



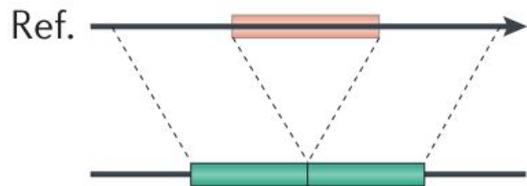
Structural Variant detection

Gabrièle Adam - INRAE

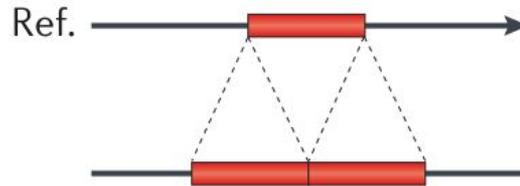
Définition

- Consensus actuel : Réarrangement génomique >50bp
- Différents types de variants structuraux :
 - Réarrangements déséquilibrés (variation du nombre de copie - CNV)
 - Délétion
 - Duplication
 - Réarrangements équilibrés
 - Insertion
 - Inversion
 - Translocation

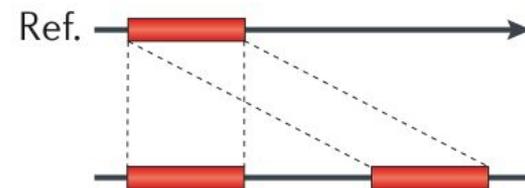
Deletion



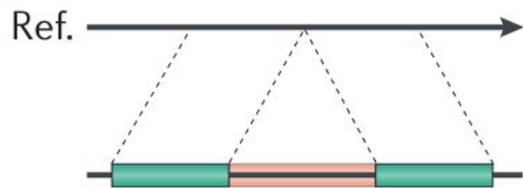
Tandem duplication



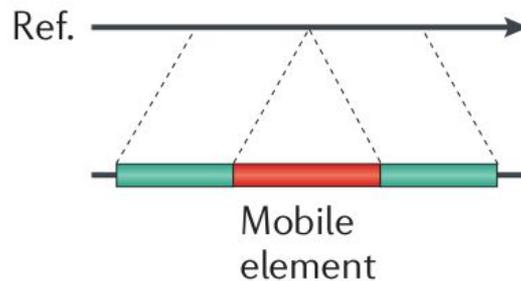
Interspersed duplication



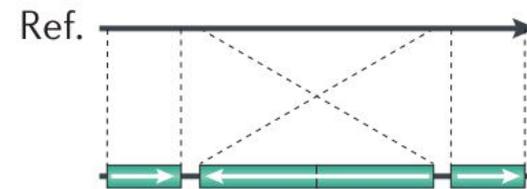
Novel sequence insertion



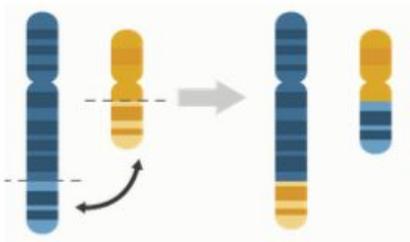
Mobile-element insertion



Inversion

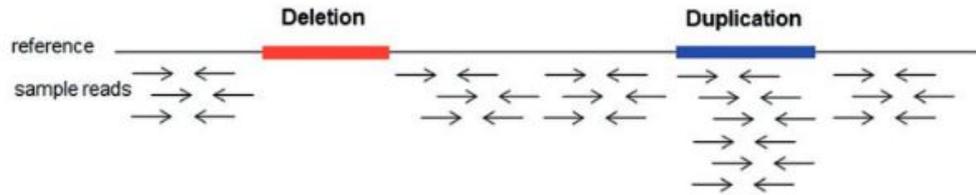


Translocation

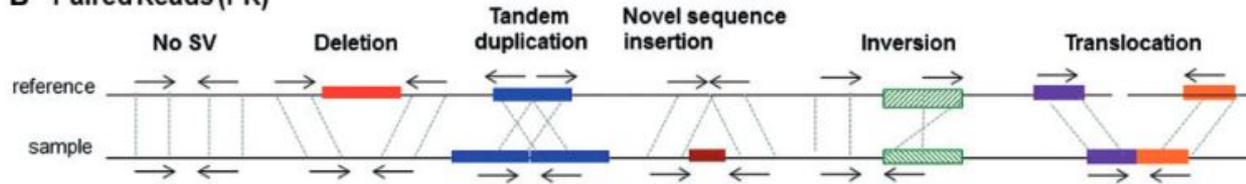


Principe de détection des SVs

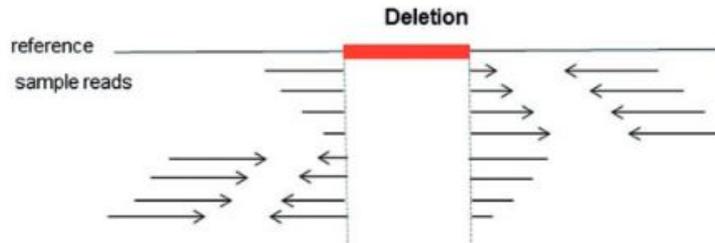
A Read Depth (RD)



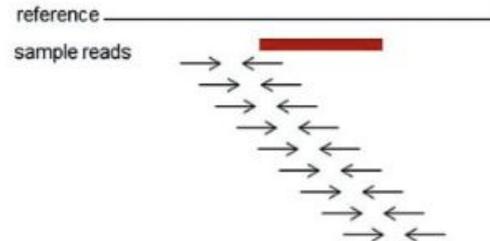
B Paired Reads (PR)



C Split Reads (SR)



D. De Novo Assembly (AS)



Review > Brief Funct Genomics. 2015 Sep;14(5):305-14. doi: 10.1093/bfgp/evl014. Epub 2015 Apr 15.

A decade of structural variants: description, history and methods to detect structural variation

Georgia Escaramis, Elisa Docampo, Raquel Rabionet

Short reads ou long reads?

Short reads (Illumina) : selon l'outil et la qualité des données

- **faible recall** : 10 à 70% des SVs détectés
- **faible précision** : jusqu'à 90% de Faux Positifs
- Difficulté à caractériser des SVs complexes (alignement imprécis dans les régions répétées et faible résolution)

!! Un calling consensus avec plusieurs outils de détection peut être utile avec des données short reads !!

Long reads (PacBio/MinION) :

- Meilleure caractérisation des altérations des régions répétées
- Une faible profondeur de couverture suffit (15-30x)

Quel outil choisir ?

Critères de choix :

- Ai-je des données short reads ou long reads ?
- Ai-je de nombreux échantillons ?
- Quel type de SV est-ce que je recherche ?
- Est-ce que la profondeur de couverture est suffisante ?
- Que privilégier : sensibilité et / ou spécificité
- Quel est le format de sortie de l'outil ?

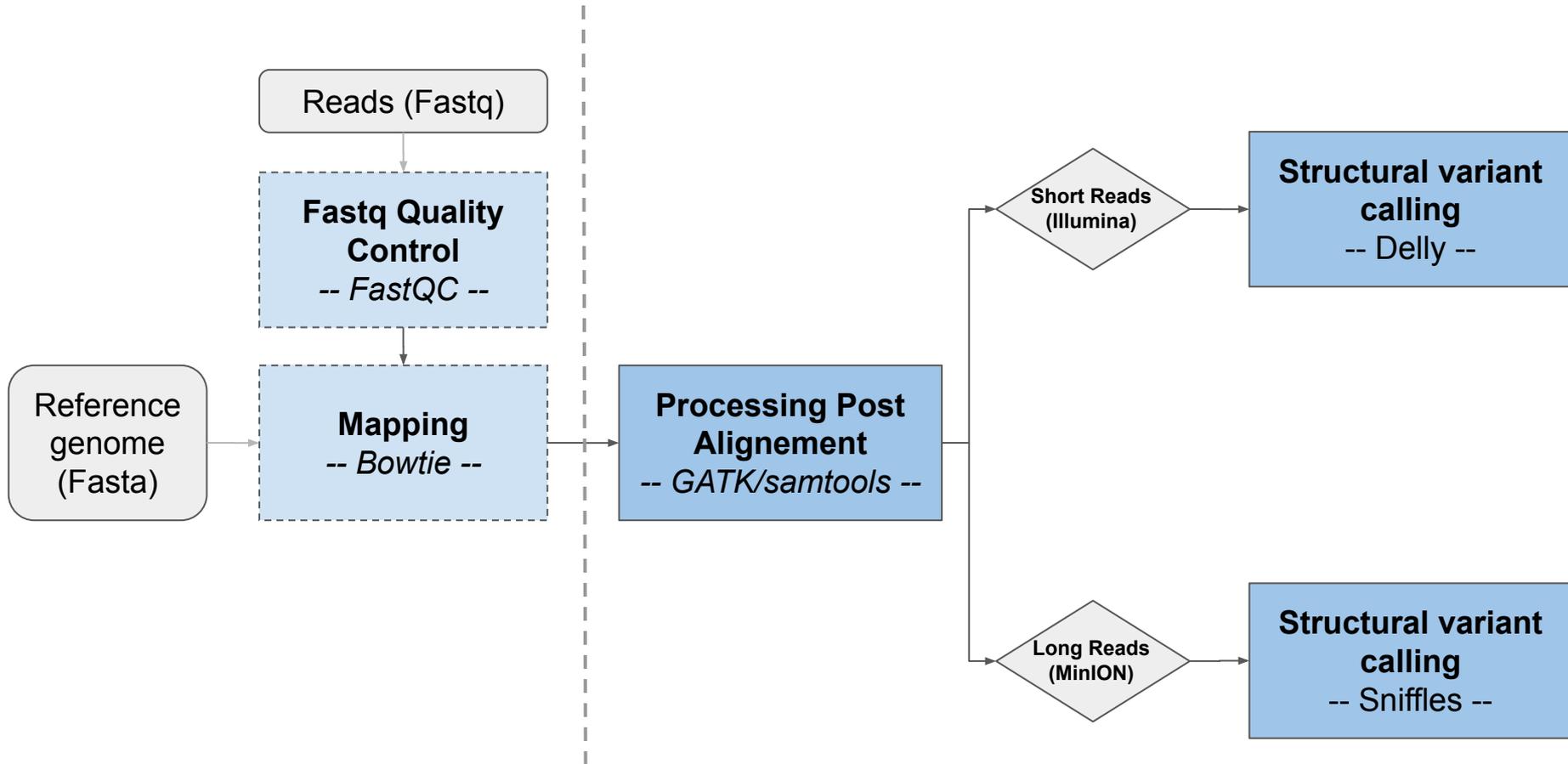
Détection de SV pour données short reads

	SV Callers	SV Types					Data	Anomalously Mapped Reads Used					Techniques				BP Resolution (Y/N)	References					
								Discovery Stage					Validation Stage										
		CNV	INS	DEL	DUP	INV		TRA	RD	SC	PR	OEA	UM	RD	SC	PR			OEA	UM	CL	SA	CA
CNV	BIC-seq	x					PE;SE	x					x							x	N	[110]	
	cn.MOPS	x					PE;SE	x					x							x	N	[44]	
	cnD	x					PE	x					x							x	N	[88]	
	CNVeM	x					PE	x					x							x	N	[105]	
	CNVnator	x					PE;SE	x					x							x	N	[3]	
	CNV-seq	x					PE;SE	x					x							x	N	[111]	
	JointSLM	x					PE;SE	x					x							x	N	[59]	
	RDXplorer	x					SE	x					x							x	N	[115]	
	SegSeq	x					PE;SE	x					x							x	N	[15]	
CNVer	x					PE						x					x		x	N	[62]		
SV	LUMPY			x	x	x	PE	x	x	x			x	x	x			x	x		N	[50]	
	MetaSV		x	x	x	x	PE	x	x	x			x	x	x			x	x	x	Y	[65]	
	SVM2		x	x			PE	x	x	x			x	x	x			x			N	[16]	
	Breakpointer		x	x			SE	x					x	x				x			N	[95]	
	Meerkat		x	x	x	x	PE	x	x	x			x			x		x	x		Y	[112,113]	
	Scalpel		x	x			PE		x	x	x									x	Y	[68]	
	SVMerge		x	x	x	x	x		x	x	x			x	x	x			x	x	Y	[109]	
	SoftSV		x	x	x	x	PE		x	x				x	x				x	x	Y	[9]	
	BreakKmer		x	x	x	x	PE				x				x						Y	[2]	
	ClipCrop		x	x	x	x	x	PE						x					x	x	Y	[97]	
	CREST		x	x	x	x	PE;SE		x					x							x	Y	[104]
	Gustaf		x	x	x	x	PE;SE		x					x						x	Y	[99]	
	Socrates		x	x	x	x	PE;SE			x				x						x	x	Y	[86]
	Bellerophon						PE				x			x	x					x	x	Y	[30]
	BreakDancer		x	x	x	x	PE			x				x						x		N	[14]
	CLEVER		x	x			PE			x						x					x	N	[60]
	DELLY			x	x	x	PE			x				x		x				x	x	Y	[80]
	FACTERA			x			PE			x				x						x	x	Y	[69]
	GASV		x	x	x	x	PE			x						x						N	[90]
	GASVpro		x	x	x	x	PE			x				x								N	[91]
	GenomeSTRIP			x			PE			x					x						x	N	[29]
	HYDRA			x	x	x	PE			x					x							Y	[78]
	HYDRA-Multi			x	x	x	PE			x					x							Y	[58]
	inGAP-SV		x	x	x	x	PE			x				x								N	[76]
	MoDIL		x	x			PE			x											x	N	[51]
	PEMer		x	x	x	x	PE			x										x		N	[45]
	PeSV-Fisher			x	x	x	PE;MP			x				x								N	[21]
	PRISM		x	x	x	x	PE			x						x				x	x	Y	[37]
	RetroSeq		x				PE			x				x						x	x	Y	[40]
	SVDetect		x	x	x	x	PE;MP			x					x							N	[116]
	SVMiner		x	x			PE			x				x		x						N	[31]
	Ulysses		x	x	x	x	MP			x				x		x						N	[25]
	VariationHunter		x	x			PE			x					x							N	[32]
	NovelSeq		x				PE				x	x								x	x	Y	[27]
	PINDEL		x	x			PE			x											x	Y	[114]
	SLOPE		x	x			PE;SE			x						x				x	x	Y	[1]
	SOAPindel		x	x			PE			x					x	x	x					Y	[55]
	Splitread		x	x			PE			x										x		Y	[39]
	BreakSeq		x	x			PE							x						x		Y	[47]
	SMUFIN		x	x		x	x	PE												x	x	Y	[66]

Outils en long reads

- PBHoney, 2014
- SMRT-SV, 2015
- Hysa, 2016 (hybrid avec short reads)
- NanoSV, 2017
- Sniffles, 2018

Workflow



Partie TP

Data : souche de *Zymoseptoria tritici* séquencées à la fois en Illumina et en MinION.

→ chaque set de reads a été aligné sur le génome de référence avec les outils dédiés

→ les données ont été réduites aux premiers 500kb du chr10

Tools :

- **Delly** (*Bioinformatics*, Volume 28, Issue 18, 15 September 2012, Pages i333-i339, <https://doi.org/10.1093/bioinformatics/bts378>)
- **Sniffles** (*Nature Methods* volume 15, pages 461-468 (2018) , <https://www.nature.com/articles/s41592-018-0001-7>) with NGMLR mapping

Jeux de données #2 : SVs

Zymoseptoria tritici : Champignon ascomycète, pathogène du blé tendre, responsable d'une maladie foliaire (septoriose).

- Principale maladie du blé (jusqu'à 50% de perte de rendement).
- Haploïde, génome de 40 Mb séquencé en 2011 : 13 chromosomes essentiels + 8 chromosomes accessoires
- Souche séquencée avec **deux technologies** : Illumina et Minlon

Your turn !
Retrouvez les délétions de grande taille

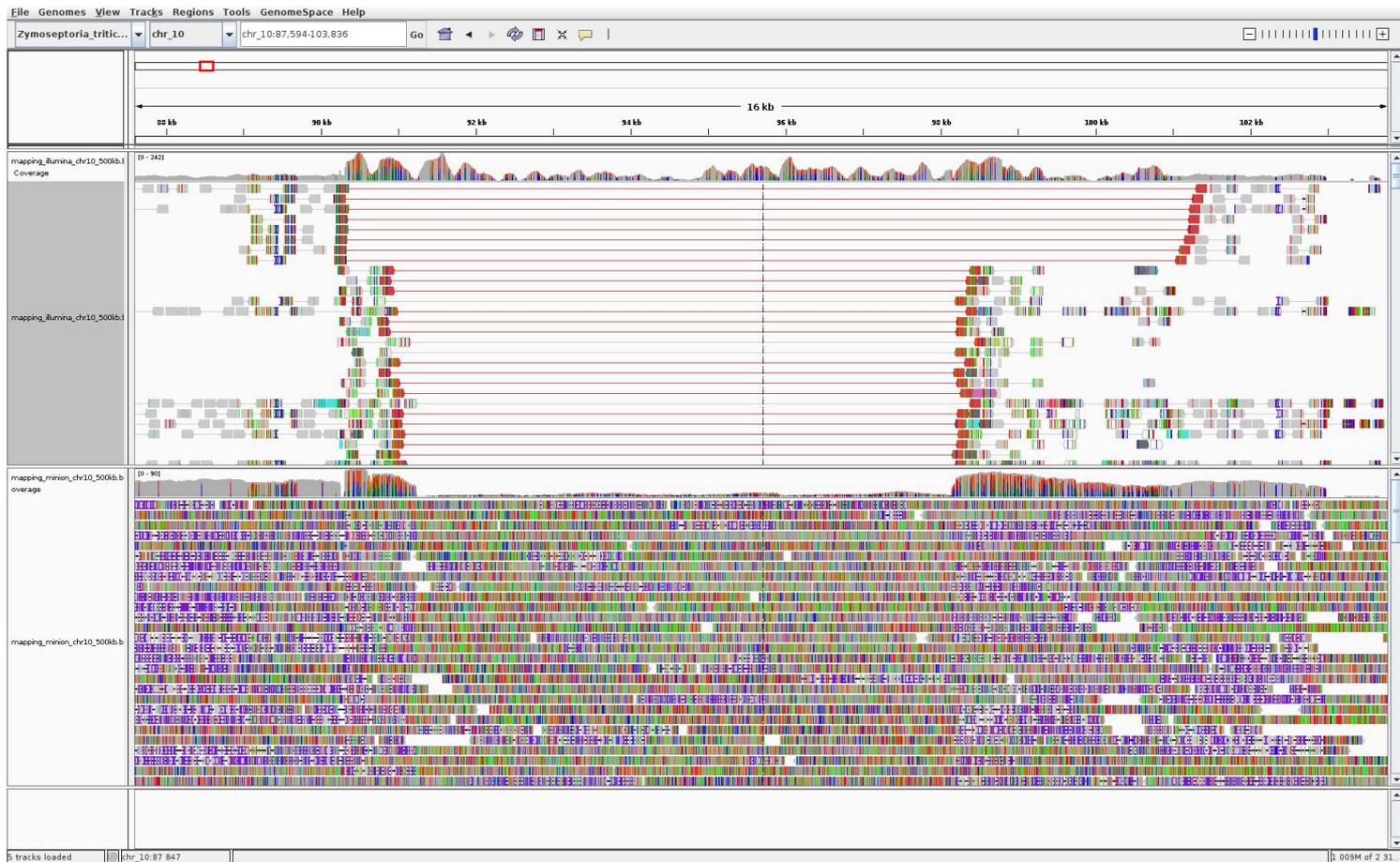


Aller au `jupyterNoteBook`

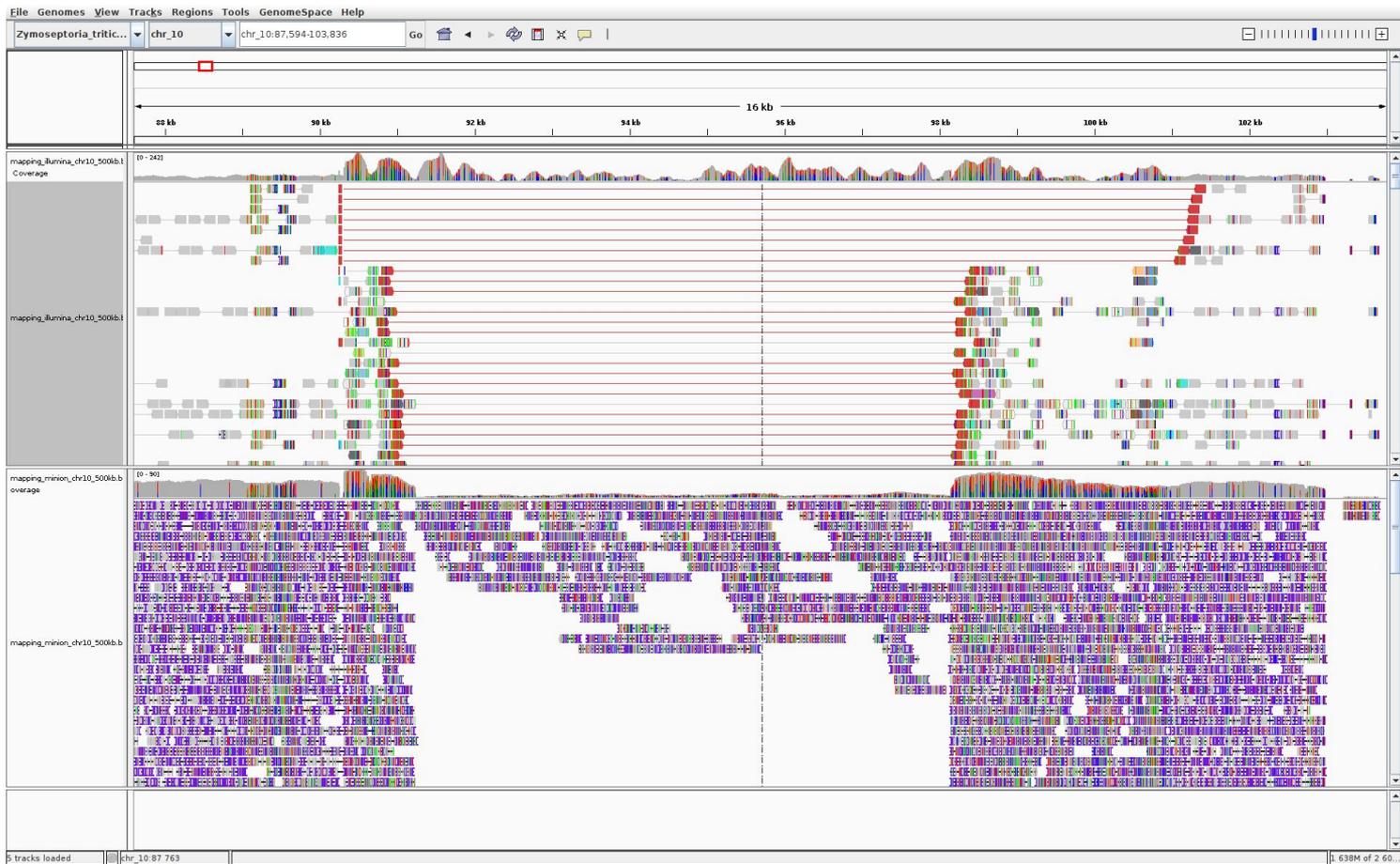
Visualisation sous IGV (Bonus)

- Télécharger en local les fichiers BAM et leurs index à travers votre session Jupyter
 - `Zymoseptoria_tritici.fa/fai`
 - `mapping_illumina_chr10_500kb.bam/bam.bai`
 - `mapping_minion_chr10_500kb.bam/bam.bai`
- Charger le génome de référence
- Ouvrir les fichiers BAM correspondant aux deux analyses (short et long reads)

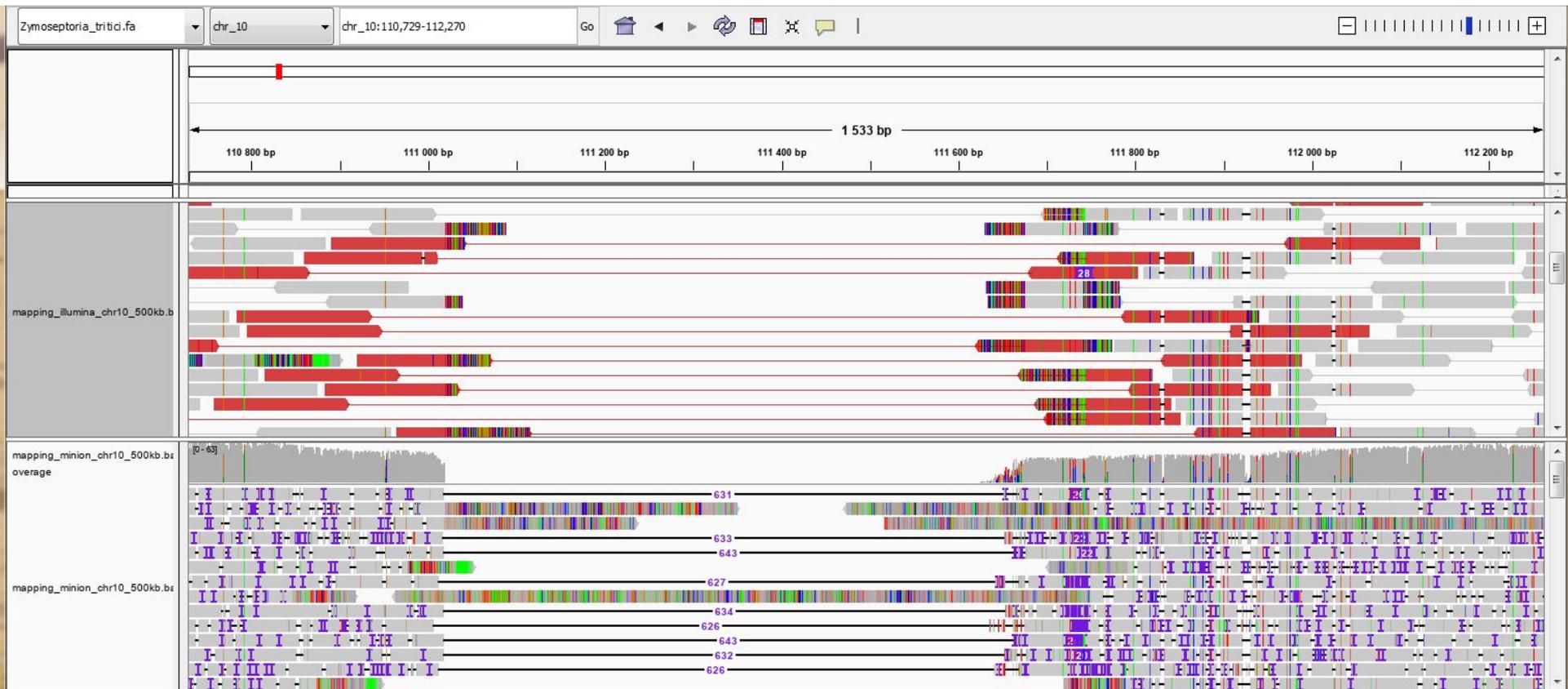
deletion 90309-101040 (illumina), 91233-98159 (Minion)



deletion 90309-101040 (illumina), 91233-98159 (Minion)



deletion 111021-111676



deletion 191291-191343



deletion 343161-343273



Comparaison des résultats de Delly et Sniffles

Delly (illumina)				
start	stop	precision	PE	SR
29522	29580	PRECISE	0	20
57127	57600	PRECISE	3	16
80015	80622	PRECISE	15	20
90255	90309	PRECISE	0	7
90309	101040	IMPRECISE	8	0
111021	111676	IMPRECISE	20	0
191291	191343	PRECISE	0	18
-	-	-	-	-
264986	265063	PRECISE	0	12
-	-	-	-	-
360628	361052	PRECISE	0	20
383682	477911	IMPRECISE	7	0
425686	426624	IMPRECISE	28	0
465858	466080	PRECISE	0	20
468192	468342	PRECISE	0	20
477523	479732	PRECISE	0	20
477526	479732	IMPRECISE	41	0

Sniffles (Minion)		
start	stop	precision
-	-	-
57126	57598	IMPRECISE
-	-	-
-	-	-
91233	98159	IMPRECISE
111020	111655	PRECISE
-	-	-
257001	257165	IMPRECISE
-	-	-
343161	343273	PRECISE
360638	361061	PRECISE
383681	477805	IMPRECISE
425682	426487	IMPRECISE
-	-	-
468192	468341	PRECISE
477525	479731	PRECISE
-	-	-

IGV OK

IGV ~OK

IGV doubt

IGV NO

Conclusion

- La détection des SVs **manque de précision** et engendre des faux positifs et faux négatifs
 - **Nécessité de croiser différents outils/technologies**
 - **Nécessité de bien utiliser les métriques des outils**
 - **Nécessité d'une bonne profondeur (variant hétérozygote)**
- Vérifier **visuellement les résultats sur IGV** permet d'augmenter la confiance dans les SVs détectés