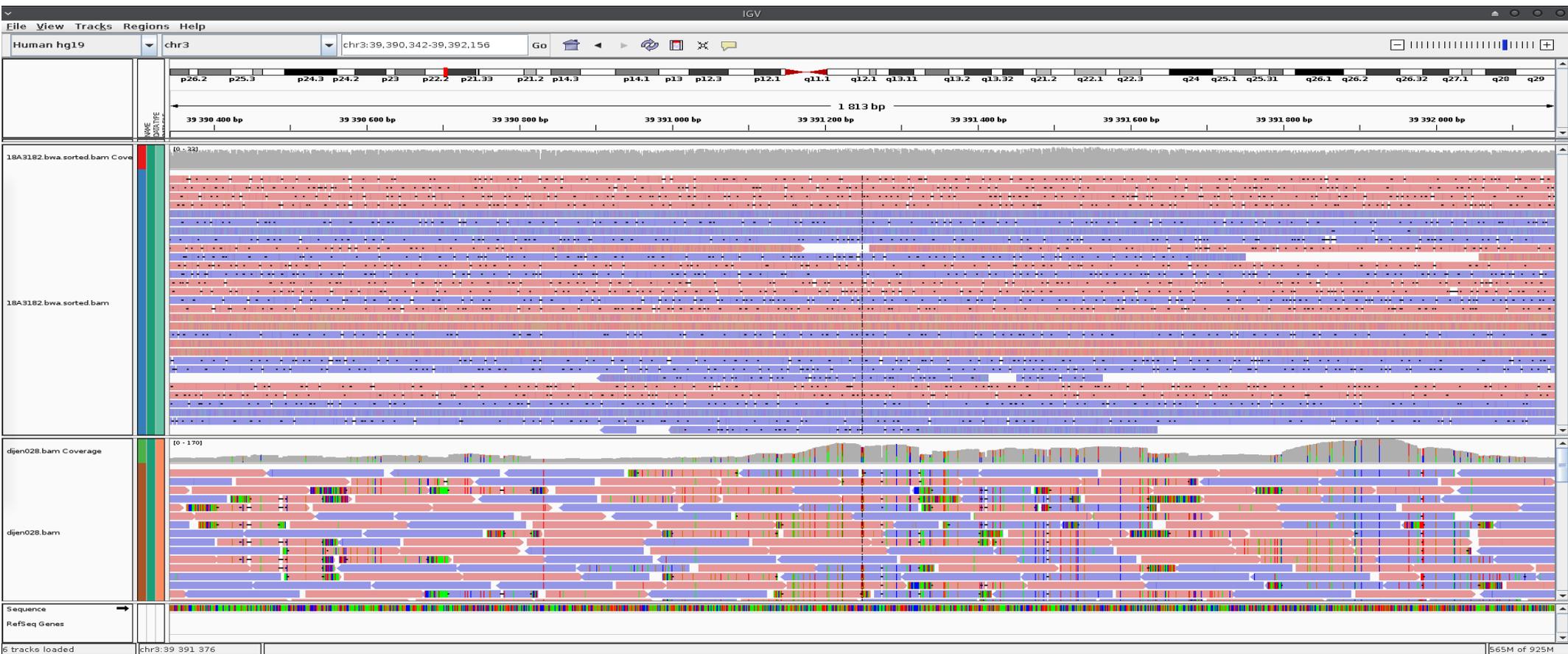


Recherche de variants

18A3182

chr3:39390342-39392156

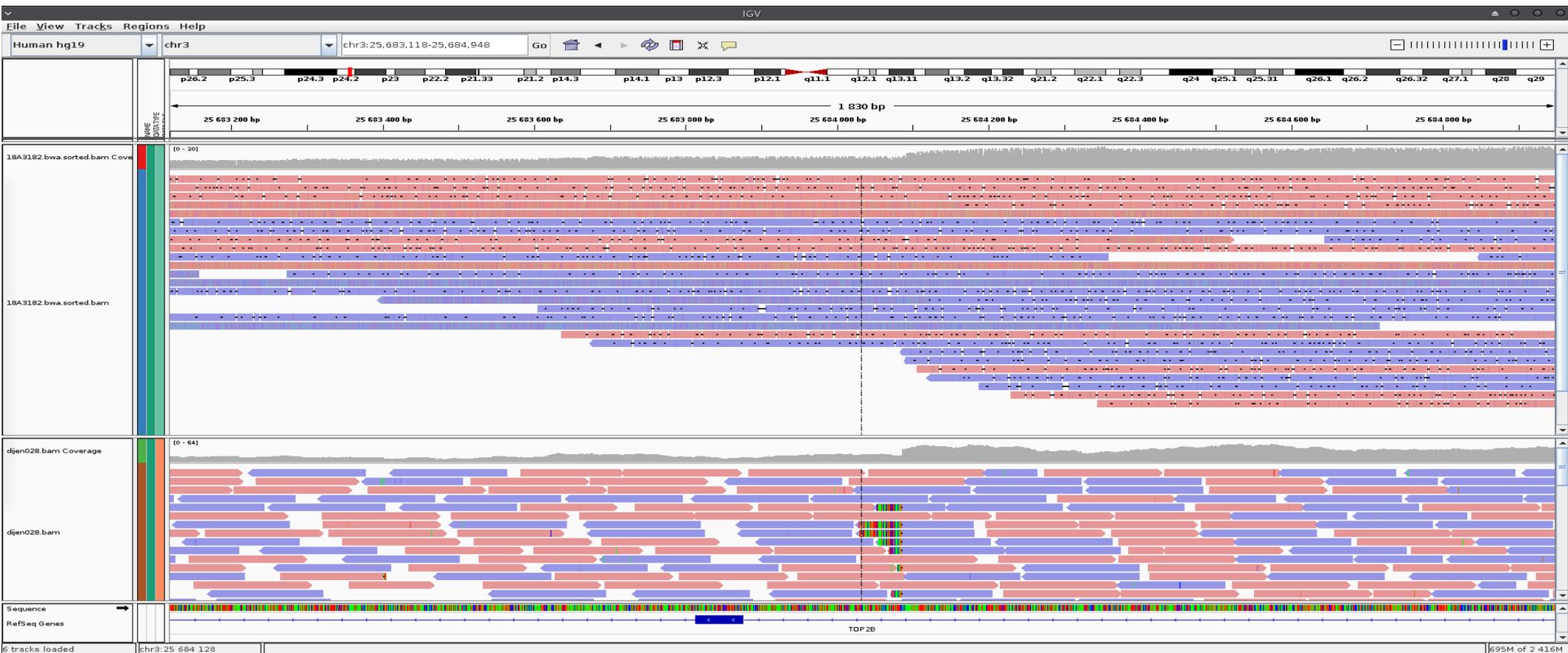
⚠ NOT FOUND



18A3182

chr3:25569542-25684356

⚠ NOT FOUND

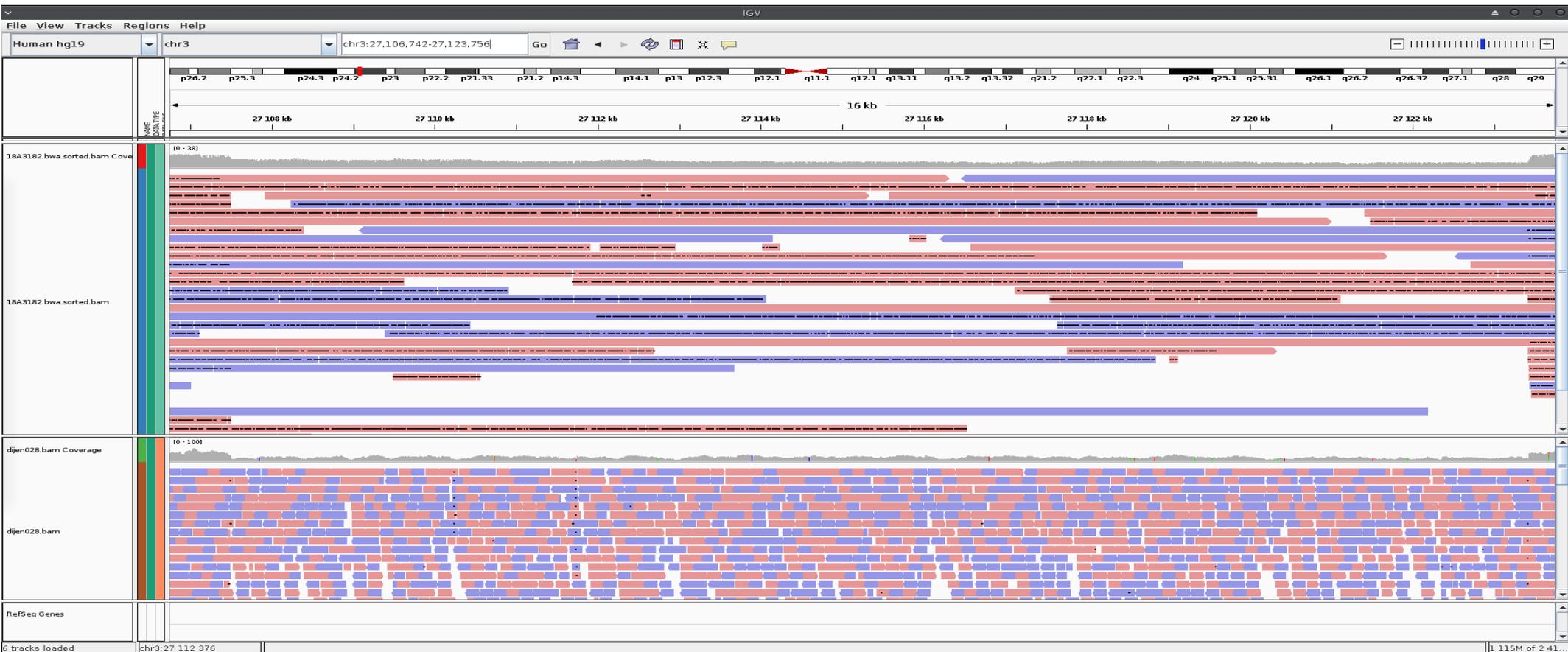


Détection d'une translocation vers chr12:58449408

18A3182

chr3:27106742-27123756

NOT FOUND



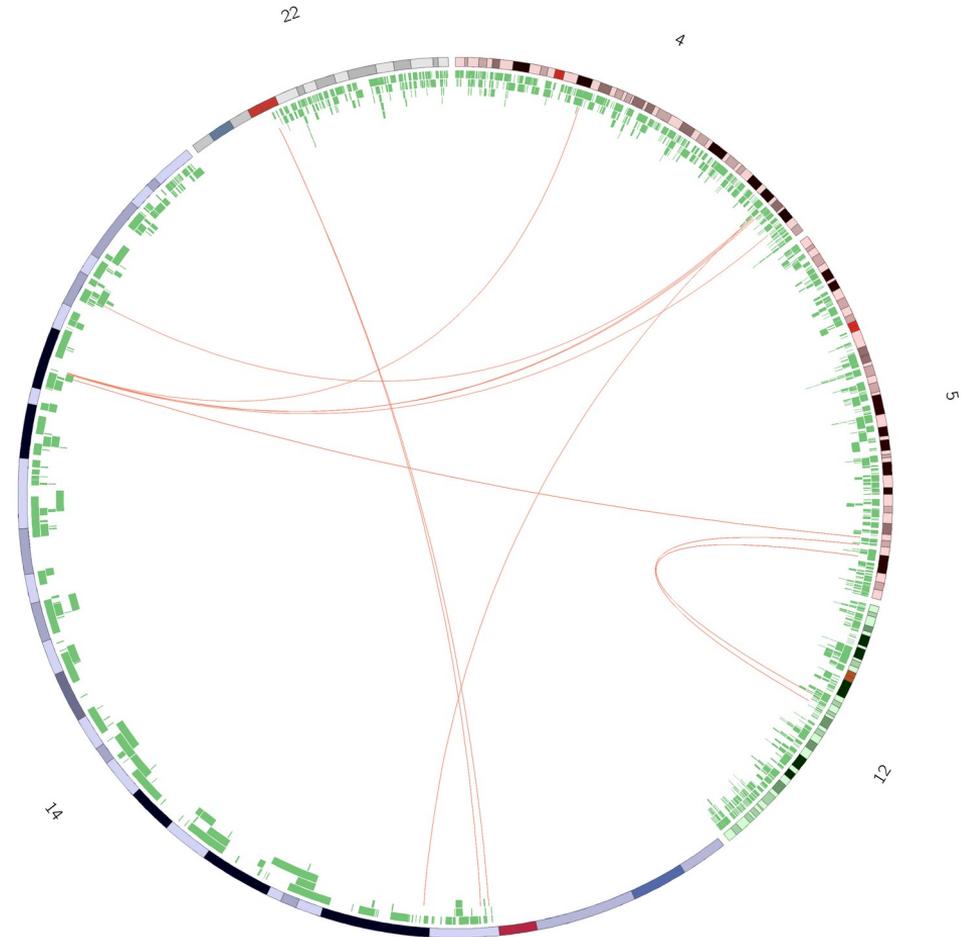
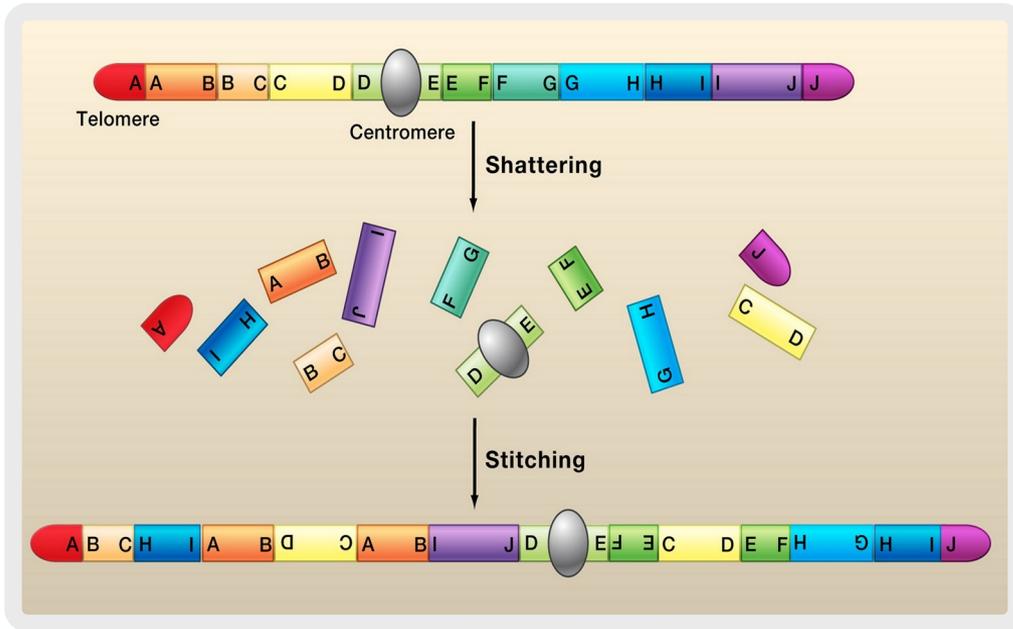
Détection d'une dup ...

Assemblages

		Falcon	
		Patient 1	Patient 2
Nombre de bases	de	2 802 857 930 pb	2 583 611 581 pb
Nombre de contigs		5 693	9 968
N25		2 569 884 pb	967 665 pb
N50		1 322 242 pb	545 808 pb
N75		579 117 pb	262 933 pb
Temps d'analyse réel		248h	218h

		wtdbg2	
		Patient 1	Patient 2
Nombre de bases	de	2 863 677 286 pb	2 894 105 539 pb
Nombre de contigs		10 238	15 636
N25		2 942 467 pb	1 449 401 pb
N50		1 738 674 pb	817 228 pb
N75		778 409 pb	383 290 pb
Temps d'analyse réel		18h	15h

Chromothripsis



Réassemblage d'un chromosome de chromothripsis (*Simon Verdez*)

Temps CPU consommé pendant le stage

Janvier 2019	Février 2019	Mars 2019	Avril 2019	Mai 2019
315h	133 517h	83 646h	60 109h	80 967h

Espace Disque mobilisé

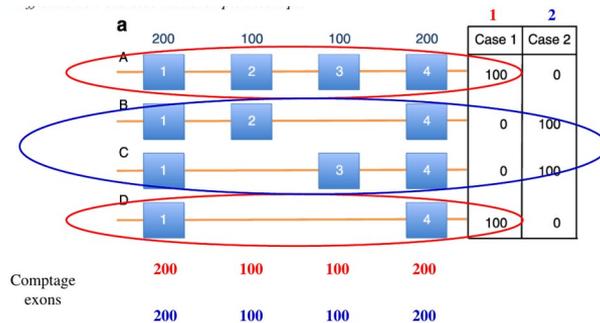
Alignement	Assemblage	Correction
7 To	1,5 To	1 To

Conclusion

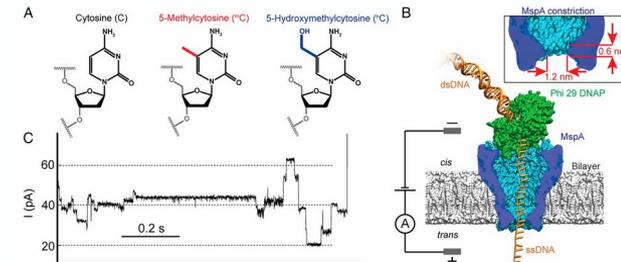
- Pas (encore?) la révolution attendue ...
- Technologies prometteuses
 - CNV & SV
 - Temps réel
- Actuellement inutilisable dans un contexte diagnostique (génétique humaine)
 - Taux d'erreur à la base : inutilisable pour les SNVs
 - Cout important (20X PacBio = 6000 €)
 - Développement bioinformatique important
 - Stabilité (10 versions de protocoles en 1 an ONT)

Pour aller plus loin ...

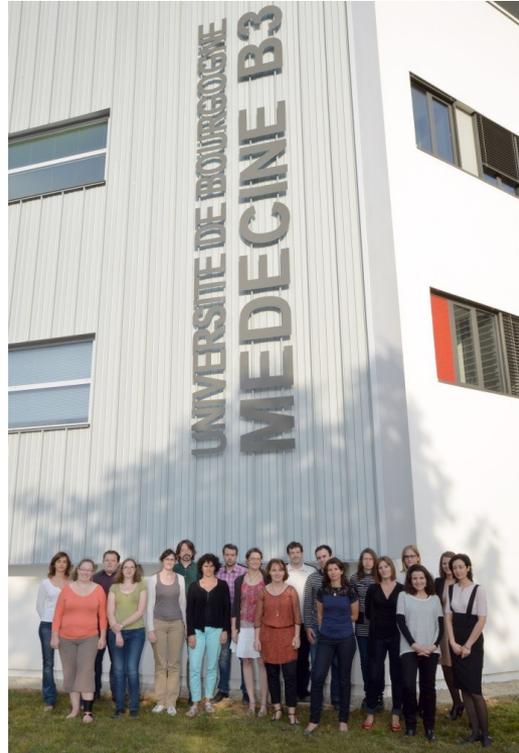
- ARN



- Méthylation



Merci de votre attention !



Questions Discussions

