

Mise en place d'un outil *in silico* de détection des délétions de l'ADN mitochondrial

HOFFMANN Virginie
Stage M2 Bioinformatique
Université de Nantes
2015-2016

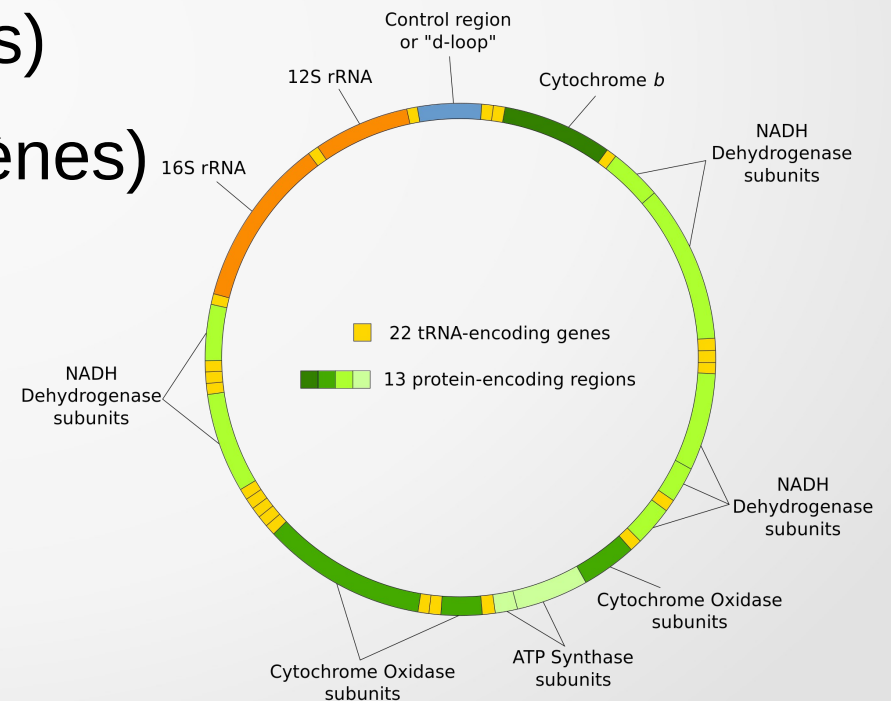
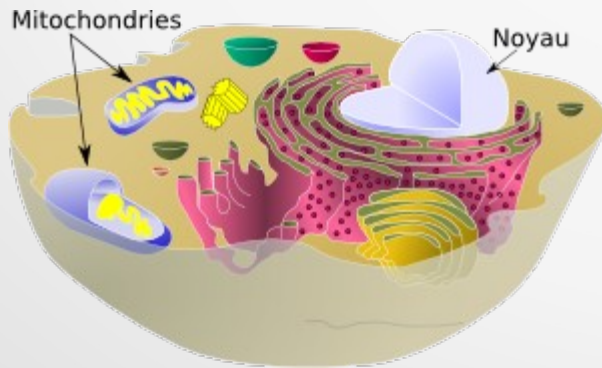
Sous l'encadrement du Pr. V. Procaccio et Dr. D. Goudenège
(financement GRIOTE)



Introduction

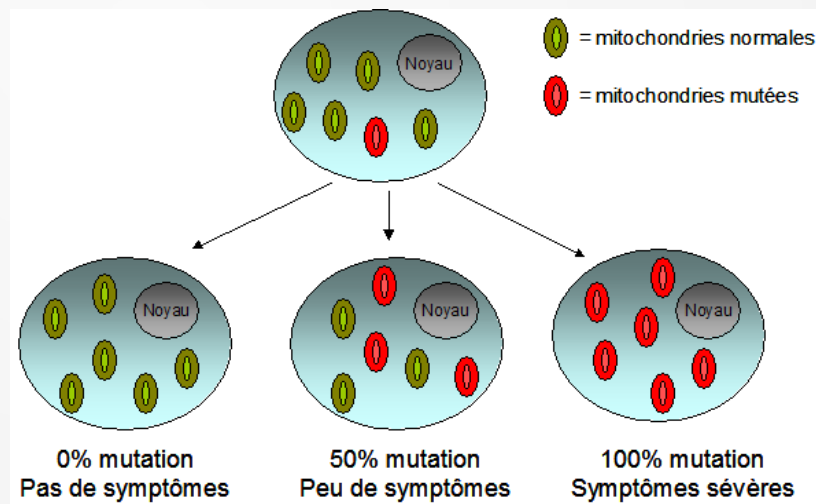
Introduction : la mitochondrie

- organe intracellulaire
- centrale énergétique de nos cellules
- sous le contrôle de deux génomes
 - l'ADN mitochondrial (37 gènes)
 - l'ADN nucléaire (+ de 1500 gènes)



Introduction : hétéroplasmie

- génome mitochondrial présent en de nombreuses copies
- coexistence entre forme sauvage et mutée
 - hétéroplasmie
 - effet seuil



Les maladies mitochondriales

- maladies rares (1/5000)
- mais phénotypes très hétérogènes
 - dysfonctionnement de la chaîne respiratoire
 - impact sur les organes les plus demandeurs en énergie (cerveau, coeur, muscle ...)

Les maladies mitochondriales

- mitopathies principalement issues de mutations
 - mutations ponctuelles : SNP, insertion, délétion
 - variants structuraux : inversion, transposition, CNV* (insertion, délétion, duplication)
- Les CNVs les plus reportés sur l'ADNmt sont :
 - délétion simple dites sporadiques
 - délétions multiples, liées à une mutation de la machinerie de réplication

(*) copy number variation

Diagnostic des mitopathies

Avant le NGS

- Détection des mutations ponctuelles
 - Par Sanger
- Détection des variants structuraux
 - Par PCR
 - Par Southern Blot

Diagnostic des mitopathies

Avant le NGS

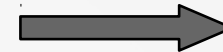
- Détection des mutations ponctuelles
 - Par Sanger
- Détection des variants structuraux
 - Par PCR
 - Par Southern Blot

Techniques peu sensibles
pour des faibles taux
d'hétéroplasmie (> 10%)
+
Difficultés dans la
détermination des positions
des CNVs
+
Techniques longues et
difficiles

Diagnostic des mitopathies

Avec le NGS


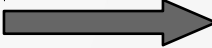
- Détection des mutations ponctuelles
 - Par des outils dit de « calling »
(recherche de variants versus une référence)



Bonne sensibilité
et spécificité

Diagnostic des mitopathies

Avec le NGS

- Détection des mutations ponctuelles
 - Par des outils dit de « calling »
(recherche de variants versus une référence) Bonne sensibilité et spécificité
- Détection des variants structuraux
 - Par des outils dédiés CNVs
(basés sur différentes méthodes) Moins performante (et technique dépendante)

Introduction : NGS & CHU d'Angers

- Technologie Life
 - Séquenceur Proton & S5-XL
 - Variation de pH
 - Librairie 'single-end'
- Pipeline « maison » pour mutations ponctuelles
 - NiourK
- **Manque un module de détection des CNVs sensible et spécifique**

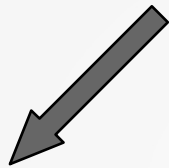


Problématique

**Développement d'un module de détection des CNVs
adapté à l'analyse du génome mitochondrial**

Test des outils de prédiction de CNVs existants

Au départ, on a voulu testé des outils adaptés à nos données Life (CANOES¹ & CoNIFER²)



Données expérimentales

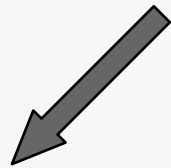
Aucune donnée pour des taux d'hétéroplasmie en dessous de 10 %

(1) Backenroth D, Homsy J, Murillo LR, Glessner J, Lin E, Brueckner M, Lifton R, Goldmuntz E, Chung WK, Shen Y, (2014) CANOES: Detecting rare copy number variants from whole exome sequencing data, Nucleic Acids Research, doi: 10.1093/nar/gku345

(2) Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, NHLBI Exome Sequencing Project, Quinlan AR, Nickerson DA and Eichler EE, (2012). Copy number variation detection and genotyping from exome sequence data. Genome Research, doi:10.1101/gr.138115.112

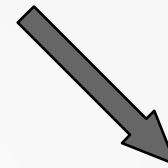
Test des outils de prédiction de CNVs existants

test des outils adaptés aux données Life (CANOES¹ & CoNIFER²)



Données expérimentales

Aucune donnée pour des taux d'hétéroplasmie en dessous de 10 %



Données simulées

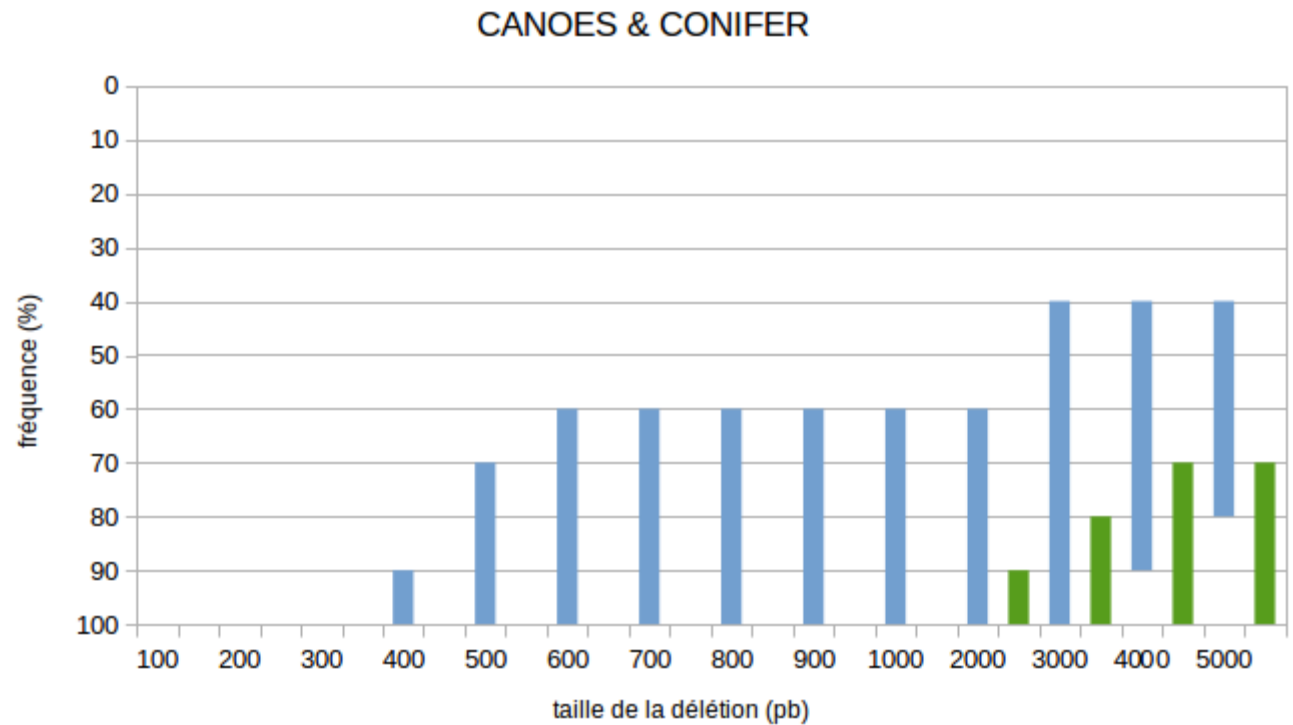
Permet un contrôle total sur les positions et les fréquences des CNVs



Tests sur les données simulées par KaSIMir

- Pas de détection
fréquence < 40 %
taille < 400pb

■ CANOES
■ CoNIFER



➔ Nécessité d'un outil bioinformatique plus performant

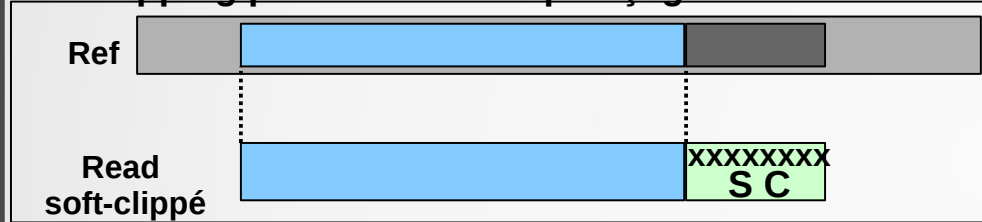
Partie I : le module



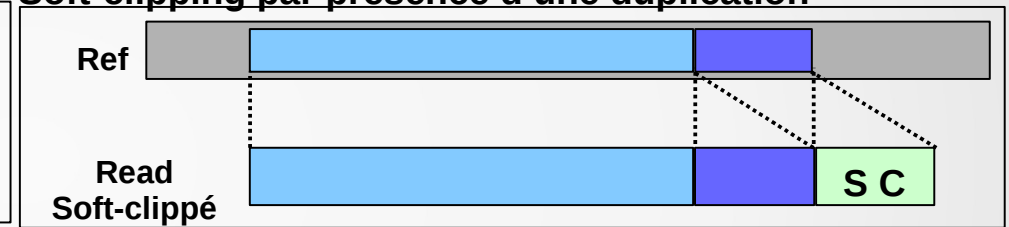
Partie I : Soft-clipping

- Soft-clipping : « rognage » des parties non alignées des reads

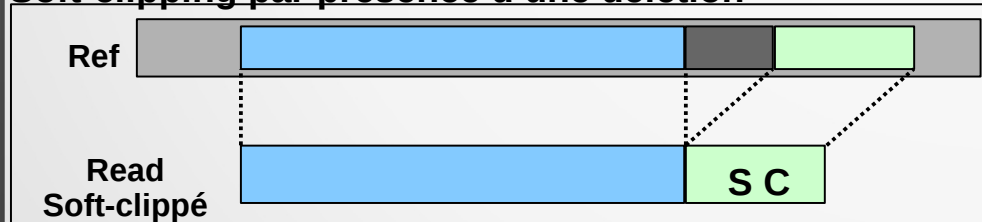
Soft-clipping par erreur de séquençage



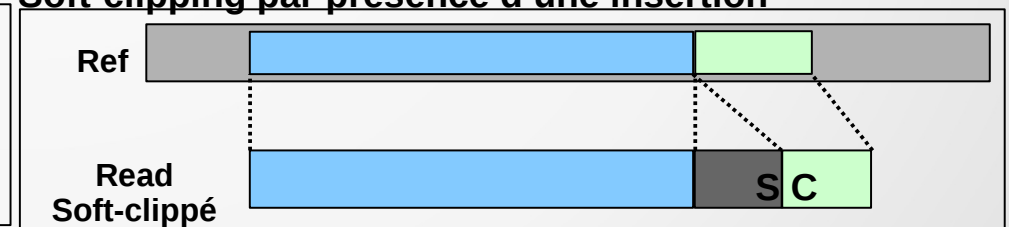
Soft-clipping par présence d'une duplication



Soft-clipping par présence d'une délétion



Soft-clipping par présence d'une insertion



- Lien entre CNVs et Soft-clipping

Partie I : Soft-clipping

- Soft-clipping : « rognage » des parties non alignées des reads

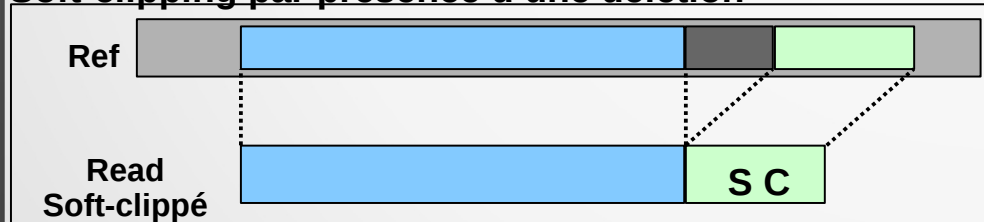
Soft-clipping par erreur de séquençage



Soft-clipping par présence d'une duplication



Soft-clipping par présence d'une délétion



Soft-clipping par présence d'une insertion



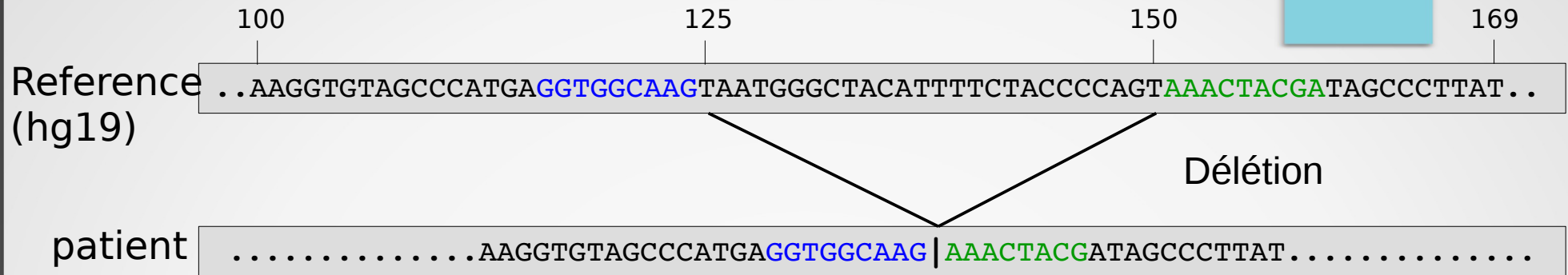
- Lien entre CNVs et Soft-clipping

Partie I : Déroulement d'eKLIPse

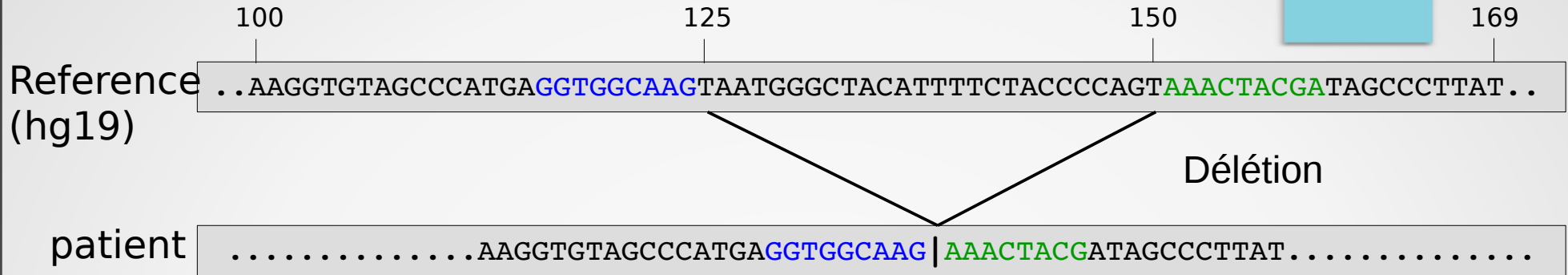
Reference (hg19) ..AAGGTGTAGCCCATGAGGTGGCAAGTAATGGGCTACATTTTCTACCCAGTAAACTACGATAGCCCTTAT..

100 125 150 169

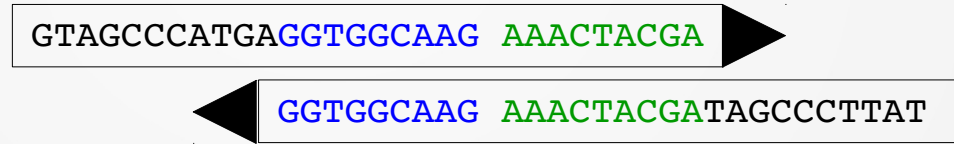
Partie I : Déroulement d'eKLIPse



Partie I : Déroulement d'eKLIPse



↓ Séquencage



Partie I : Déroulement d'eKLIPse

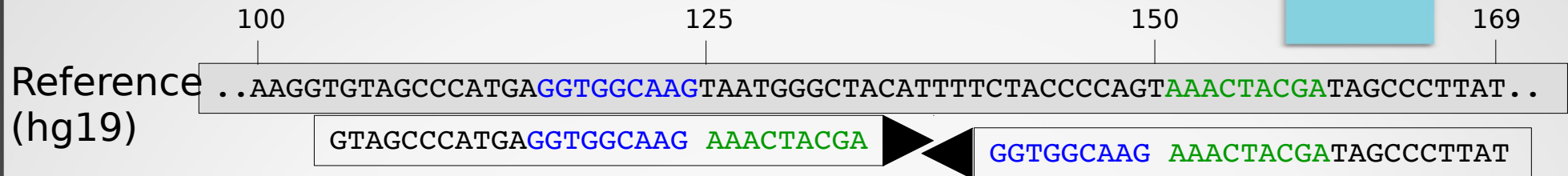
Reference (hg19) ..AAGGTGTAGCCCATGAGGTGGCAAGTAATGGGCTACATTTTCTACCCAGTAAACTACGATAGCCCTTAT..

100 125 150 169

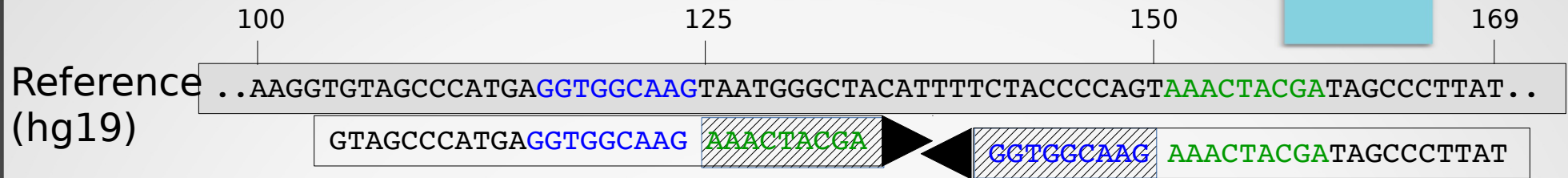
GTAGCCCATGAGGTGGCAAG AACTACGA

GGTGGCAAG AACTACGATAGCCCTTAT

Partie I : Déroulement d'eKLIPse

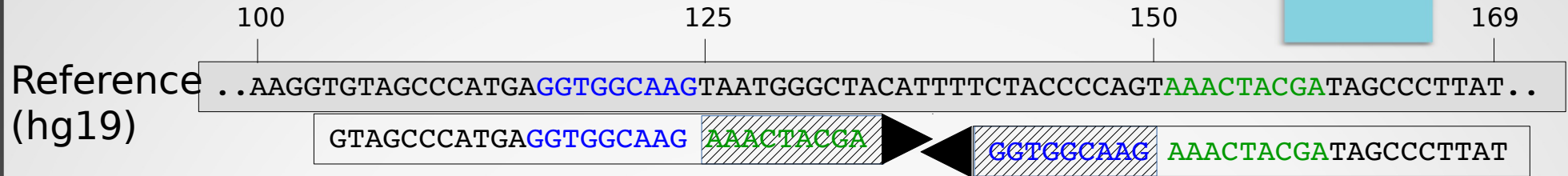


Partie I : Déroulement d'eKLIPse



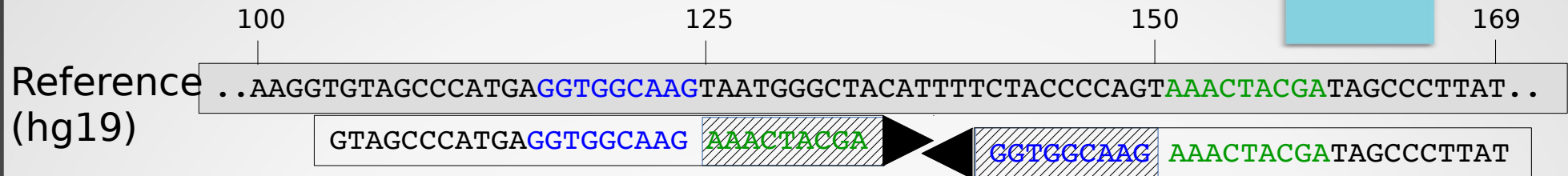
Soft-clipping

Partie I : Déroulement d'eKLISe

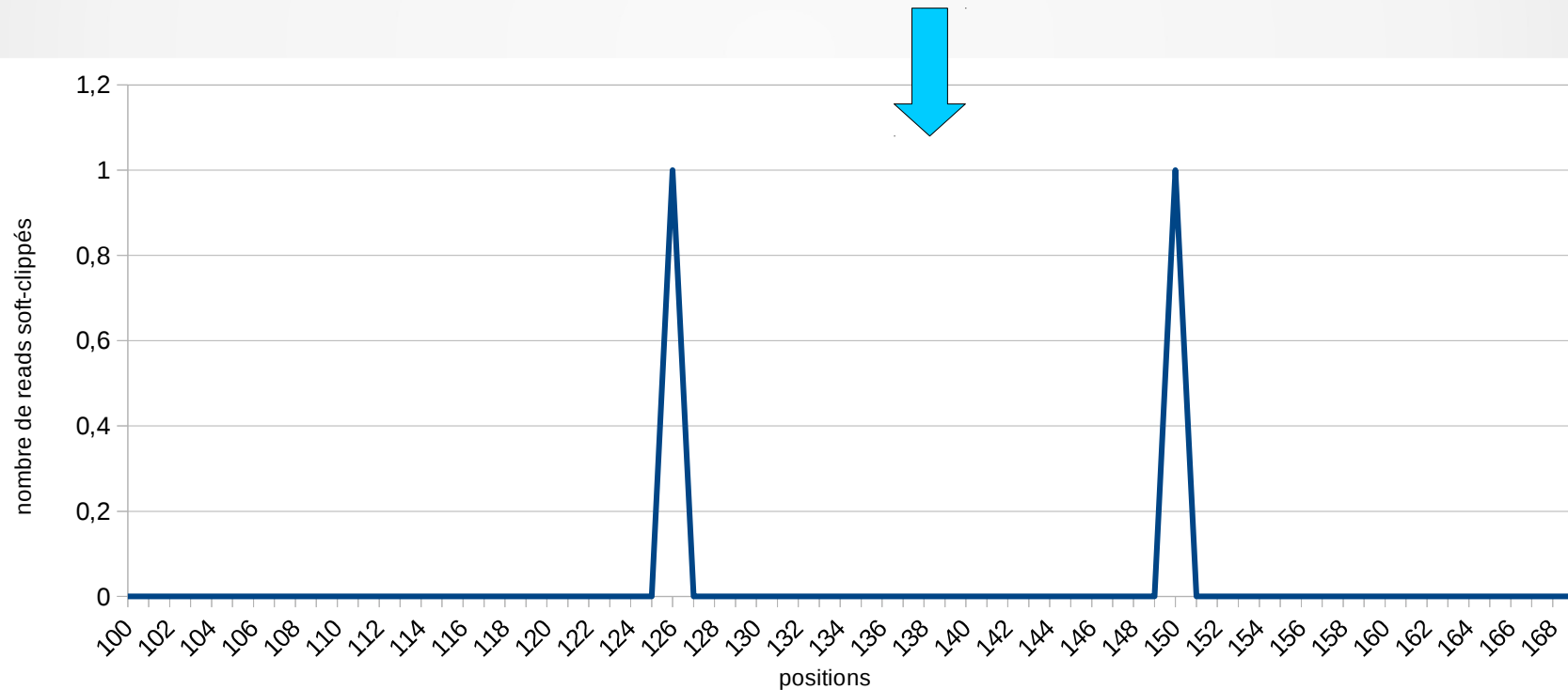


- 1 Distribution du soft-clipping

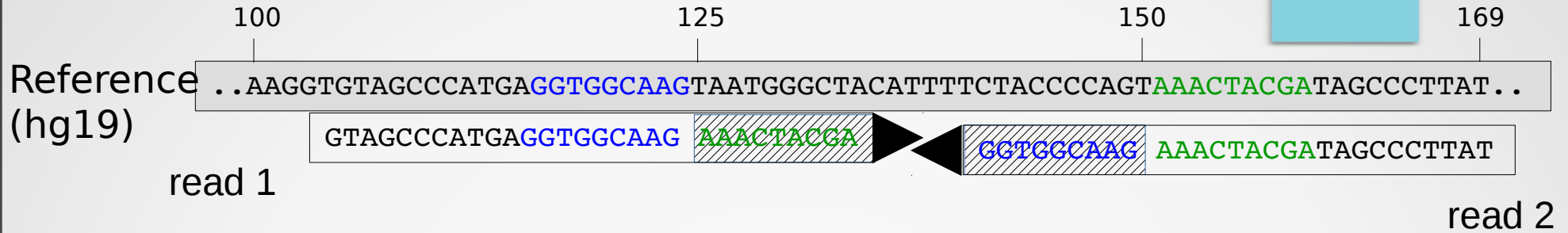
Partie I : Déroulement d'eKLIPse



1 Distribution du soft-clipping



Partie I : Déroulement d'eKLIPse



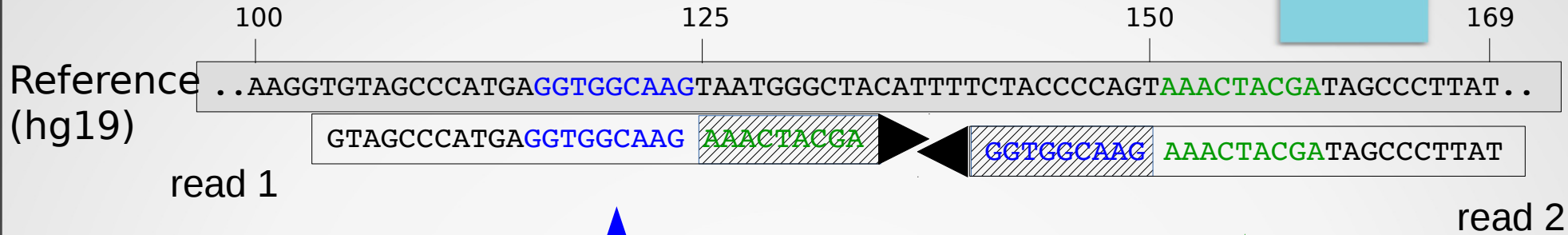
- 2 Récupération des séquences soft-clippées (≥ 25 pb)



```
>read1_scR  
AAACTACGA  
  
>read2_scL  
GGTGGTAAG
```

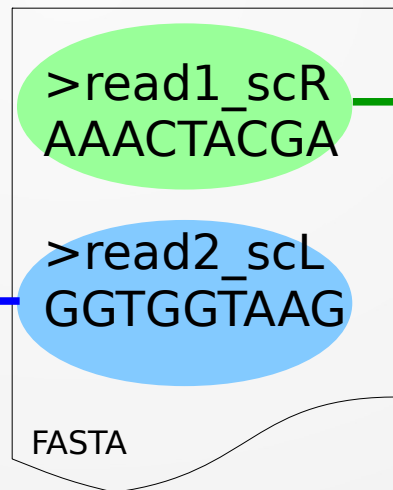
FASTA

Partie I : Déroulement d'eKLIPse



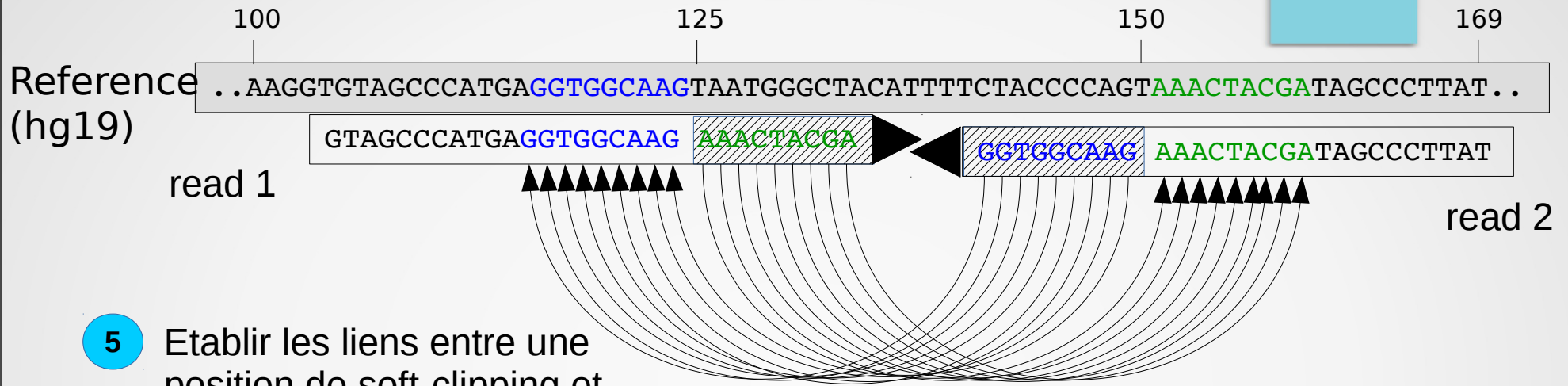
3 BLAT de ces séquences soft-clippées.

id = 100 %
cov =
100 %



id = 100 %
cov =
100 %

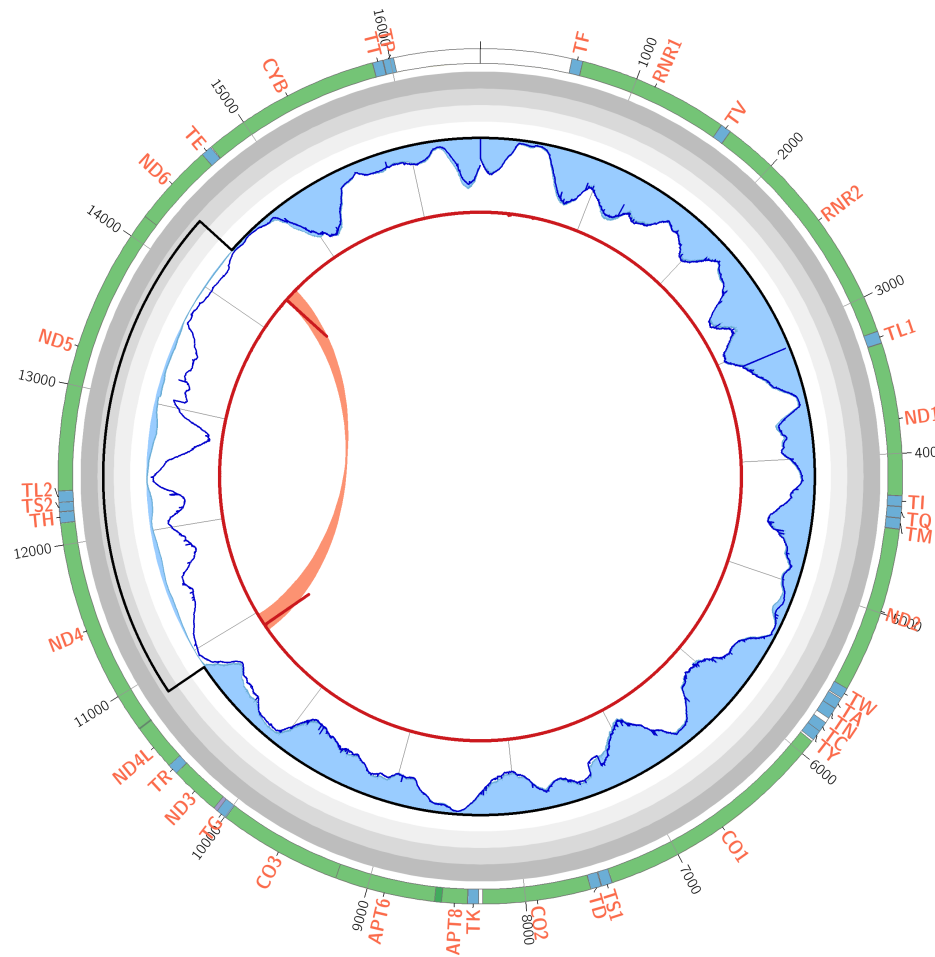
Partie I : Déroulement d'eKLIPse



- 5 Etablir les liens entre une position de soft-clipping et la position du BLAT de sa séquence

➔ Au final, eKLIPse dispose des pics de soft-clipping et les liens BLAT associés

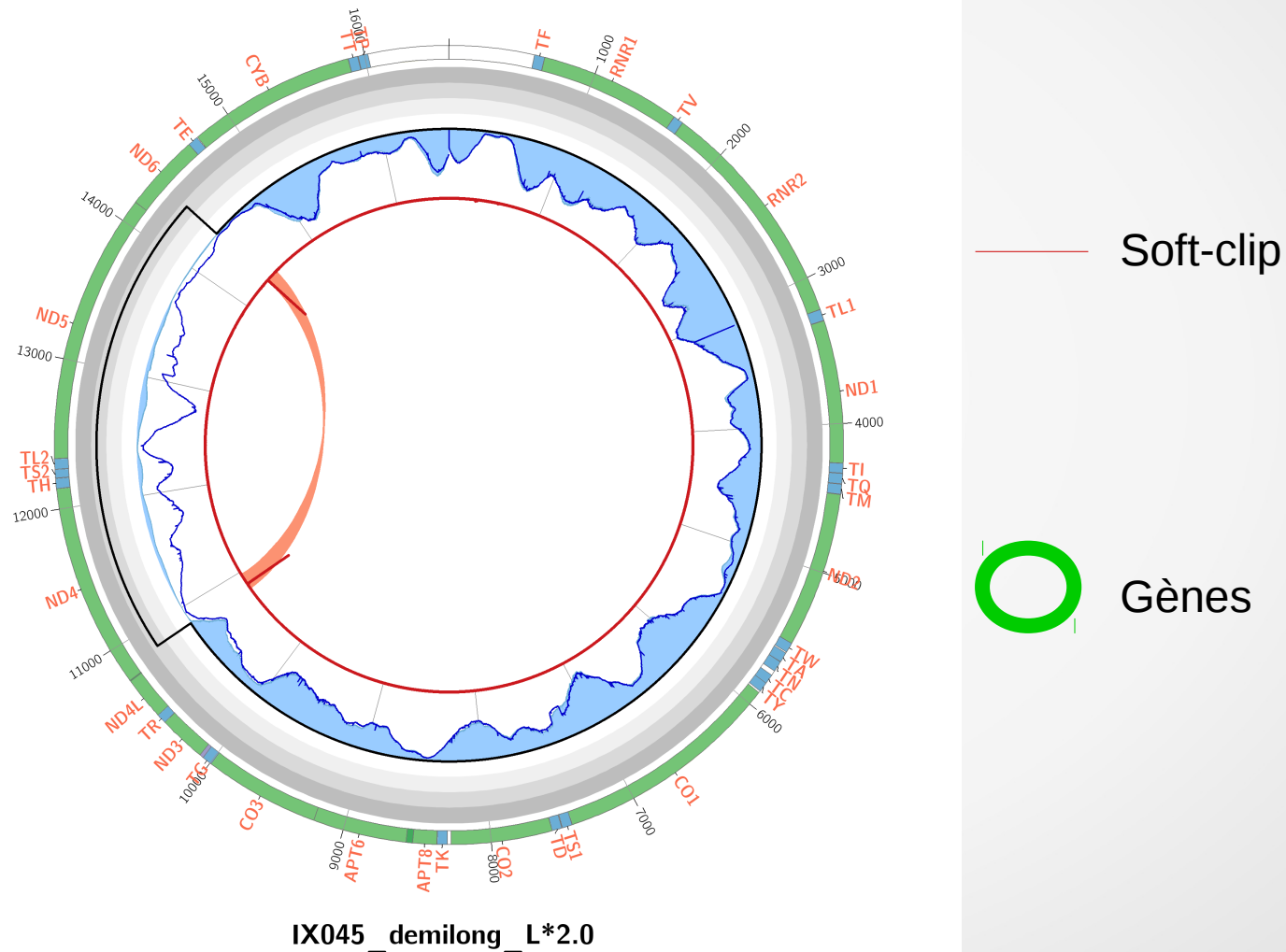
Partie I : Représentation Circos³



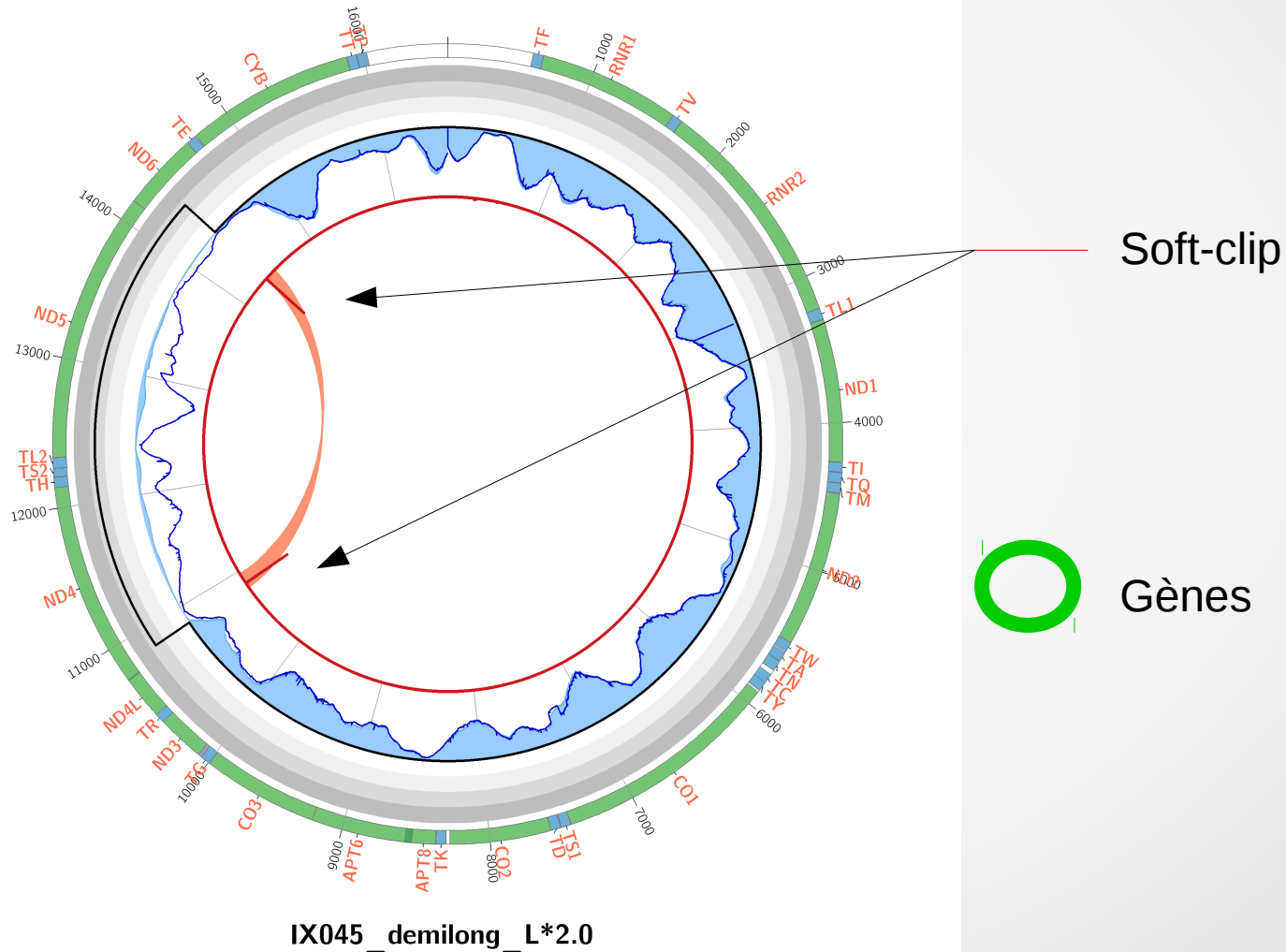
IX045_demilong_L*2.0

 Gènes

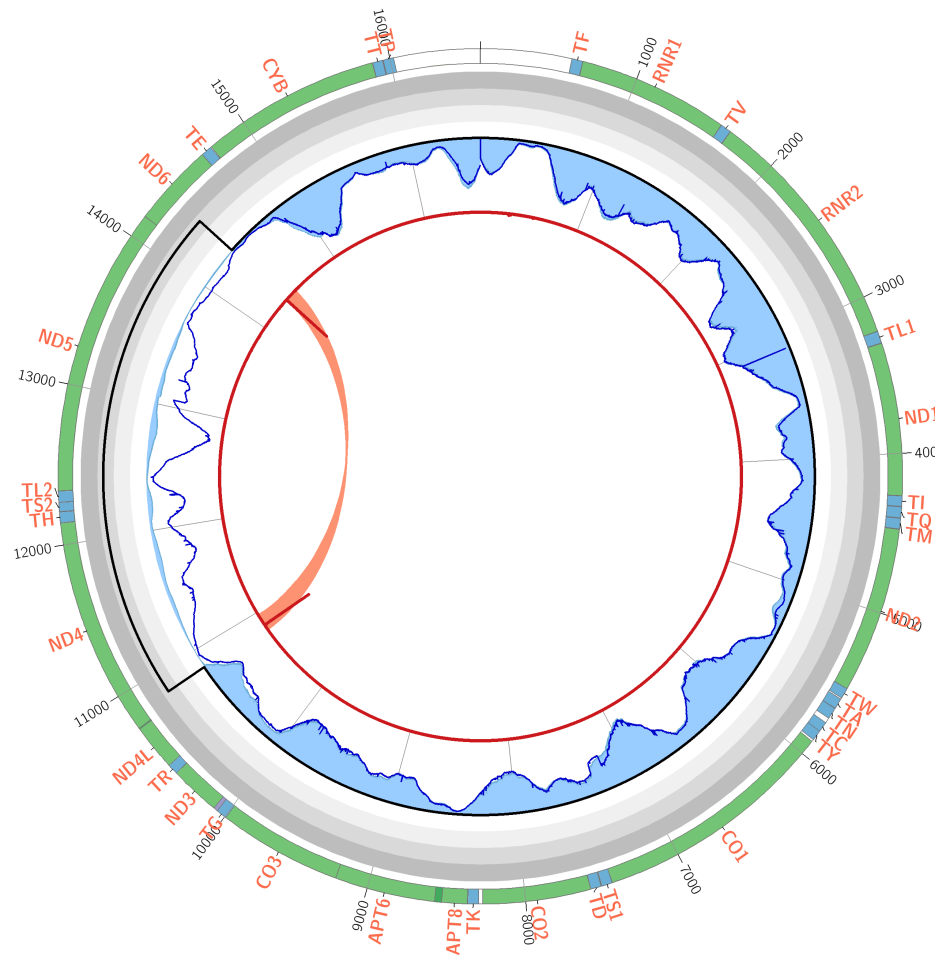
Partie I : Représentation Circos³



Partie I : Représentation Circos³



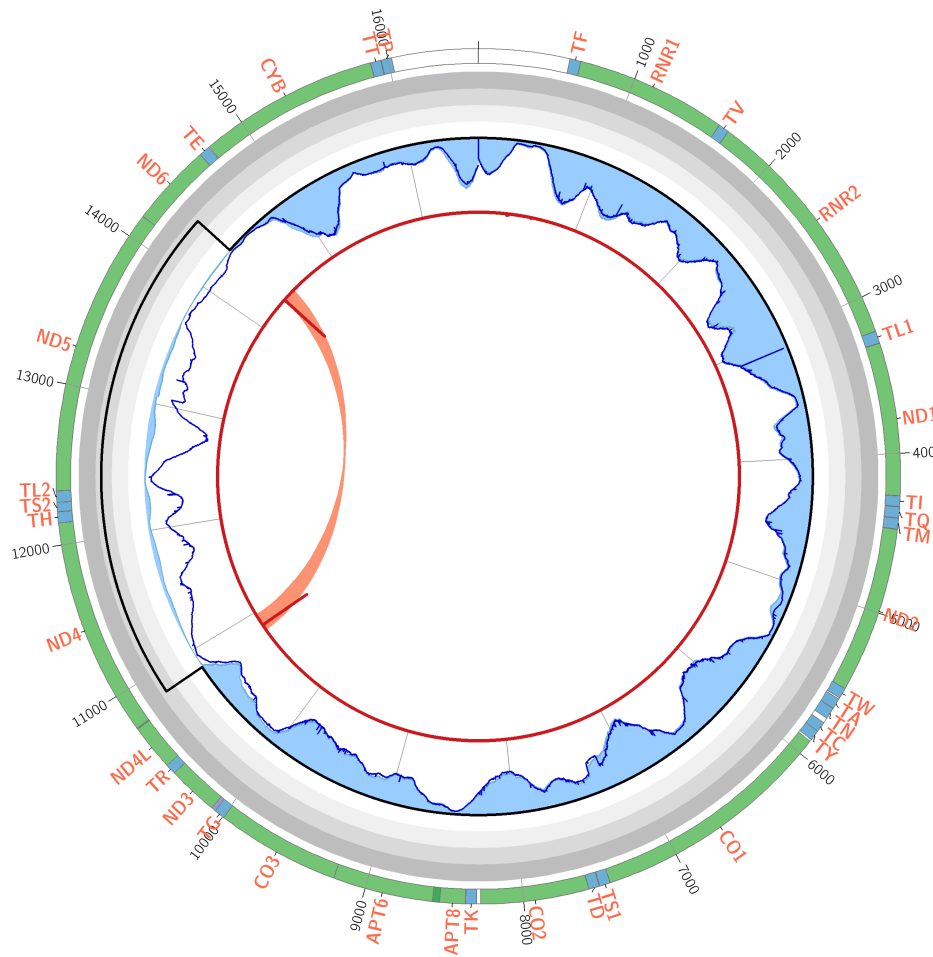
Partie I : Représentation Circos³



IX045_demilong_L*2.0



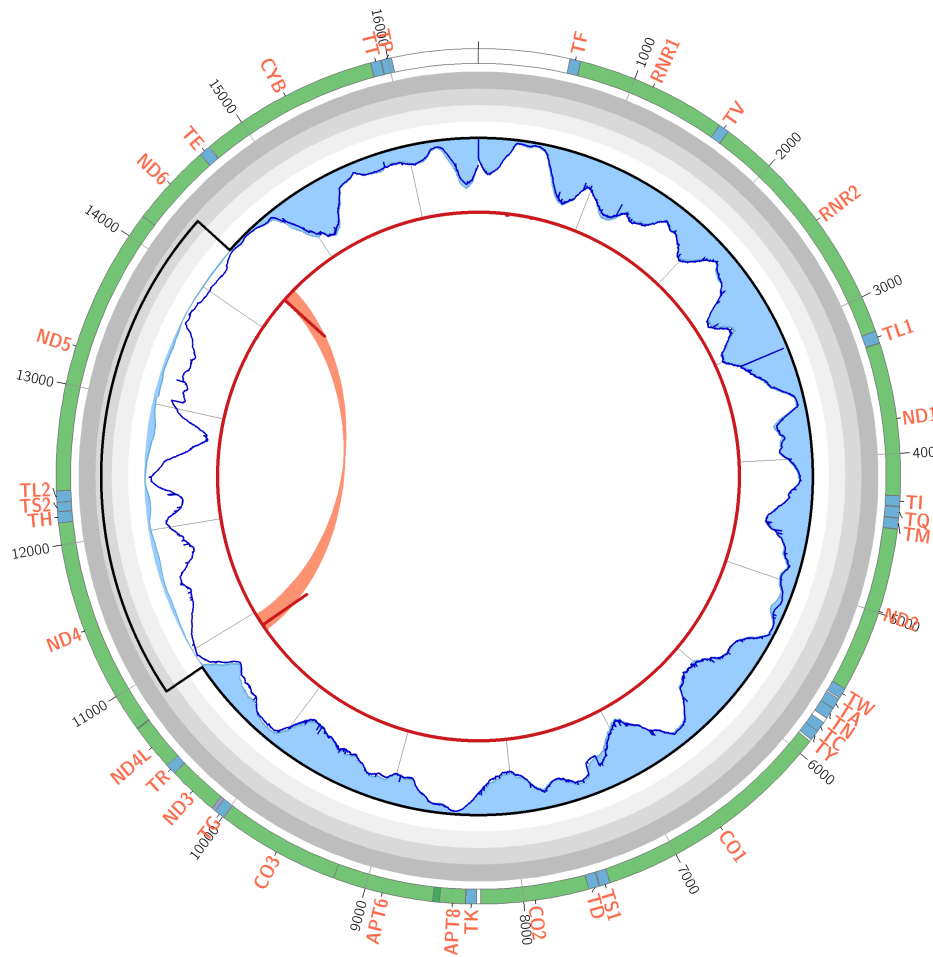
Partie I : Représentation Circos³



IX045_demilong_L*2.0

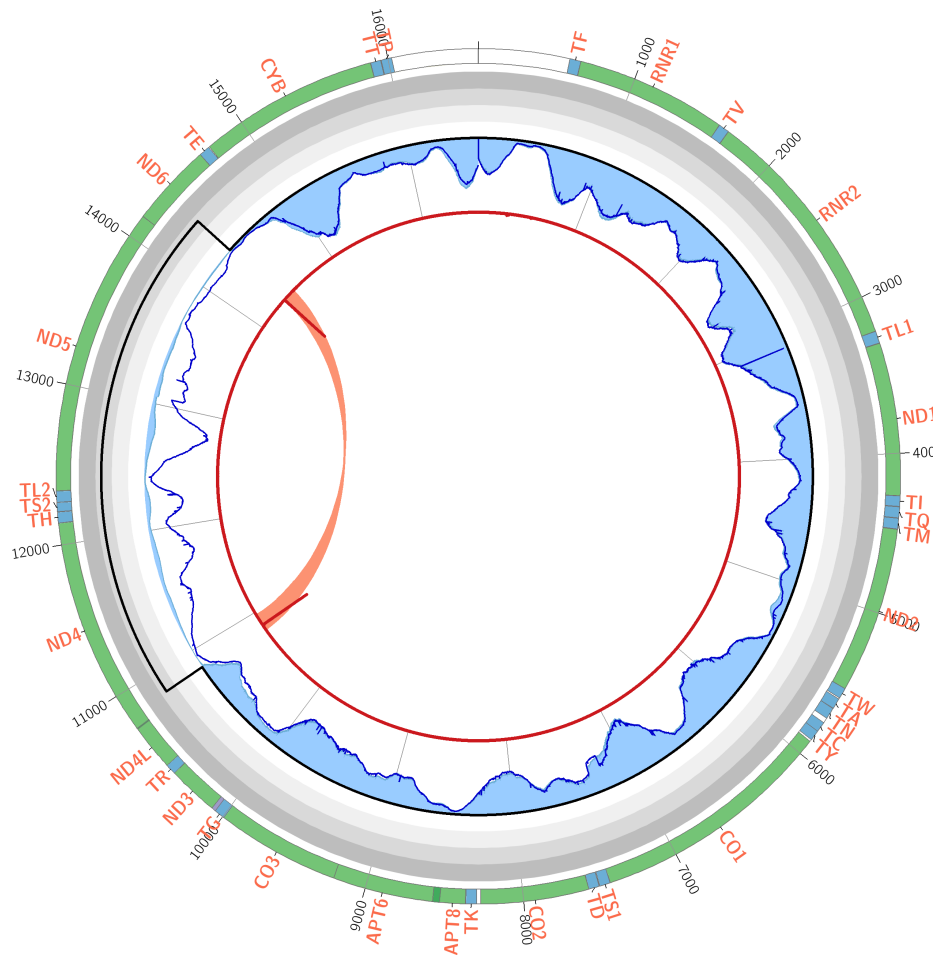
- Soft-clip
- Liens « blat »
- Médiane couverture
- Gènes

Partie I : Représentation Circos³



IX045_demilong_L*2.0

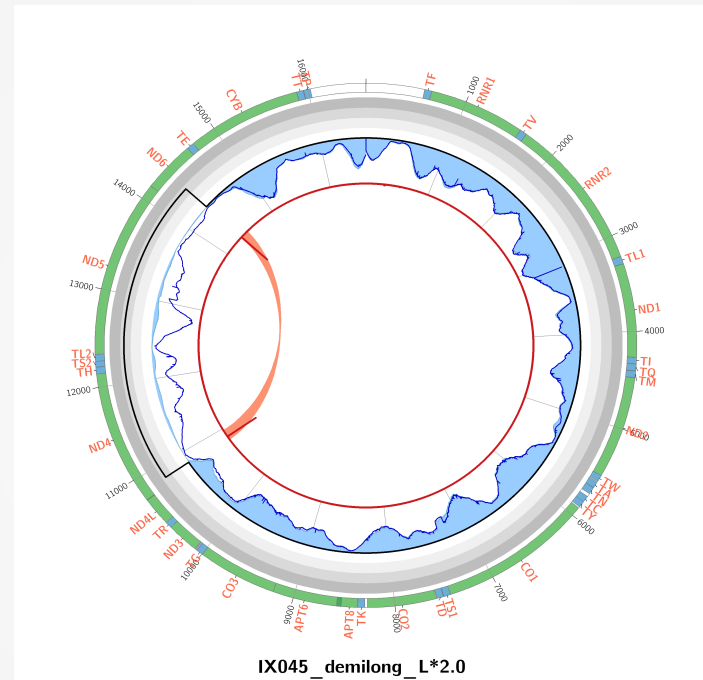
Partie I : Représentation Circos³



IX045_demilong_L*2.0

- Soft-clip
- Liens « blat »
- Médiante couverture
- Couverture patient
- Fréquence cumulée
- Gènes

Partie I : Résultats eKLIPse





IonXpress_045										
Region	Pos	Blat Interval				Nb blat	Freq peak	Region of possible del		
		Start_sc	End_sc	Start_ali	End_ali			Start	End	Freq del
Peak_0	10851	10843	10988	14357	14504	443	65,48%	10851	14355	65,87%
Peak_1	14355	14355	14405	10793	10842	48	66,27%	10851	14355	65,87%

Partie I : Test et validation d'eKLIPse

- Test sur
 - Délétion commune de 4kb



Partie I : Test et validation d'eKLIPse

- Test sur
 - Délétion commune de 4kb 
 - Un jeu de données simulées par KaSIMir 
 - Localisation précise des délétions
 - Bonne estimation du taux d'hétéroplasmie
 - Marge d'erreur sur ce taux à $\pm 5 \%$



Partie II : Cas pratique

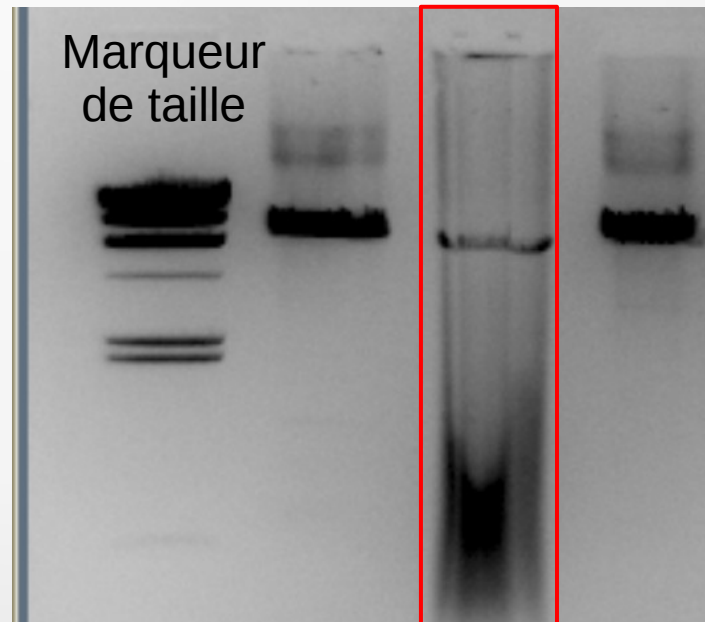
Partie II : Cas pratique

- Madame P. suivi au CHU de Brest
- Maladie Ophtalmologique
- Origine mitochondriale

Partie II : Cas pratique

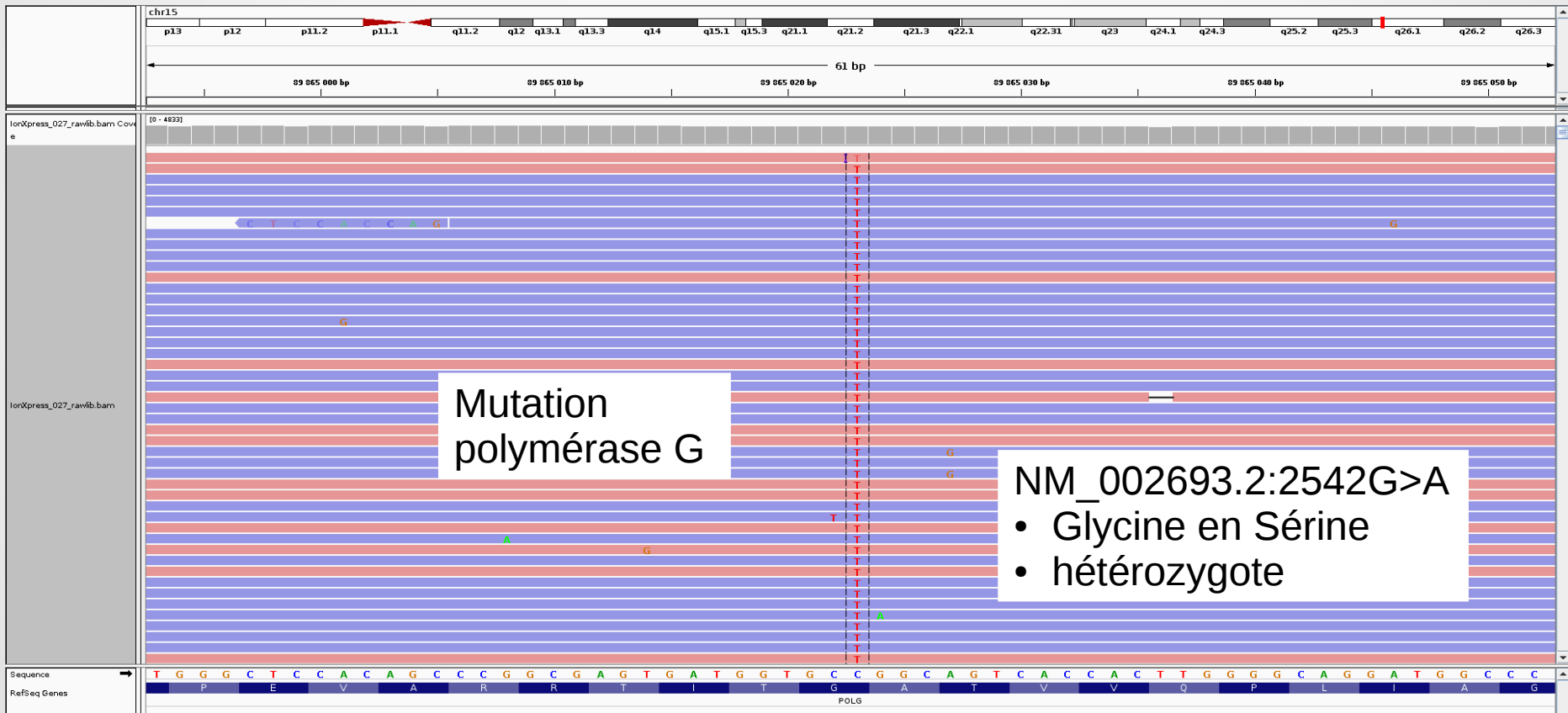
1 1^{er} séquençage NGS sur l'ADN mitochondrial

- ➔ Aucune mutation ponctuelle observée
- ➔ Profil de migration PCR caractéristique de délétions multiples



Partie II : Cas pratique

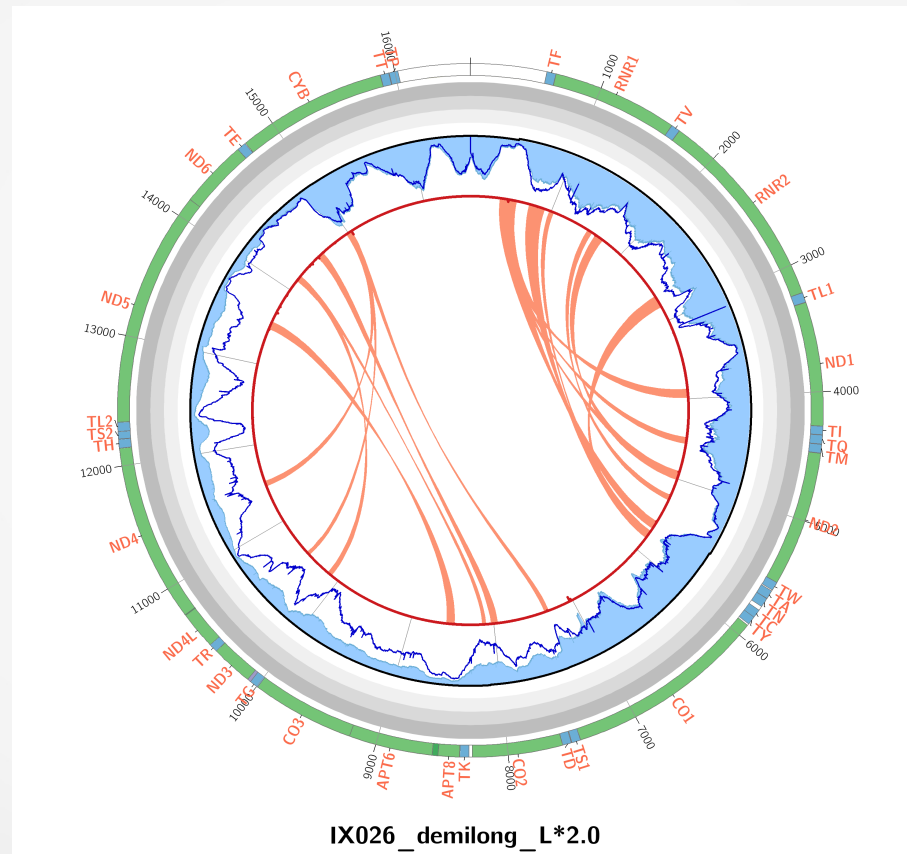
2 2^{ème} séquençage NGS sur gènes nucléaires



➔ Mutation reportée comme impliquée dans l'accumulation des délétions multiples de l'ADNmt.

Partie II : Cas pratique

2 Résultats d'eKLIPse sur le 1^{er} séquençage (ADNmt)



Partie II : Cas pratique

2 Résultats d'eKLIPse sur le 1^{er} séquençage (ADNmt)

IonXpress_026								Region of possible del		
Region	Pos	Blat Interval				Nb blat	Freq peak	Start	End	Freq del
		Start_sc	End_sc	Start_ali	End_ali					
Region_6	462	462	573	3880	3988	55	3,66%	462	3879	2,15%
Region_1	4476	4476	4552	386	460	16	0,63%	462	4476	2,14%
Region_0	5582	5582	5681	362	460	21	0,86%	462	5582	2,26%
Region_9	462	460	555	5584	5678	28	3,66%	462	5582	2,26%
Region_2	3879	3879	3950	394	463	14	0,64%	535	3879	0,41%
Region_3	4912	4912	4991	394	471	16	1,25%	535	4912	0,71%
Region_12	823	823	930	5723	5825	45	0,55%	823	5721	0,44%
Region_4	5721	5721	5820	726	823	12	0,33%	914	5721	0,18%
Region_11	980	980	1079	5723	5822	9	0,17%	980	5721	0,25%
Region_8	1499	1499	1587	5217	5304	21	0,21%	1499	5206	0,40%
Region_7	1613	1613	1718	4905	5007	67	0,90%	1613	4904	0,94%
Region_10	2740	2740	2828	5730	5817	10	0,13%	2740	5721	0,23%
Region_5	5728	5728	5807	2663	2741	6	0,09%	2807	5728	0,07%
Region_18	7333	7320	7370	14995	15044	19	1,11%	7333	14973	0,57%
Region_13	14529	14529	14611	7955	8034	13	2,68%	8092	14529	1,43%
Region_16	8100	8100	8181	14171	14251	6	0,10%	8100	14150	0,36%
Region_14	8481	8481	8580	13460	13558	10	0,73%	8481	13446	0,91%
Region_15	10110	10110	10187	14142	14218	9	0,35%	10110	14132	0,42%
Region_19	10477	10477	10519	15003	15044	11	0,25%	10477	15002	0,14%
Region_17	11478	11478	11548	14974	15043	6	0,34%	11478	14973	0,19%

Partie II : Cas pratique

2 Résultats d'eKLIPse sur le 1^{er} séquençage (ADNmt)

IonXpress_026								Region of possible del		
Region	Pos	Blat Interval				Nb blat	Freq peak	Start	End	Freq del
		Start_sc	End_sc	Start_ali	End_ali					
Region_6	462	462	573	3880	3988	55	3,66%	462	3879	2,15%
Region_1	4476	4476	4552	386	460	16	0,63%	462	4476	2,14%
Region_0	5582	5582	5681	362	460	21	0,86%	462	5582	2,26%
Region_9	462	460	555	5584	5678	28	3,66%	462	5582	2,26%
Region_2	3879	3879	3950	394	463	14	0,64%	535	3879	0,41%
Region_3	4912	4912	4991	394	471	16	1,25%	535	4912	0,71%
Region_12	823	823	930	5723	5825	45	0,55%	823	5721	0,44%
Region_4	5721	5721	5820	726	823	12	0,33%	914	5721	0,18%
Region_11	980	980	1079	5723	5822	9	0,17%	980	5721	0,25%
Region_8	1499	1499	1587	5217	5304	21	0,21%	1499	5206	0,40%
Region_7	1613	1613	1718	4905	5007	67	0,90%	1613	4904	0,94%
Region_10	2740	2740	2828	5730	5817	10	0,13%	2740	5721	0,23%
Region_5	5728	5728	5807	2663	2741	6	0,09%	2807	5728	0,07%
Region_18	7333	7320	7370	14995	15044	19	1,11%	7333	14973	0,57%
Region_13	14529	14529	14611	7955	8034	13	2,68%	8092	14529	1,43%
Region_16	8100	8100	8181	14171	14251	6	0,10%	8100	14150	0,36%
Region_14	8481	8481	8580	13460	13558	10	0,73%	8481	13446	0,91%
Region_15	10110	10110	10187	14142	14218	9	0,35%	10110	14132	0,42%
Region_19	10477	10477	10519	15003	15044	11	0,25%	10477	15002	0,14%
Region_17	11478	11478	11548	14974	15043	6	0,34%	11478	14973	0,19%

➡ Prédiction de délétions multiples par eKLIPse

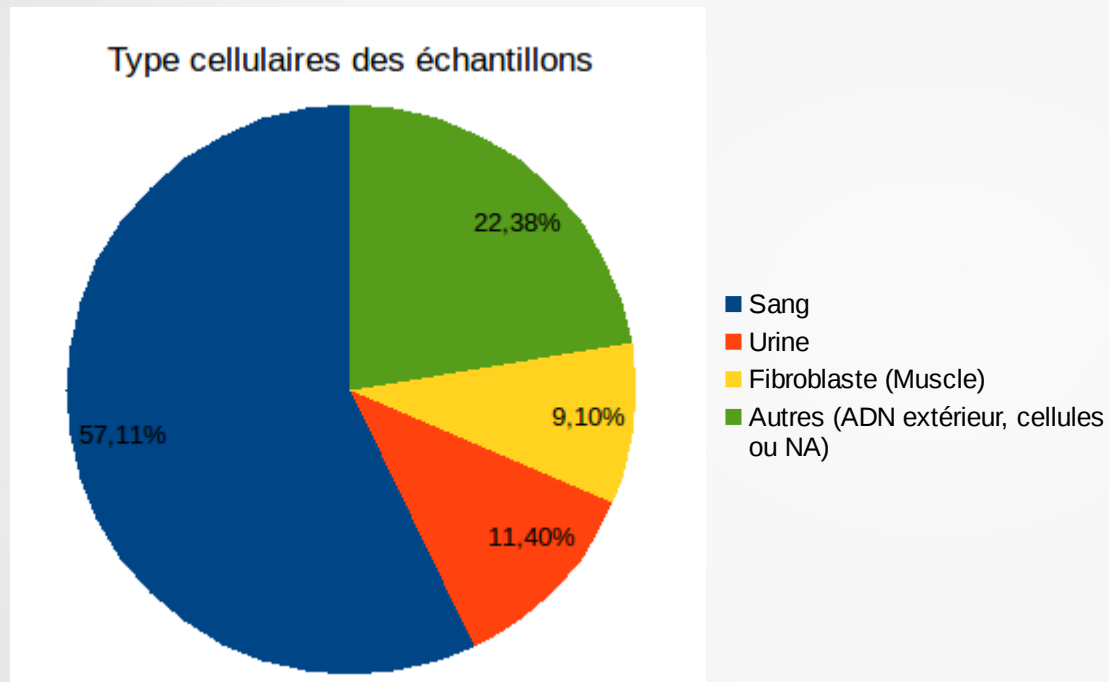
➡ Corrélation entre hétéroplasmie et types cellulaires ?



Partie III : Importance du type cellulaire sur les délétions

Partie III : les échantillons

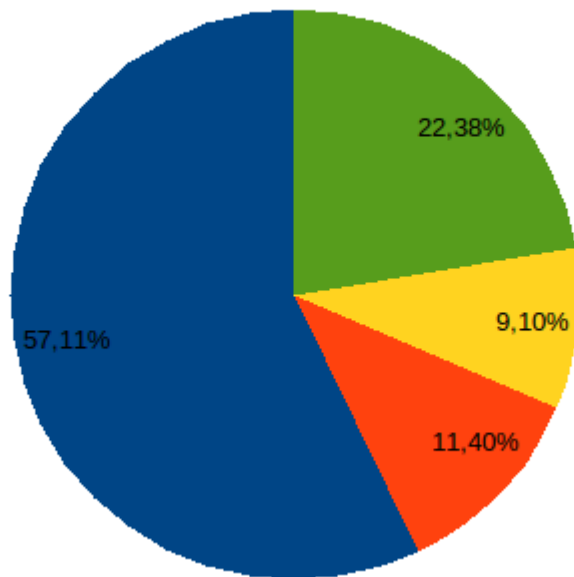
- Analyse de 58 runs soit 955 patients



Partie III : les échantillons

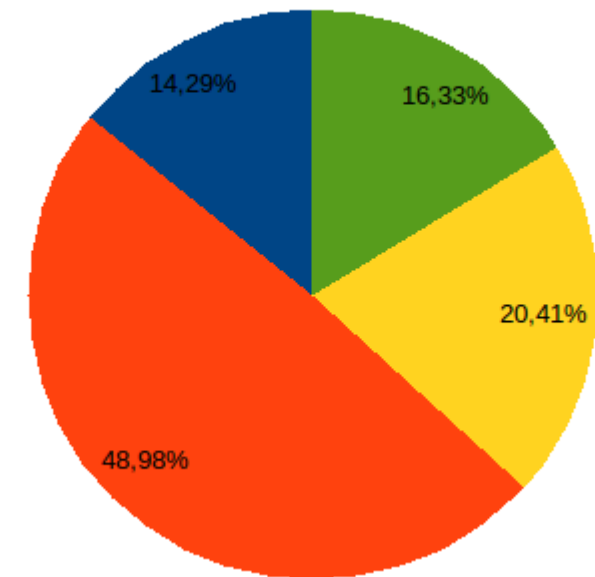
- Analyse de 58 runs soit 955 patients

Type cellulaires des échantillons



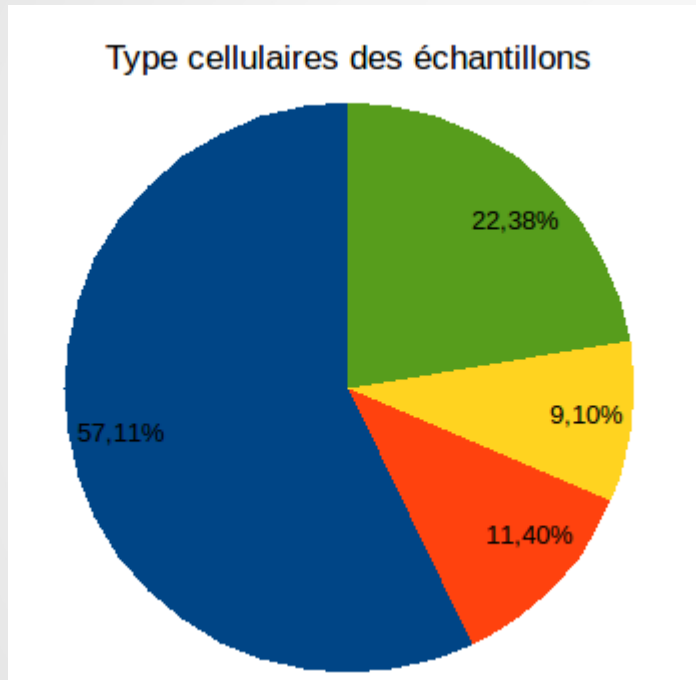
■ Sang
■ Urine
■ Fibroblaste (Muscle)
■ Autres (ADN extérieur, cellules ou NA)

Délétions détectées par eKLIPse

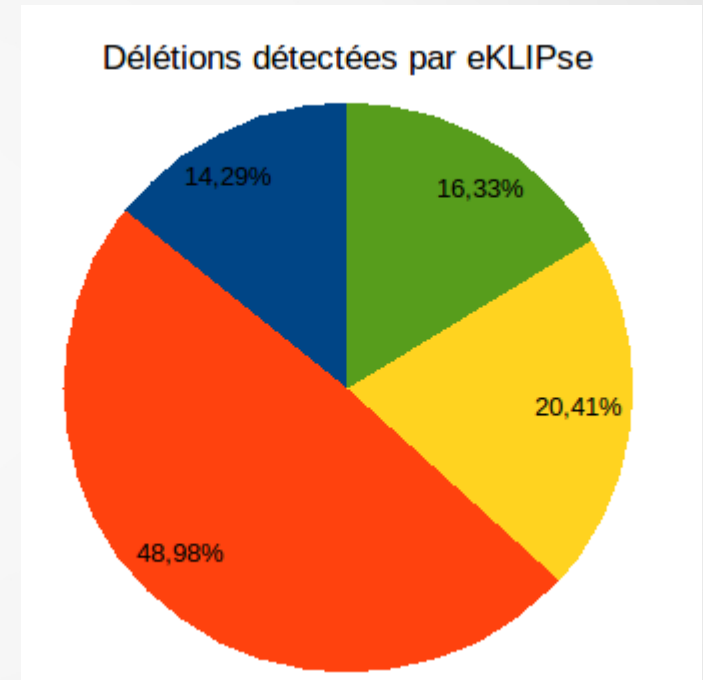


Partie III : les échantillons

- Analyse de 58 runs soit 955 patients



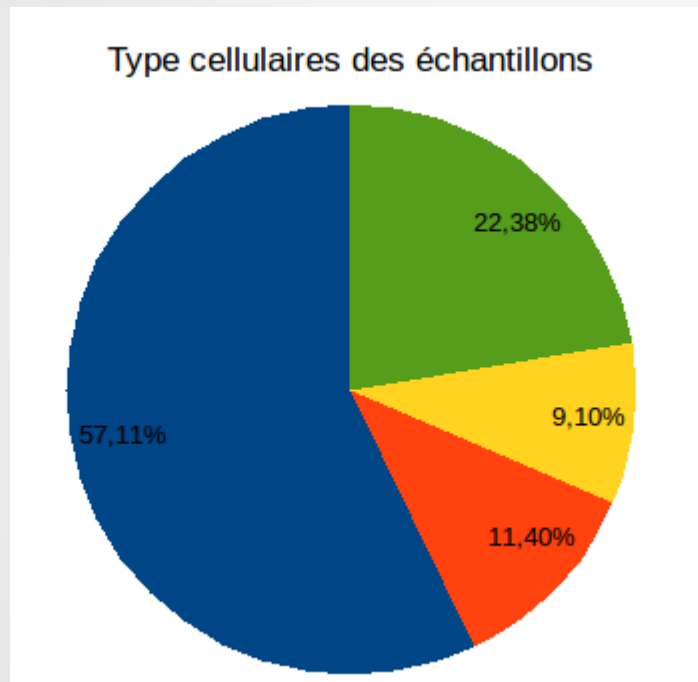
■ Sang
■ Urine
■ Fibroblaste (Muscle)
■ Autres (ADN extérieur, cellules ou NA)



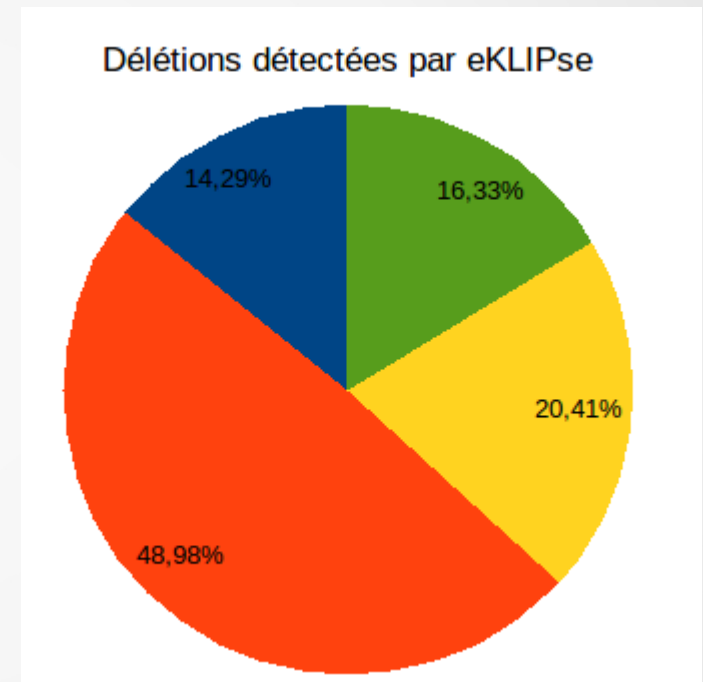
- Sang > fort taux de renouvellement
- Urine/Muscle > faible taux de renouvellement

Partie III : les échantillons

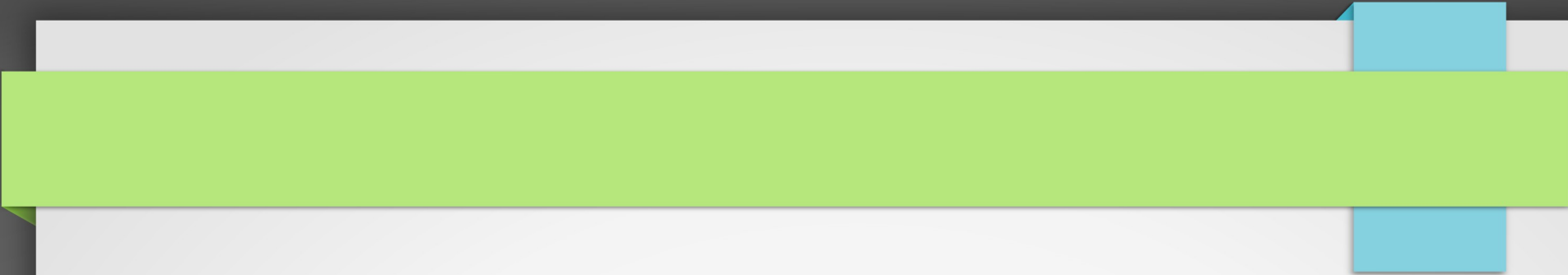
- Analyse de 58 runs soit 955 patients



■ Sang
■ Urine
■ Fibroblaste (Muscle)
■ Autres (ADN extérieur, cellules ou NA)

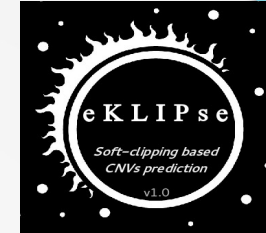


- Favoriser le type urinaire pour la détection de délétion
- Moins invasif et possibilité de suivi dans le temps

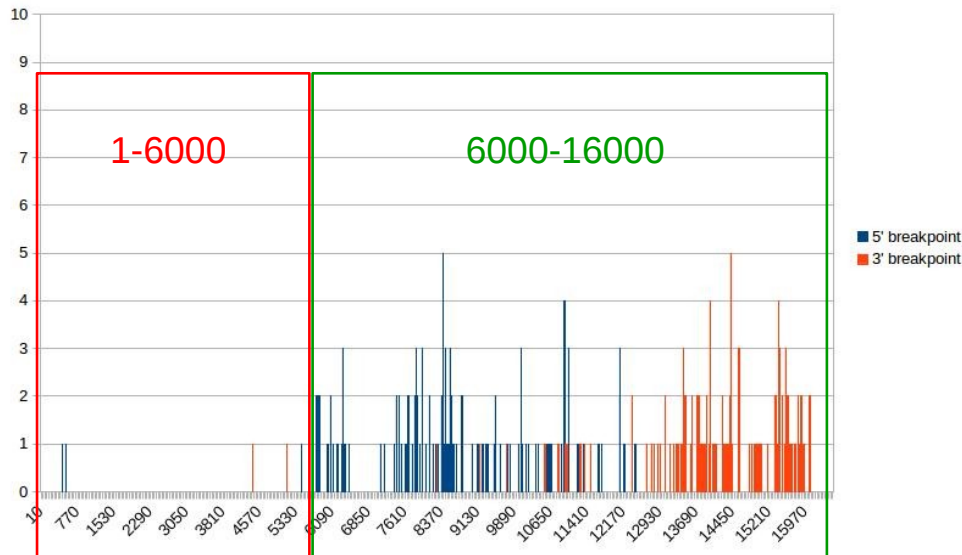


Partie IV : Analyse des positions des délétions

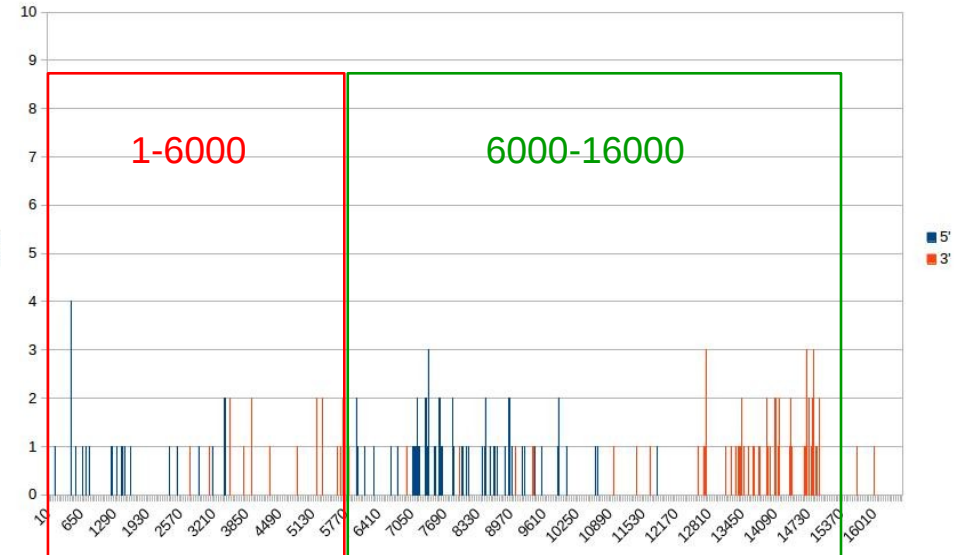
Partie IV : Répartition des délétions



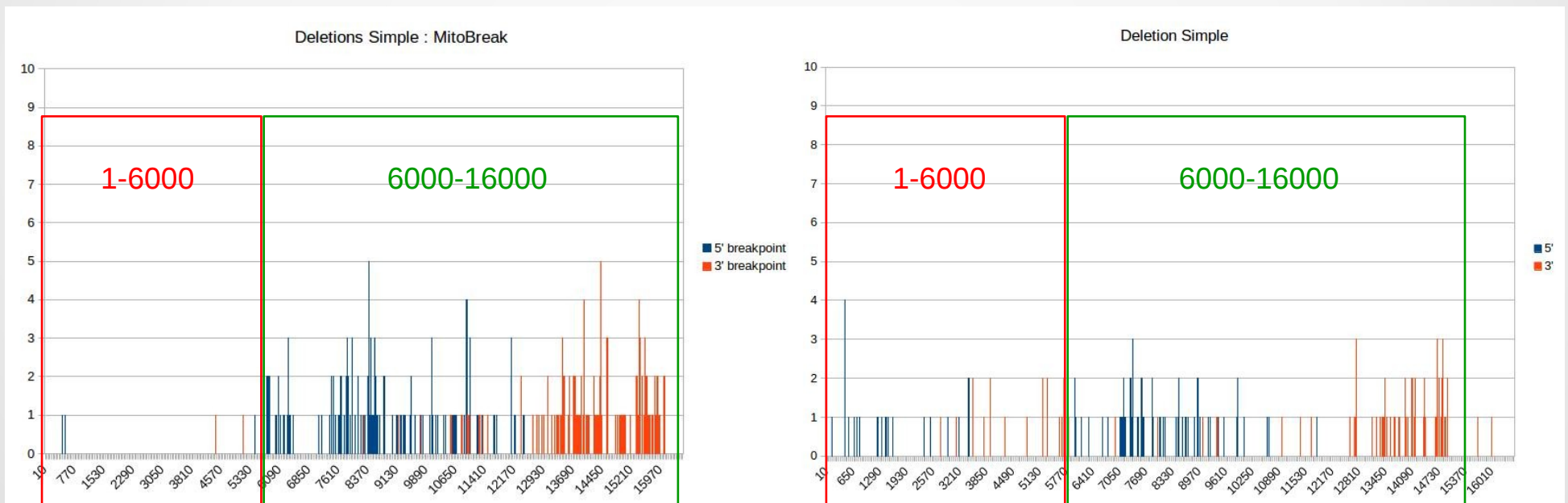
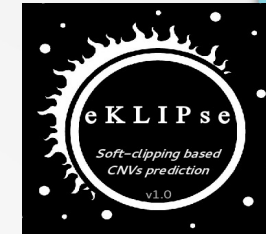
Deletions Simple : MitoBreak



Deletion Simple

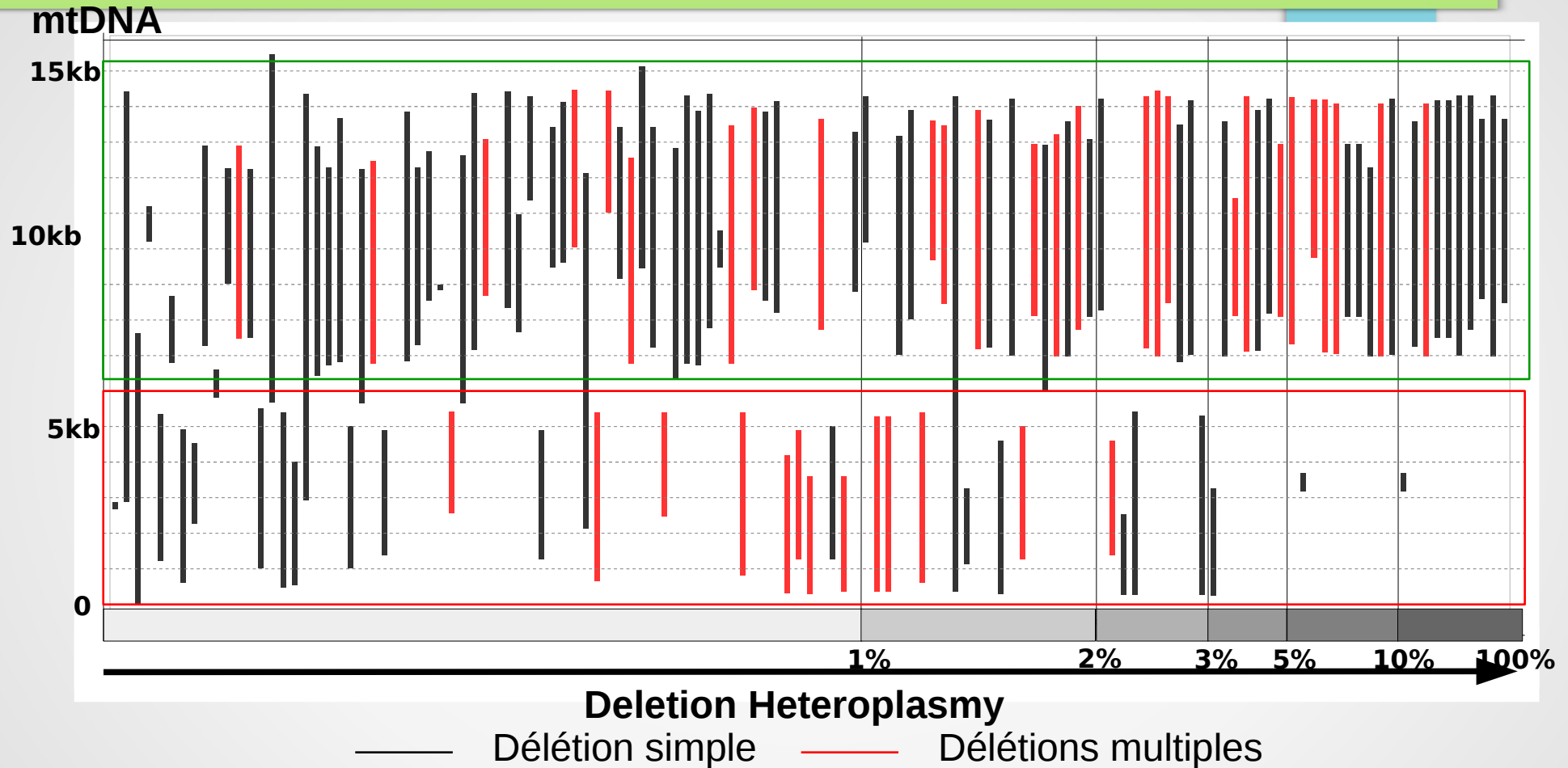


Partie IV : Répartition des délétions



- eKLIPse détecte des taux d'hétéroplasmie PLUS faibles

Partie IV : Répartition des délétions



- Confortation dans nos prédictions > même profile entre 6000 et 1600
- Mais plus sensible > présence de délétion à faible hétéroplasmie entre 1 et 6000



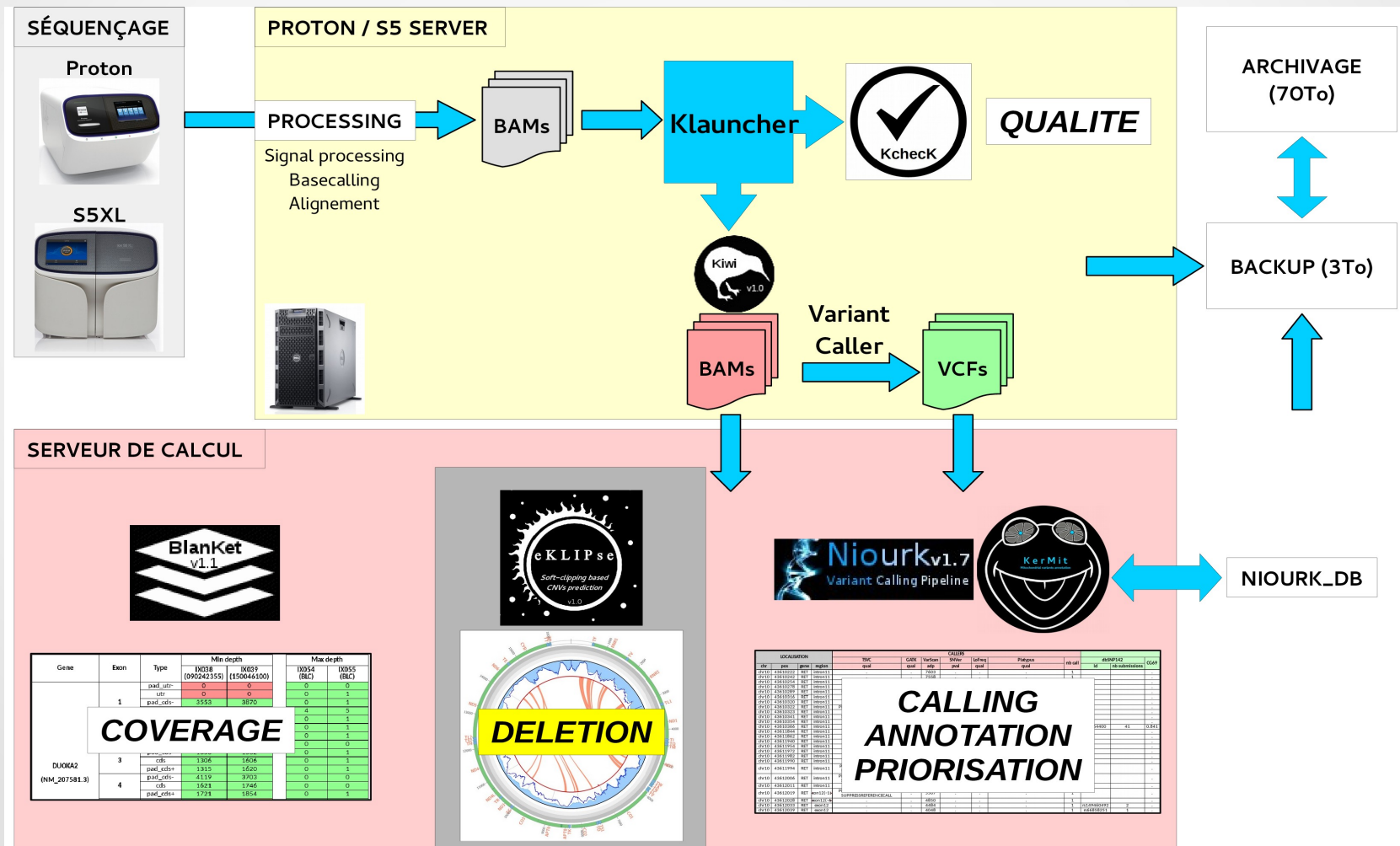
Conclusion & Perspectives

Conclusion

- Développement de KaSIMir permettant un contrôle total lors de la simulation de CNVs
- Développement d'eKLIPSe améliorant le diagnostic clinique NGS
 - Soft-clipping
 - « Délétion simple et multiple »
 - Localisation précise des positions de délétion
 - Prédiction de délétion à des taux d'hétéroplasmie très faibles (<1%)
 - Marge d'erreur sur les fréquences de $\pm 5 \%$
- Analyse de 955 patients dont 52 portant des délétions
 - eKLIPse a permis d'améliorer le diagnostique pour au moins XXX patients où les délétions n'avaient pas été détectées par l'analyse classique.

Conclusion

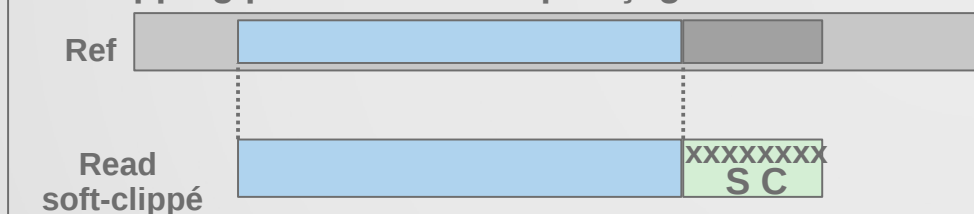
- EKLIPse intégré au pipeline de diagnostic NGS



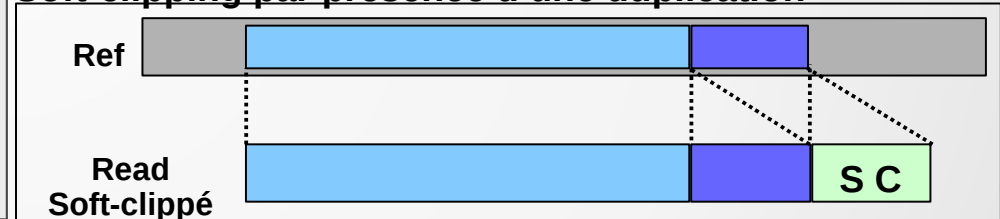
Perspectives

- Amélioration de KaSIMir
- Extension d'eKLIPSe aux 2 autres types de CNVs (insertion et duplication)

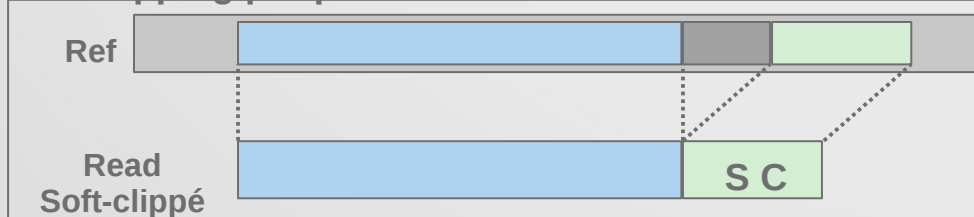
Soft-clipping par erreur de séquençage



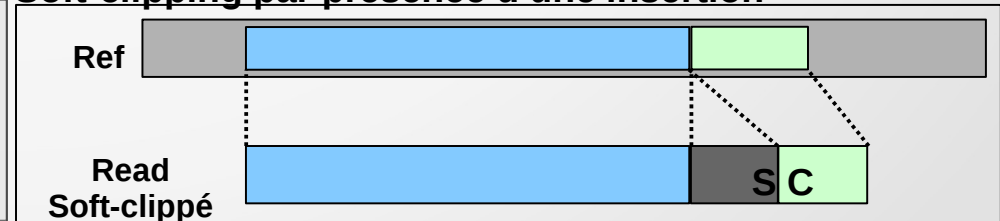
Soft-clipping par présence d'une duplication



Soft-clipping par présence d'une délétion



Soft-clipping par présence d'une insertion



Perspectives

- Valorisation
 - par un article scientifique
 - par la diffusion de KaSIMir aux laboratoires de diagnostic, afin de tester et valider leur pipeline NGS, ceci dans une démarche d'accréditation.
 - par la diffusion d'eKLIPse sous forme d'un programme téléchargeable et/ou sous forme d'un plugin inclus dans la suite logicielle Life.

Remerciements

UMR CNRS 6214 – INSERM 1083 – BNMI / Département Génétique-Biochimie du CHU d'Angers

Dr. David Goudenège

Pr. Vincent Procaccio

Pr. Daniel Henrion

Dr. Marie-Claire Malinge

Dr. Patrizia Amati-Bonneau

Dr. Delphine Mirabeau-Prunier

**LINA, Laboratoire Informatique de Nantes
Atlantique**

Pr. Bourdon Jérémie

Dr. Guy Lenaers

Pr. Pascal Reynier

Dr. Céline Bris

Dr. Charlotte Veyrat-Durebex

Dr. Valérie Desquirit-Dumas

et les différentes équipes d'accueil ...

Projet GRIOTE



Merci de votre attention !