



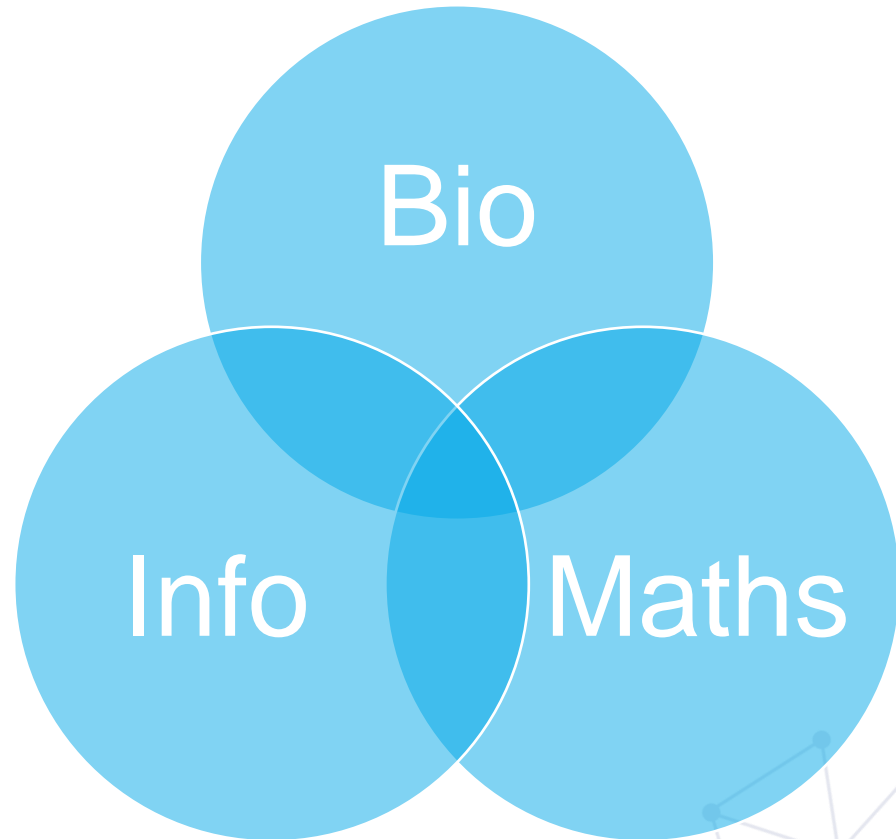
La bioinformatique en laboratoire de biologie médical

LIPINSKI Boris

Bioinformaticien – Eurofins Biomnis

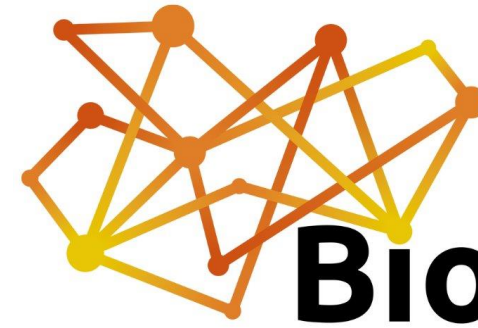
ChatGPT :

- **Domaine interdisciplinaire entre :**



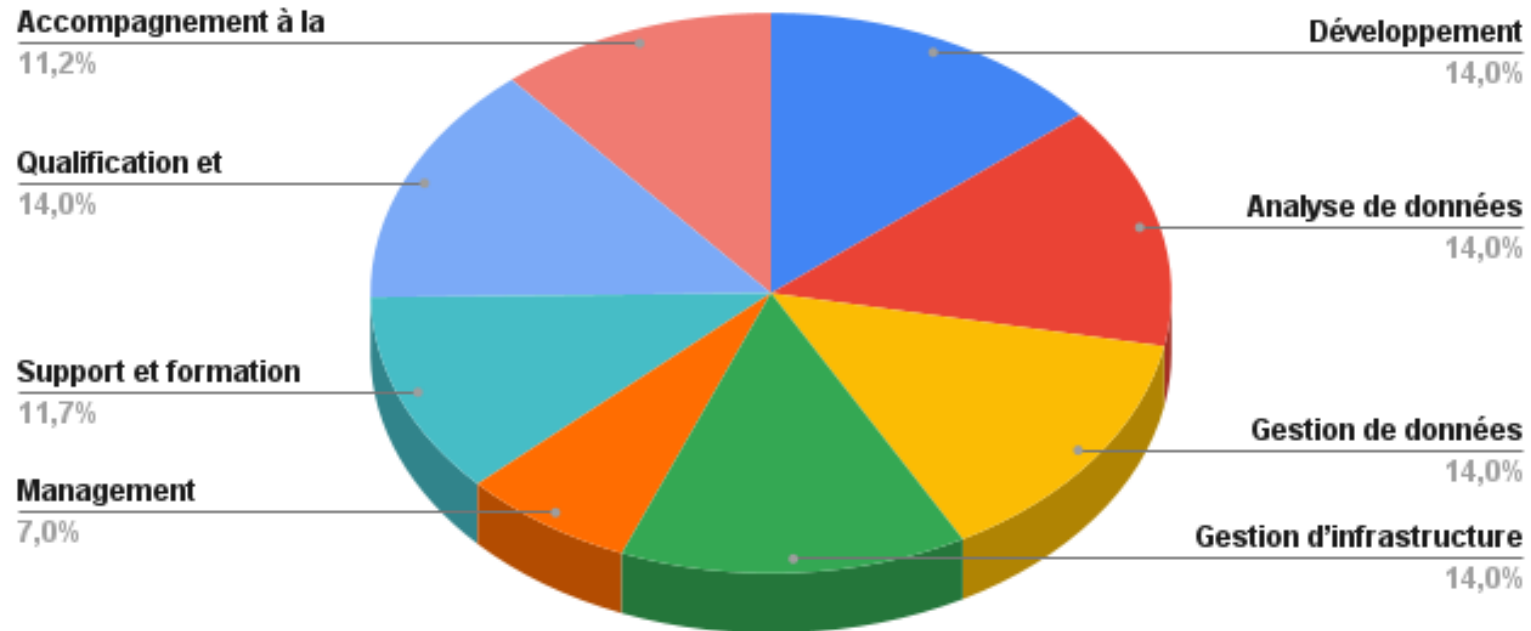
- **Analyser et interpréter les données biologiques : génomiques et protéomiques.**

BioInfoDiag :



BioInfoDiag

Profil Polyvalent



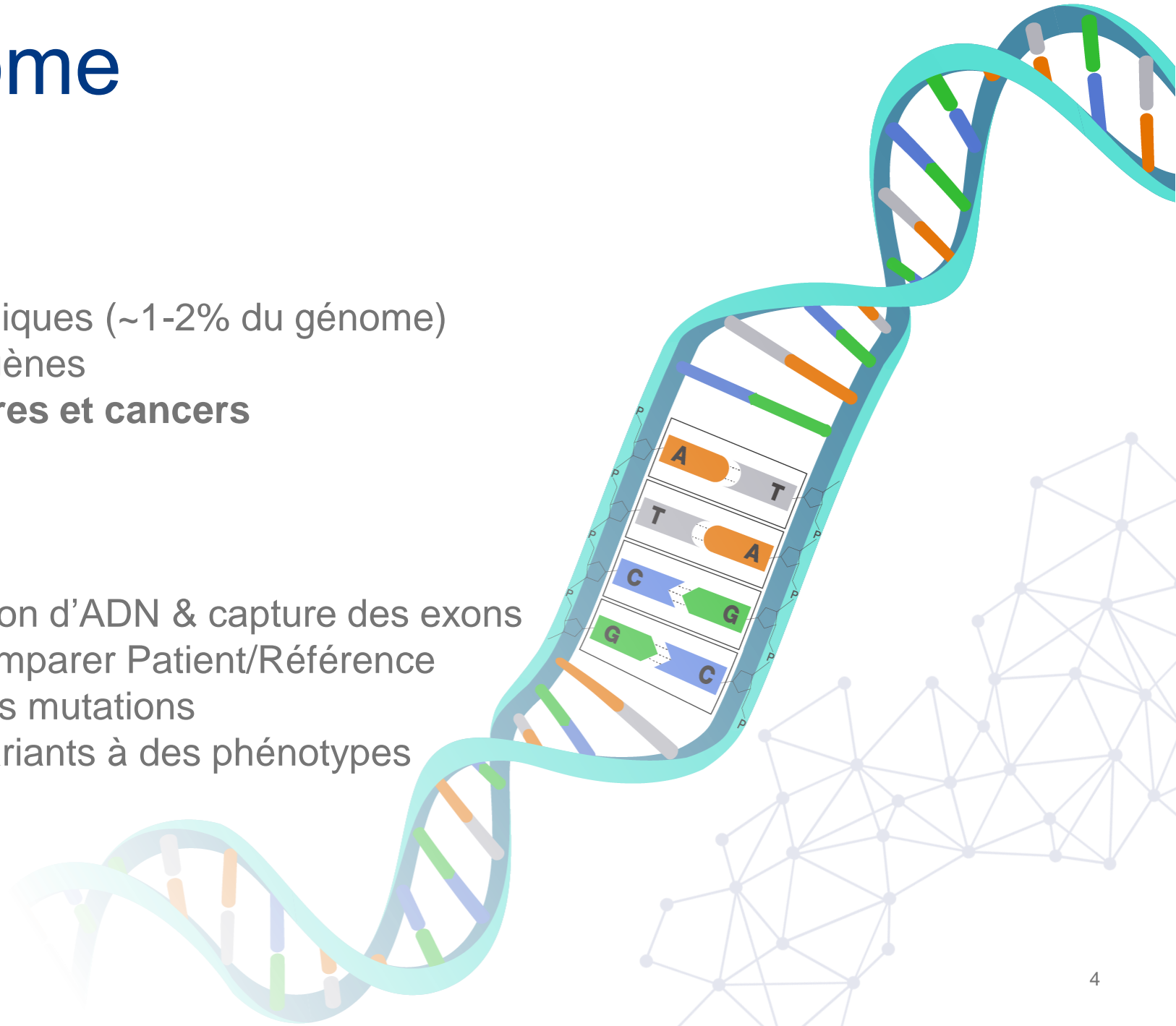
Analyse d'Exome

Objectif :

- Identifier des mutations exoniques (~1-2% du génome)
- Détecter des variants pathogènes
- Diagnostiquer maladies rares et cancers

Processus :

- **Séquençage ciblé** : Extraction d'ADN & capture des exons
- **Alignement de lecture** : Comparer Patient/Référence
- **Appel de variants** : Lister les mutations
- **Interprétation** : Relier les variants à des phénotypes



Read :

```
@ST-E00104:1299:HG7TFCX2:3:1101:6380:1643 1:N:0:TCCGGAGA  
TACAGTGGTCAACTGTACTATCGGATTTCTTATATGGGGCCCCGCGTCCCGCCCTGGTGGCCGGCTGGGGGCGACAAGTCTGAGAGAACCAGACGGAAGCGCGCTGGGACTGACACGGGACTTGGGCGGTGCTGCCCGGTGGGTC  
+  
-AFFFFAJJFJFJJJJFFJF -AAAAAFJJJJJJJJJJJJJJJJ -FFJ7FJJAJJJJJFJJJJJF<F -AJJFJJJJFAJJJJ7AJ7AJJJJJJAFFF<A -FJF7 -F<<7AJ<FJF7 -A -A -AJ -77AF7AJFJJJJ7AA<JF7AAJA - -7 -A -
```

FASTQ



Génome de référence



VCF

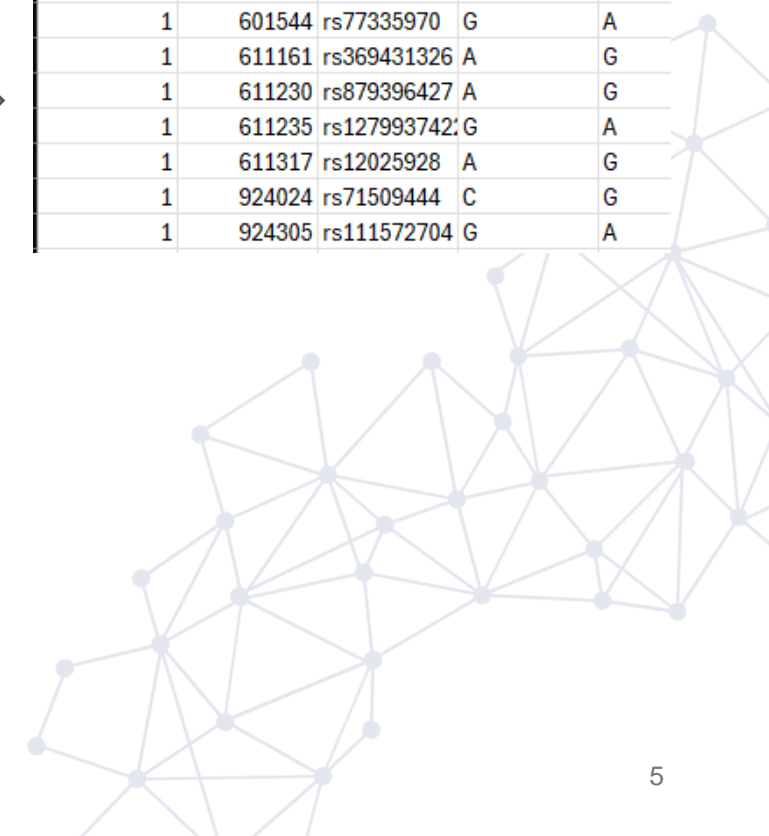
#CHROM	POS	ID	REF	ALT
1	601436	rs373360530	C	T
1	601515	rs78713263	T	C
1	601544	rs77335970	G	A
1	611161	rs369431326	A	G
1	611230	rs879396427	A	G
1	611235	rs127993742	G	A
1	611317	rs12025928	A	G
1	924024	rs71509444	C	G
1	924305	rs111572704	G	A



XX milliard de lectures



PANEL EXOME



Read :

```
@ST-E00104:1299:HG7TFCCX2:3:1101:6380:1643 1:N:0:TCCGGAGA  
TACAGTGGTCAACTGTACTATCGGATTTCTTATATGGGGGCCCGCTGCCCGCCTGGTGGCCGGCTGGGGGCGACAAGTCCTGAGAGAACCAGACGGAAGCGCGCTGGGACTGACACGGGGACTTGGCGGTGCTGCCCGGTGGGT  
+  
-AAFFFAJJFJFJJJFFJF-AAAAAFJJJJJJJJJJJJJJ-FFJ7FJJAJJJJJFJJJJJF<F-AJJFJJJJFAJJJJ7AJ7AJJJJJAFFF<A-FJF7-F<<7AJ<FJF7-A-A-AJ-77AF7AJFJJJJ7AA<JF7AAJA--7-A-
```

FASTQ



AGCA Génome de référence



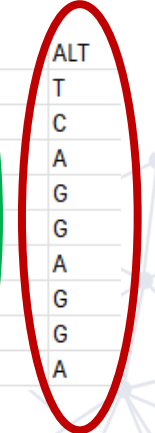
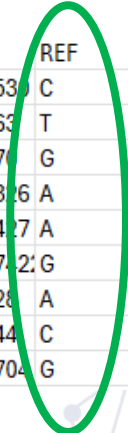
AGTA



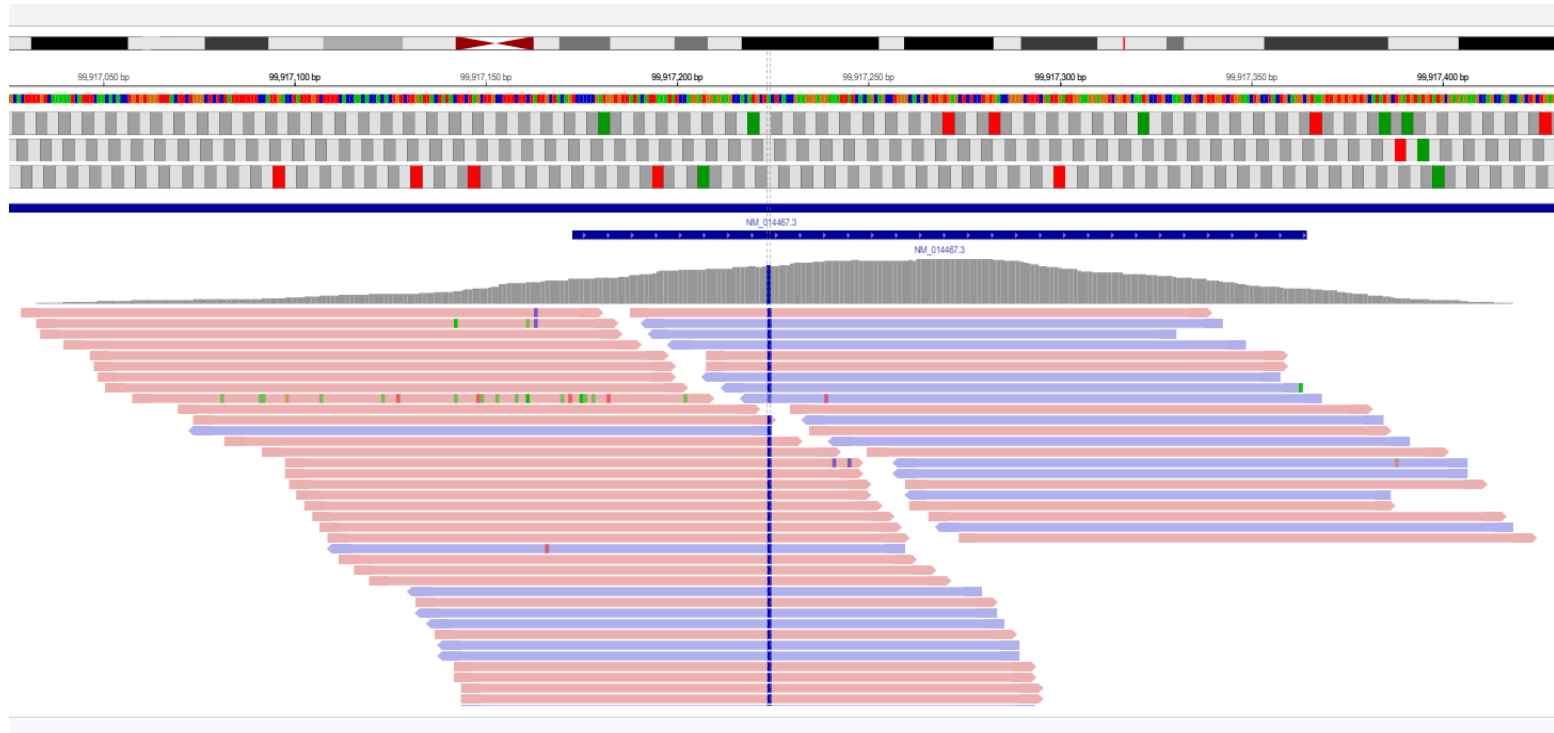
PANEL EXOME



#CHROM	POS	ID	REF	ALT
1	601436	rs37336053	C	T
1	601515	rs78713263	T	C
1	601544	rs7733597	G	A
1	611161	rs3694313	A	G
1	611230	rs8793964	A	G
1	611235	rs1279937	G	A
1	611317	rs1202592	A	G
1	924024	rs71509444	C	G
1	924305	rs111572704	G	A



XX milliard de lectures

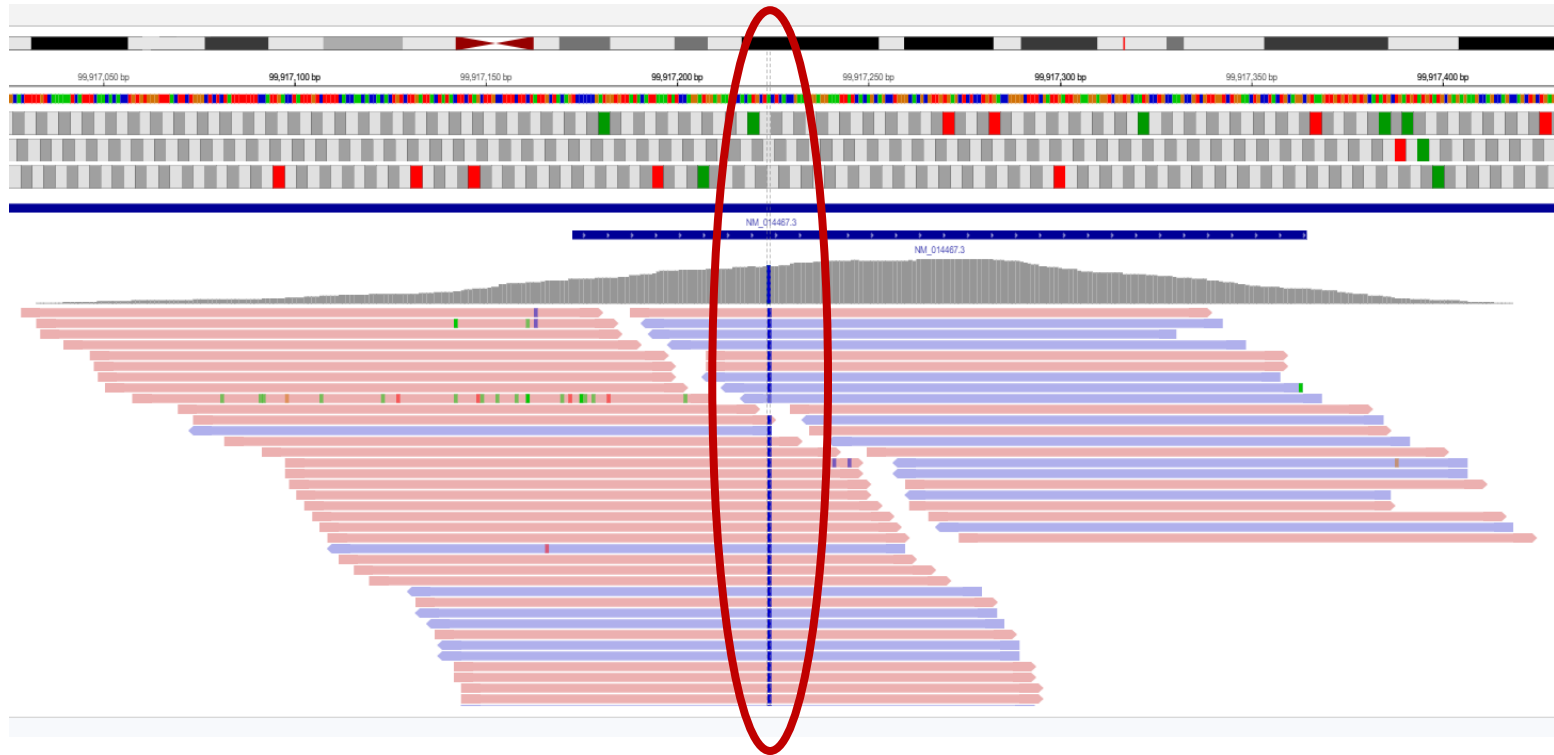


REFERENCE



XX milliard de lectures
FASTQ

A



REFERENCE



XX milliard de lectures
FASTQ

C/C
1/1



Mutation
détectée



Diagnostic

Analyse de Violations Mendéliennes

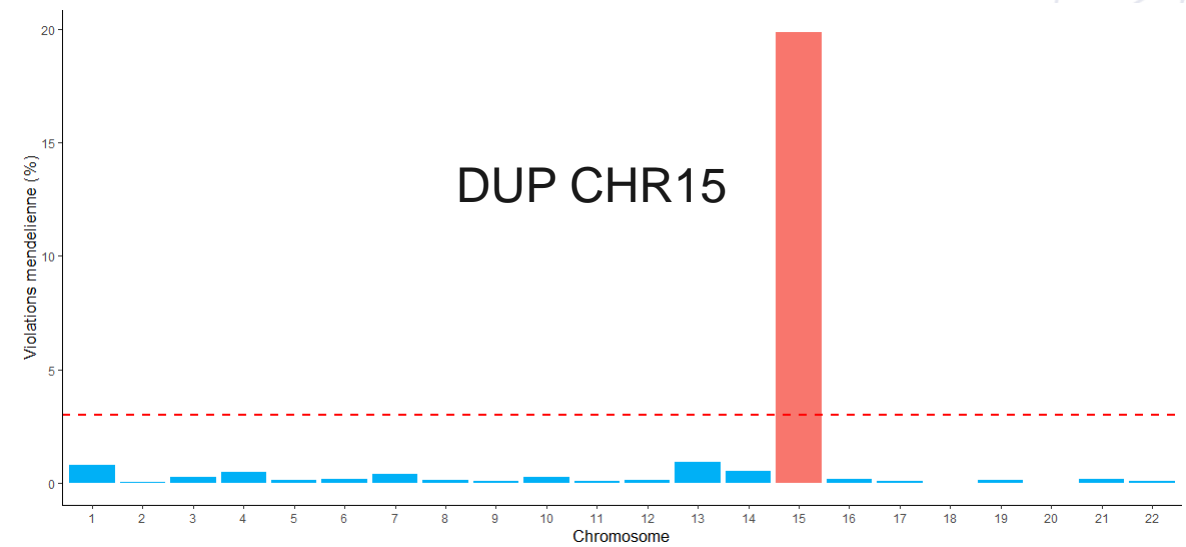
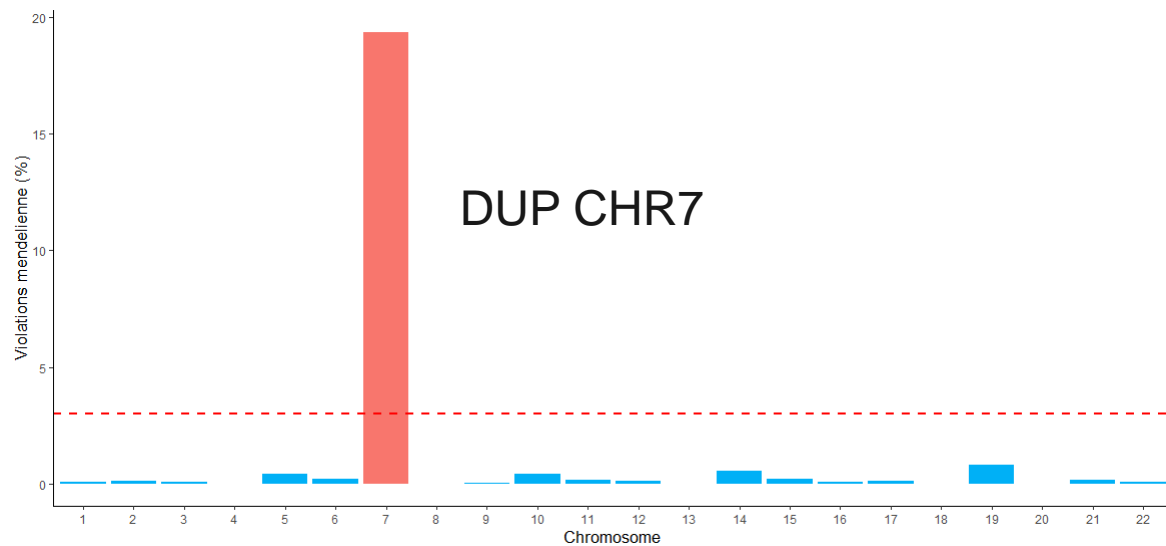
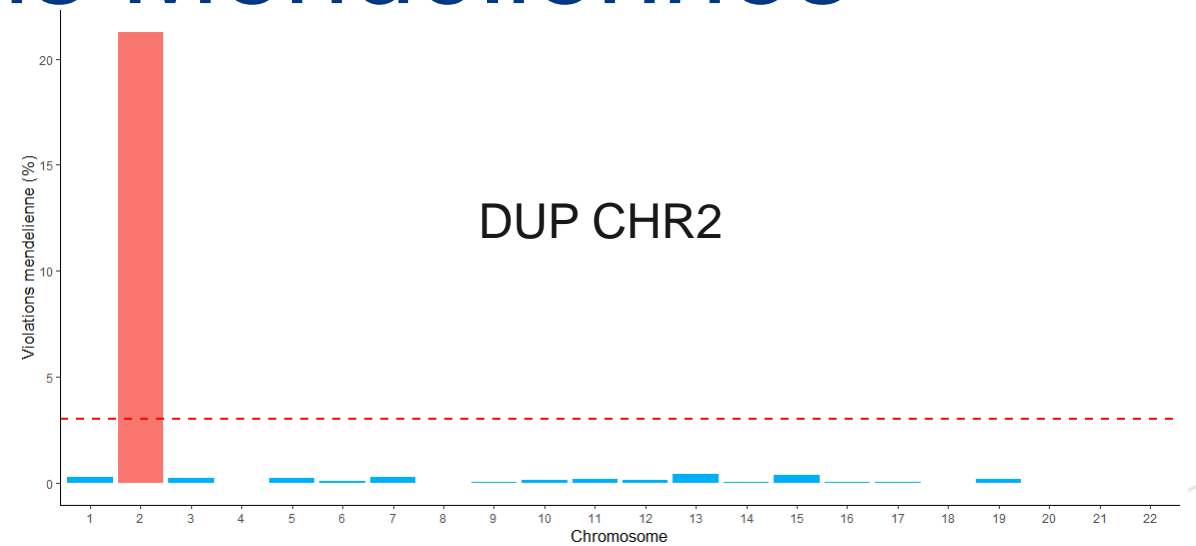
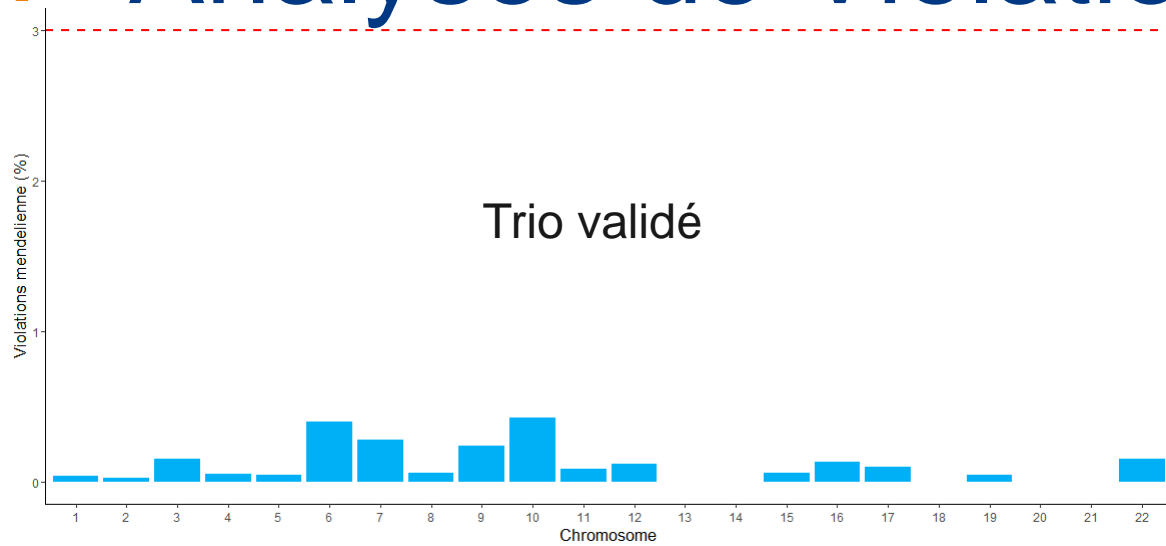
Objectif : Au sein d'un trio

- Identifier le % de génotypes non conformes aux lois mendéliennes :

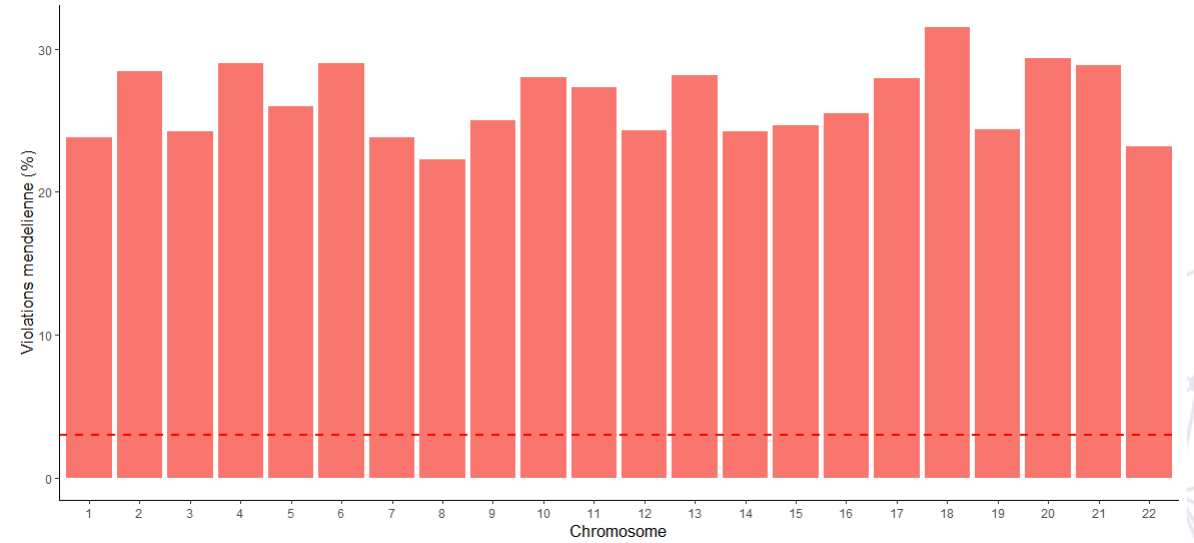
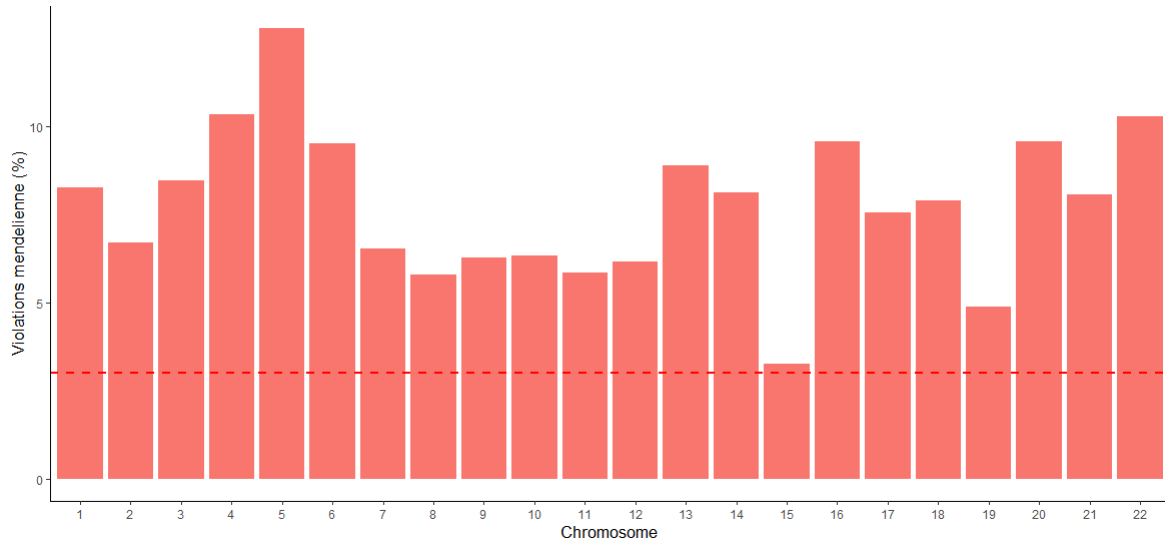
Père	Mère	Cas index	Interprétation	
1/1	+	→	1/1	Trio validé
			1/0	Trio validé
			0/0	Violation Mendélienne

- Résultats possibles :
 - Validation des familles : % faible (<0,5%)
 - Disomies uniparentales : % moyen
 - Inversion de tube / Fausse paternité : % fort (>20%)

Analyses de Violations Mendéliennes



Analyses de Violations Mendéliennes



DUP en
mosaïque



Trio invalidé

Communication Bio-Bioinfo

NORME ISO 15189 – SH-GTA-16



Création de pipeline :

- Définition du flux, reproductibilité, fidélité intermédiaire, exactitude, maîtrise de risque
- MAJ mineure : dossier de non-régression
- MAJ majeure : dossier de qualification

Formation du personnel :

- Qualification, habilitation, maintien des compétences ...



Traçabilité :

NomRun	Sequençage	Rsync Archive	bioIT-SNV				Dépôt SeqOne		bioIT-CNV			Commentaires	
	date RUN fin		Vérifié par	Date fin	≠ B,D	IV	≠ D,F	Vérifié par	Date fin	Vérifié par	Date		≠ B,K
RTV2055	19/04/2024 10:44	FP	20/04/2024 12:37	1,078414352	20/04/2024 14:46	0,08947	FP	29/04/2024	FP/BL	21/04/2024 02:41	1,664479167	FP	
RTV2056	26/04/2024 06:56	VG	27/04/2024 02:38	0,821053241	27/04/2024 04:59	0,09744	VG	29/04/2024 13:40	VG	28/04/2024 00:58	1,751608796	VG	
RTV2057	02/05/2024 10:24	FP	04/05/2024 03:30	1,712083333	04/05/2024 05:33	0,08569	FP	06/05/2024 15:43	BL	04/05/2024 15:56	2,230451389	FP	Présence de pseudo dossier dans le run => echec de prepar NC A-05-24-20991
RTV2058	09/05/2024 10:46	VG	10/05/2024 05:04	0,762581019	10/05/2024 11:35	0,27153	VG	10/05/2024 18:49	BL	11/05/2024 04:34	1,741736111	BL	Lignes en double dans les genotypes tables => collectQC en relance. Dépôt manuelle de tout les fichiers de validation d dossier de routine. NC A-04-24-20911 complétée
RTV2059	17/05/2024 11:14	BL	21/05/2024 18:00	4,282199074	27/05/2024 16:26	5,93448	BL	22/05/2024 20:36	BL	22/05/2024 11:08	4,996076389	BL	
RTV2060	23/05/2024 16:32	FP	23/05/2024 22:26	0,245972222	23/05/2024 22:49	0,01587	FP	24/05/2024 17:04	BL	24/05/2024 03:04	0,438726852	FP	S1
RTV2061	24/05/2024 18:16	FP	24/05/2024 23:11	0,205439815	24/05/2024 23:38	0,01821	FP	27/05/2024 14:47	BL	25/05/2024 02:30	0,343287037	FP	S1
RTV2062	30/05/2024 11:09	BL	31/05/2024 10:35	0,976736111	31/05/2024 13:25	0,11807	BL	31/05/2024	BL		-45442,4647	BL	Un échantillon qui a été bloqué dans Genius, et n'a pas été cet échantillon dans le run RTV2062A (VG)
RTV2063	06/06/2024 09:42	VG	07/06/2024 05:09	0,809918981	07/06/2024 07:31	0,09874	VG	09/06/2024 09:30	VG	07/06/2024 20:18	1,441539352	VG	
RTV2064	13/06/2024 11:45	FP	14/06/2024 07:30	0,822858796	17/06/2024 11:59	3,18671	FP	14/06/2024 19:03	FP	15/06/2024 04:06	1,681238426	FP	

Mise en PROD

Ex : acide gras

- But : faire un...

Sample nam	Sample ID	C15	C19	Acide myristi	Acide palmit	Acide trans-1	Acide palmit	Acide stearic	Acide elaidic	Acide trans-
20240222_00	BLA									
20240222_00	Blanc+S			0.04	0.18					0.16
20240222_00	Point 1			0.11	1.03					0.05
20240222_00	Point 2			0.26	2.42					0.12
20240222_00	Point 3			0.48	4.99					0.25
20240222_00	Point 4			0.98	10.62					0.54
20240222_00	Point 5			2.30	23.95					1.27
20240222_00	Point 6			4.62	47.92					2.45
20240222_00	Point 7			10.60	102.57	5.01	10.61	53.25		5.00
20240222_01	Sbioda 1/2			2.34	60.73		3.00	21.42	1.68	
20240222_01	CQI			6.71	58.32	0.03	6.29			
20240222_01	CQI						6.22			
20240222_01	CQI 1/2						3.43			
20240222_01	CQI 1/2						3.43			

Maitrise des risques

Procédure de qualification

Changement des références

Aller-retour avec les bios

Discussion avec la technique

Mise à jour

Ecriture de dossier de qualification

Passation aux autres bio-informaticiens

Transfert dans OL avec l'IT



Ecriture du pipeline : 2 semaines

MAIS ...



Amélioration continue des pipelines

Nom	Modifié le	Type	Taille
LC480 31234_638398762864313682	03/01/2024 10:58	Microsoft Edge H...	495 Ko
LC480 31234_638398822905197825	03/01/2024 12:38	Microsoft Edge H...	495 Ko
LC480 31234_638398884754535404	03/01/2024 14:21	Microsoft Edge H...	496 Ko
LC480 31234_638200625624750757	04/01/2024 10:56	Microsoft Edge H...	496 Ko

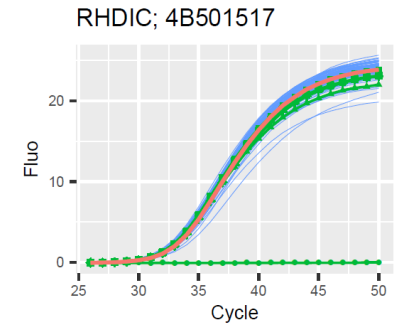
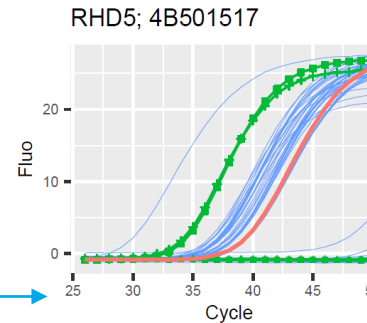
208...

INFO1

INFO2

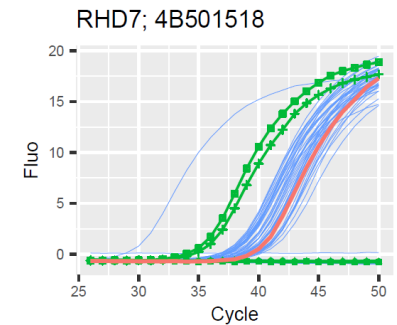
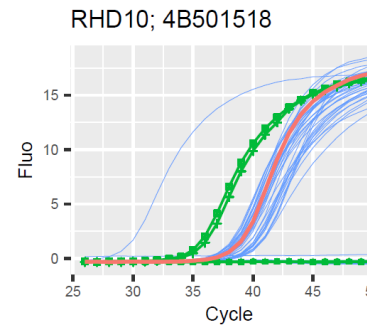


V1.0.0



color

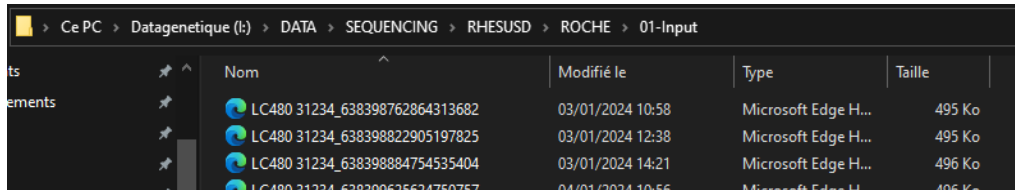
- 3587066702
- Controls
- Others



shape

- A1-BLPCR
- B1-RHD-
- B7-RHD-
- C1-RHD+
- C7-RHD+

Amélioration continue des pipelines



Nom	Modifié le	Type	Taille
LC480 31234_638398762864313682	03/01/2024 10:58	Microsoft Edge H...	495 Ko
LC480 31234_638398822905197825	03/01/2024 12:38	Microsoft Edge H...	495 Ko
LC480 31234_638398884754535404	03/01/2024 14:21	Microsoft Edge H...	496 Ko
LC480 31234_638200625624750757	04/01/2024 10:56	Microsoft Edge H...	496 Ko

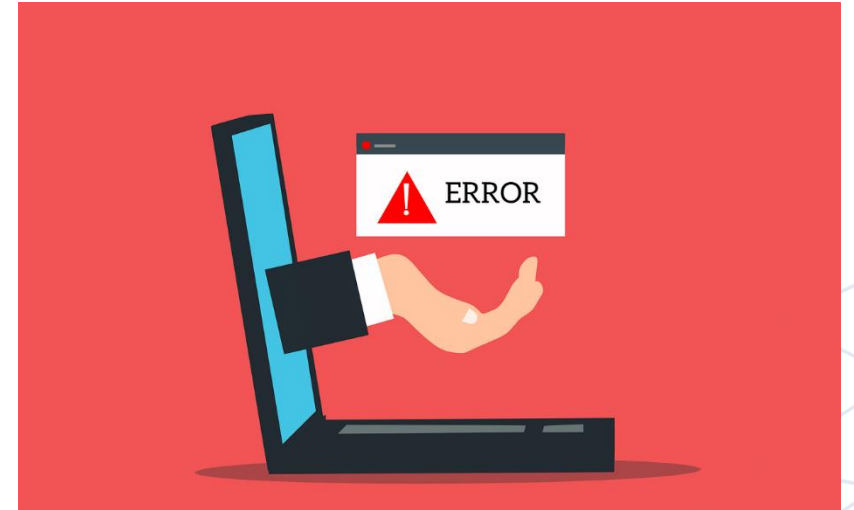
209...

INFO1

INFO2



V1.0.0



Amélioration continue des pipelines

Nom	Modifié le	Type	Taille
LC480 31234_638398762864313682	03/01/2024 10:58	Microsoft Edge H...	495 Ko
LC480 31234_638398822905197825	03/01/2024 12:38	Microsoft Edge H...	495 Ko
LC480 31234_638398884754535404	03/01/2024 14:21	Microsoft Edge H...	496 Ko
LC480 31234_638200625624750757	04/01/2024 10:56	Microsoft Edge H...	496 Ko

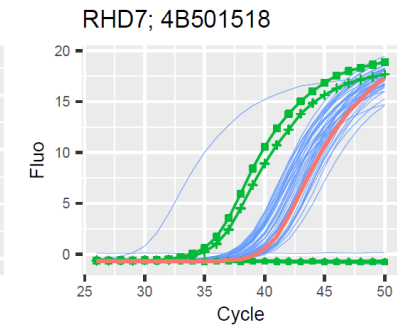
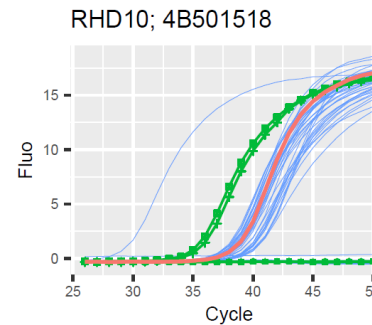
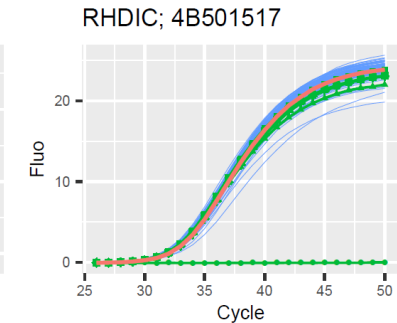
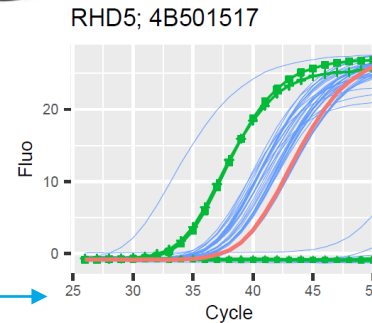
209...

INFO1

INFO2



V2.0.0



color
— 3587066702
— Controls
— Others

shape
• A1-BLPCPR
▲ B1-RHD-
■ B7-RHD-
+ C1-RHD+
■ C7-RHD+

Amélioration continue des pipelines

Exome :

2022 : 90 patients / série → 12h

2024 : 250 patients / série → 30h?

Amélioration continue des pipelines

Exome :

2022 : 90 patients / série → 12h

2024 : 250 patients / série → 30h?

Solutions possibles :

- Nouveaux serveurs de calculs
- Optimisation du pipeline
 - Mise à jour des outils
 - Parallélisation des processus
 - Optimisations des ressources des serveurs

2024 : 250 patients / série →

13h



Alerte infrastructure

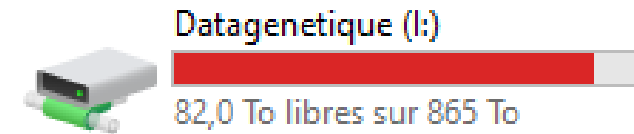
Problème :

Le besoin

- Nouveau pipeline
- Nouveau gène
- Nouveau variant
- Nouveau score



Emplacements réseau (5)

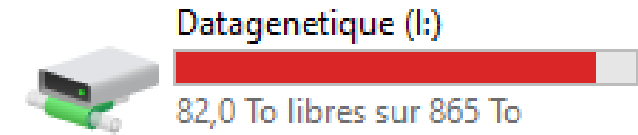


Le coût

- 1,5cts/GiB/mois
- 1 run exome = 24€/mois
- 4 runs / mois = 1 200€/ans

Alerte infrastructure

Emplacements réseau (5)



Problème :

Le besoin

- Nouveau pipeline
- Nouveau gène
- Nouveau variant
- Nouveau score

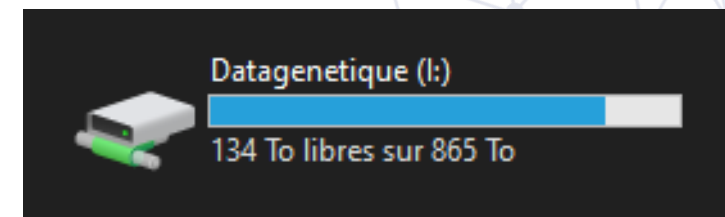


Le coût

- 1,5cts/GiB/mois
- 1 run exome = 24€/mois
- 4 runs / mois = 1 200€/ans

Solution :

- Choix des fichiers à conserver (fichier source + le code)
- Compresser les fichiers sources
- Supprimer les anciens fichiers
- Procédure de nettoyage périodique



Pipelines bioIT à Eurofins Biomnis

Génétique humaine

- Design d'amorces Sanger
- Exome (SNV, CNV, mito, MUC1,HLA, expansion, contamination ...)
- Pharmacogénétique
- Génome
- DPNI
- Nanopore ciblée
- Nanopore Génome
- Base de données intégrée

Génétique des cancers

- Panel Lympho-Myeloid (FLT3, CNV, IGVH ...)
- Panel Tumeurs Solides

Biologie Moléculaire Infectio

- 16S
- Microbiote intestinale
- Mycobactéries

Chimie

- Acide gras érythrocytaires
- PFAS

CoreLab

- Rhésus D





 eurofins

Biomnis



**Merci pour votre
attention**

Dis, ChatGPT : quel est l'avenir de la bioinformatique en LBM ?

1. Intégration de l'intelligence artificielle (IA) et de l'apprentissage automatique



Analyse prédictive : Utilisation de l'IA pour prédire les risques de maladies, la réponse aux traitements, et identifier de nouveaux biomarqueurs.



Automatisation des pipelines : Développement de pipelines entièrement automatisés capables de traiter des volumes massifs de données avec une intervention humaine minimale.

2. Médecine personnalisée et de précision



Profilage génomique avancé : Utilisation de la bioinformatique pour fournir des analyses génomiques détaillées, permettant des traitements personnalisés basés sur le profil génétique unique du patient.



Thérapies ciblées : Identification de cibles thérapeutiques spécifiques et développement de médicaments personnalisés.

3. Big Data et intégration multi-omique



Analyse intégrée : Combinaison de données génomiques, transcriptomiques, protéomiques, métabolomiques, et cliniques pour obtenir une vision holistique de la santé et des maladies.



Bases de données centralisées : Développement de bases de données globales pour stocker et partager des données multi-omiques à grande échelle

4. Technologies de séquençage avancées

- Séquençage à cellule unique : Analyse des variations génomiques et transcriptomiques à l'échelle d'une seule cellule pour des informations ultra-précises.



Séquençage long-read : Amélioration de l'assemblage des génomes et de la détection des variations structurales grâce aux technologies de séquençage long-read.

ChatGPT : quel est l'avenir de la bioinformatique en LBM ?

1. Amélioration des infrastructures et des outils bioinformatiques

 Cloud computing : Utilisation de plateformes de cloud computing pour le stockage et l'analyse des données, permettant une scalabilité et une flexibilité accrues.

 Développement de nouveaux logiciels : Création de logiciels et d'algorithmes plus rapides, précis et user-friendly pour l'analyse bioinformatique

2. Collaboration interdisciplinaire et formation

 Équipes multidisciplinaires : Renforcement des collaborations entre bioinformaticiens, biologistes, cliniciens et data scientists pour maximiser l'impact des analyses bioinformatiques.

 Formation continue : Programmes de formation pour les professionnels de la santé et les chercheurs afin de maintenir leurs compétences à jour avec les avancées bioinformatiques.

3. Applications cliniques avancées

 Diagnostics précoces : Utilisation de la bioinformatique pour détecter précocement les maladies grâce à l'analyse des données génétiques et autres biomarqueurs.

 Suivi thérapeutique : Surveillance continue de l'efficacité des traitements et adaptation en temps réel des stratégies thérapeutiques.

4. Éthique et régulation

 Confidentialité et sécurité des données : Mise en place de protocoles stricts pour assurer la confidentialité et la sécurité des données des patients.

- Réglementation et standards : Développement de standards internationaux pour l'utilisation éthique et responsable des données bioinformatiques.

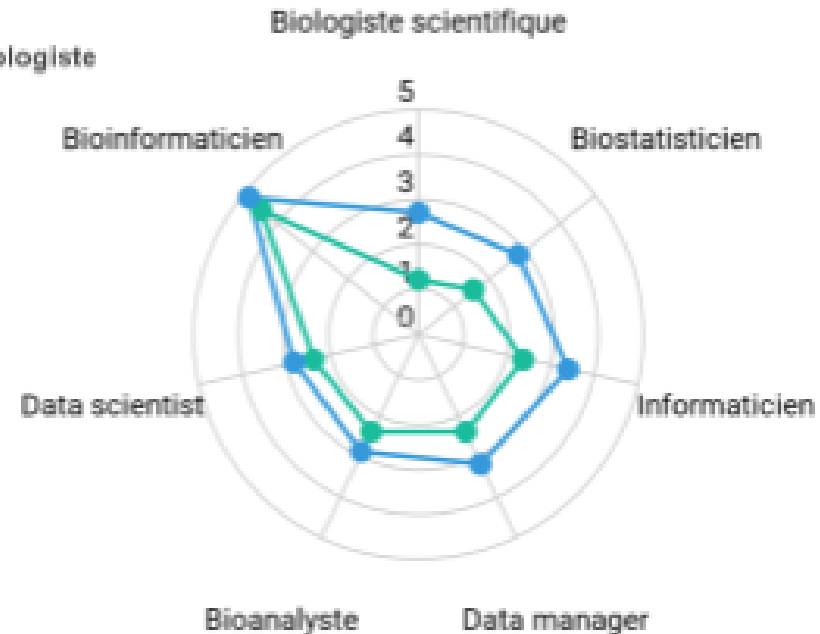


Comparaison des perceptions du métier

● BioInfo ● Biologiste

- Pour les bioinformaticiens :
description de vos activités
- Pour les biologistes :
compétences attendues

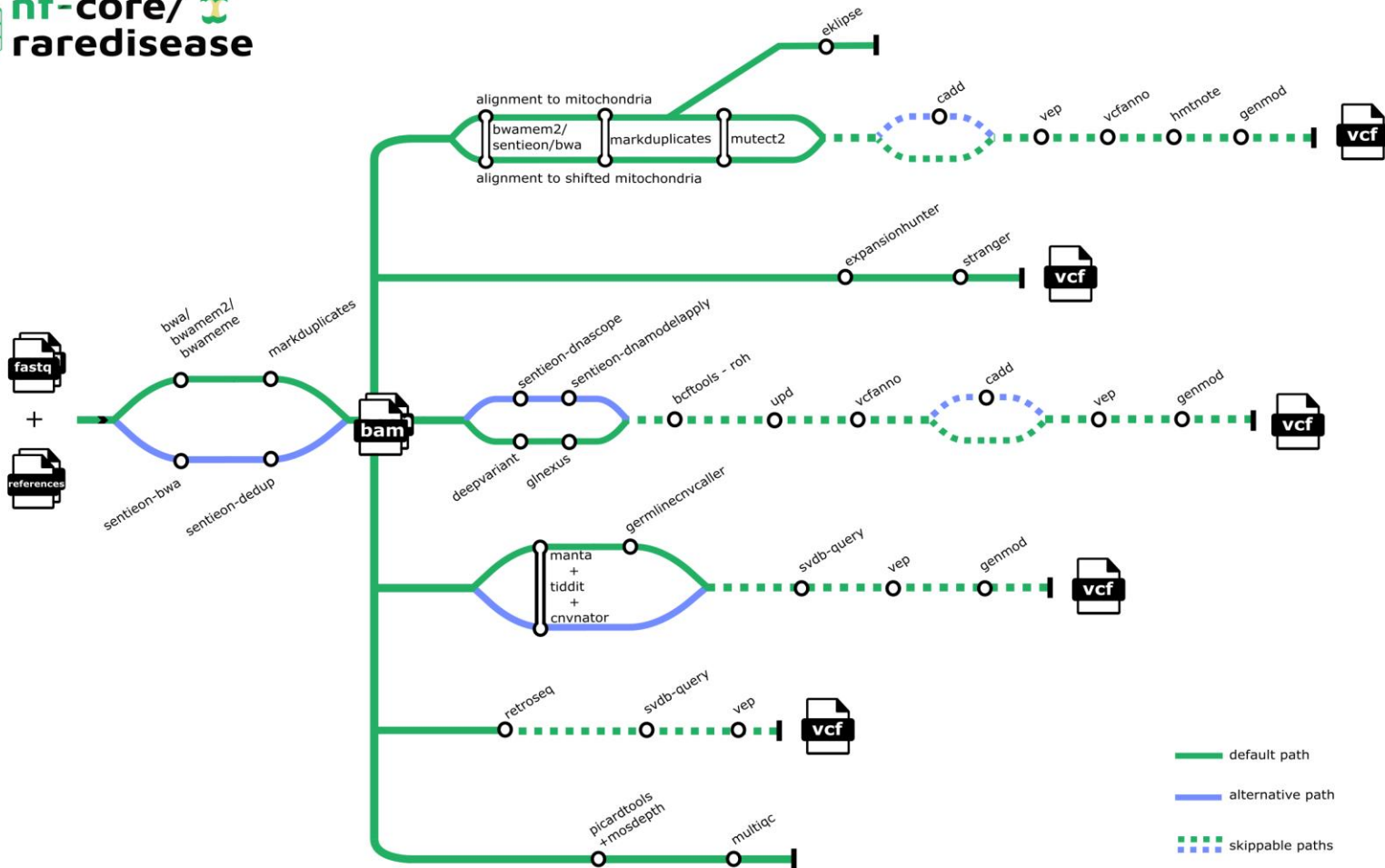
Il existe un **décalage** entre les compétences attendues par les biologistes et les activités décrites par les bioinformaticiens.

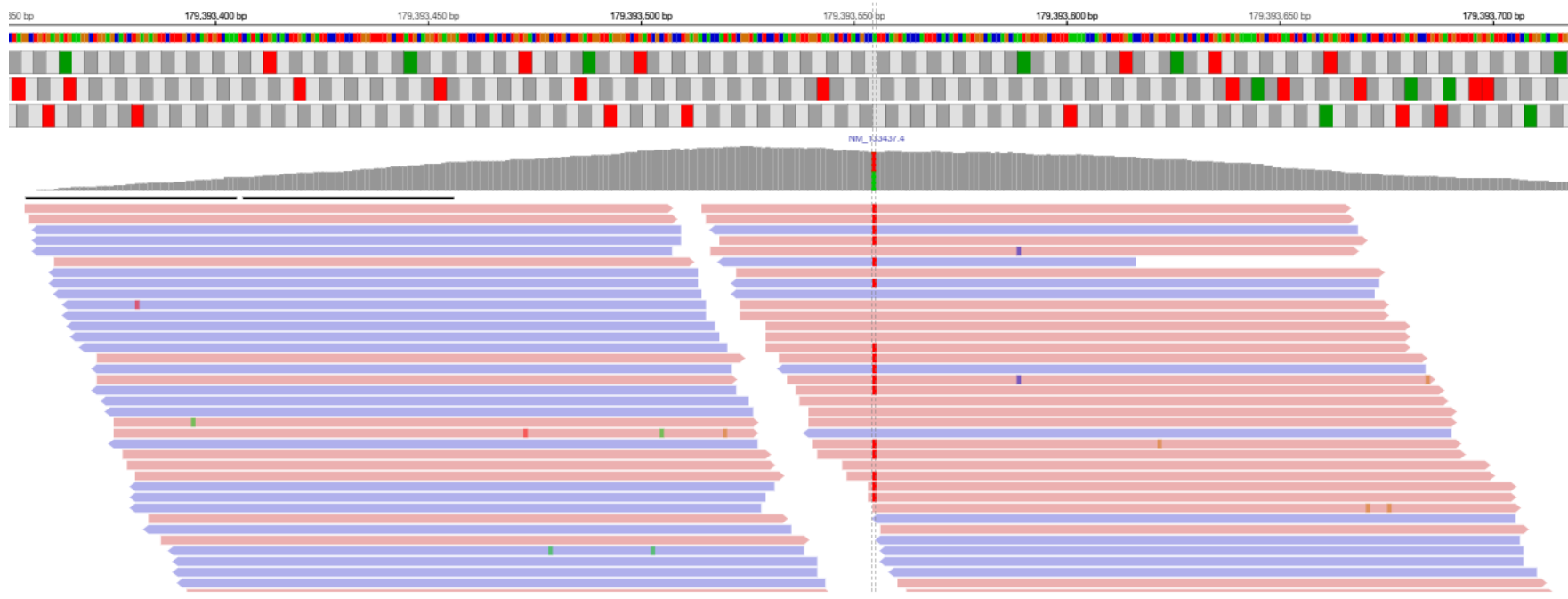


Comment ? Avec un pipeline !



FASTQ



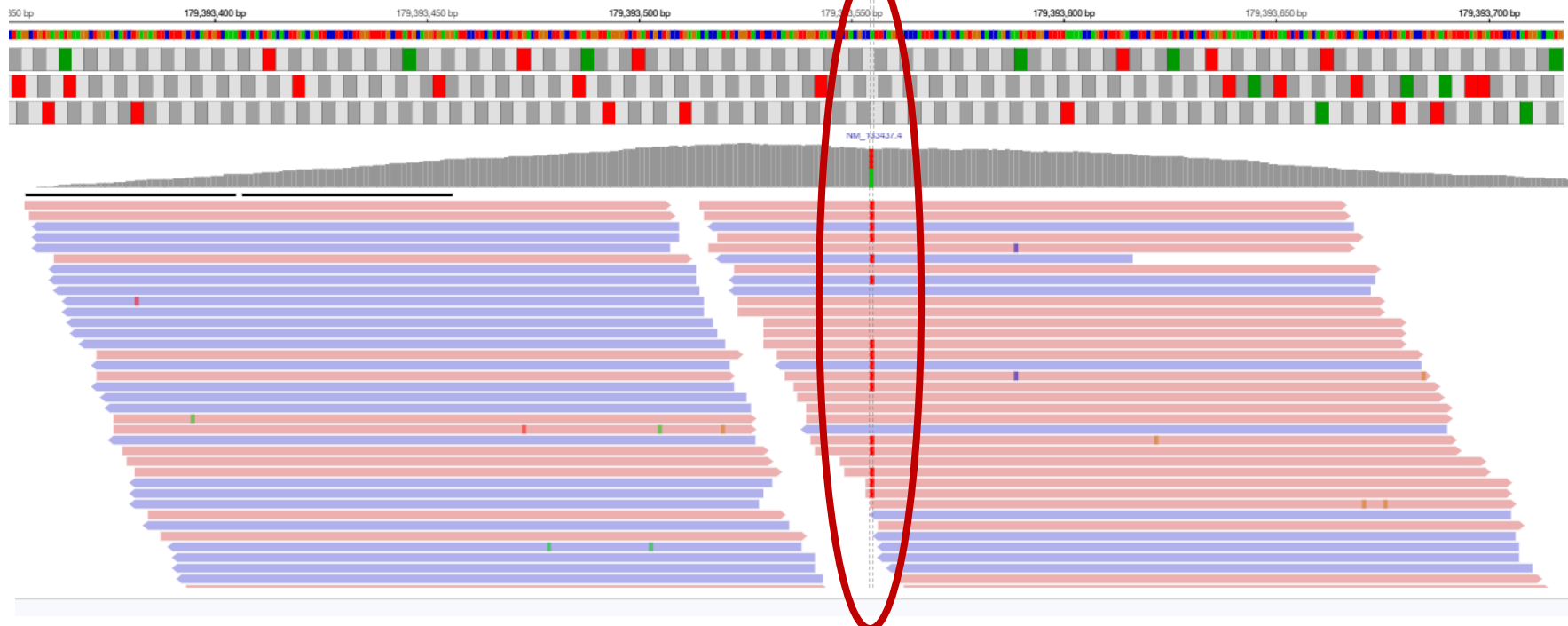


REFERENCE



XX milliard de lectures
FASTQ

A



REFERENCE



XX milliard de lectures
FASTQ

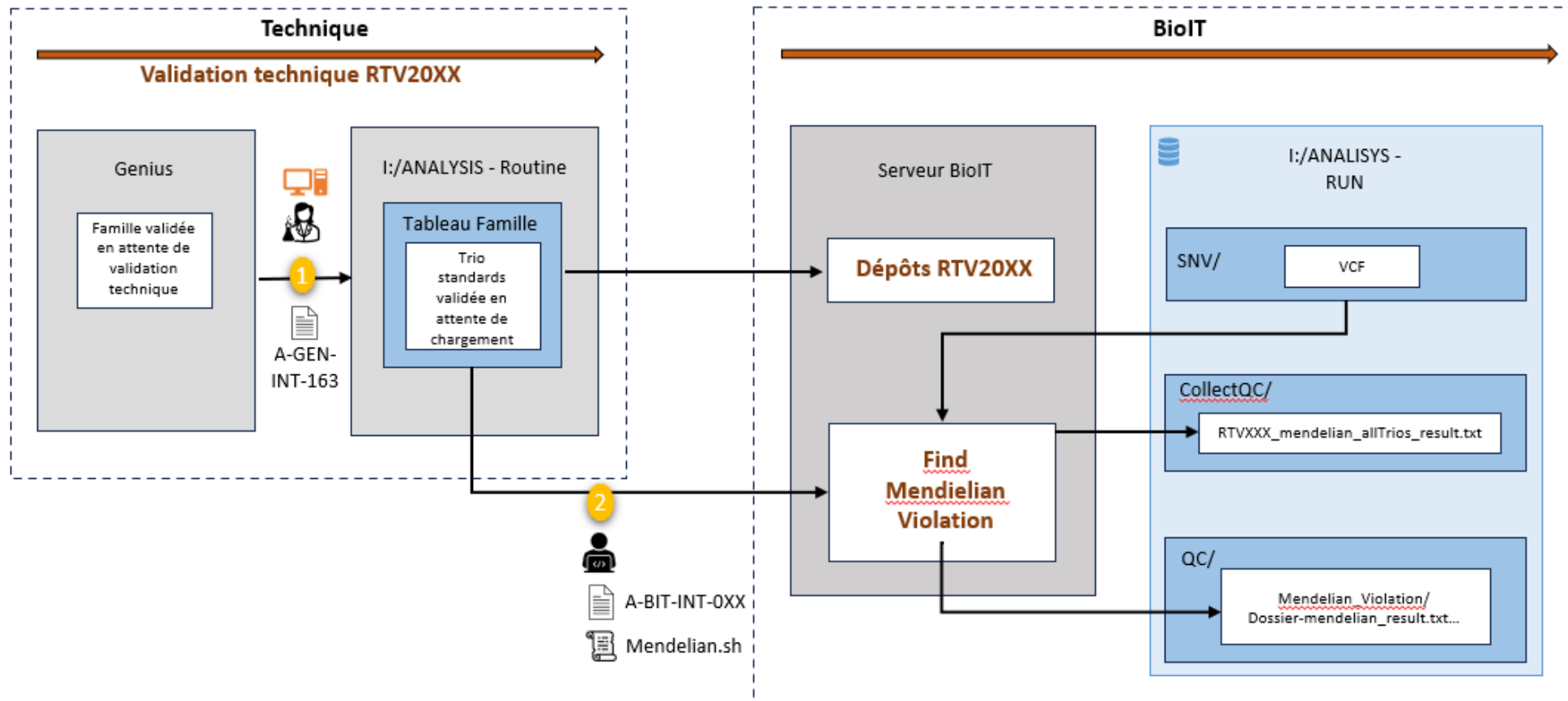
A/T
0/1



Variation hétérozygote

Analyses de Violations Mendéliennes

Schéma du flux de données du pipeline : Violations Mendéliennes



Localisation: Eurofins Biomnis Lyon Gerland

Amélioration continue des pipelines

Analyse du gène *MEFV* en séquençage long reads

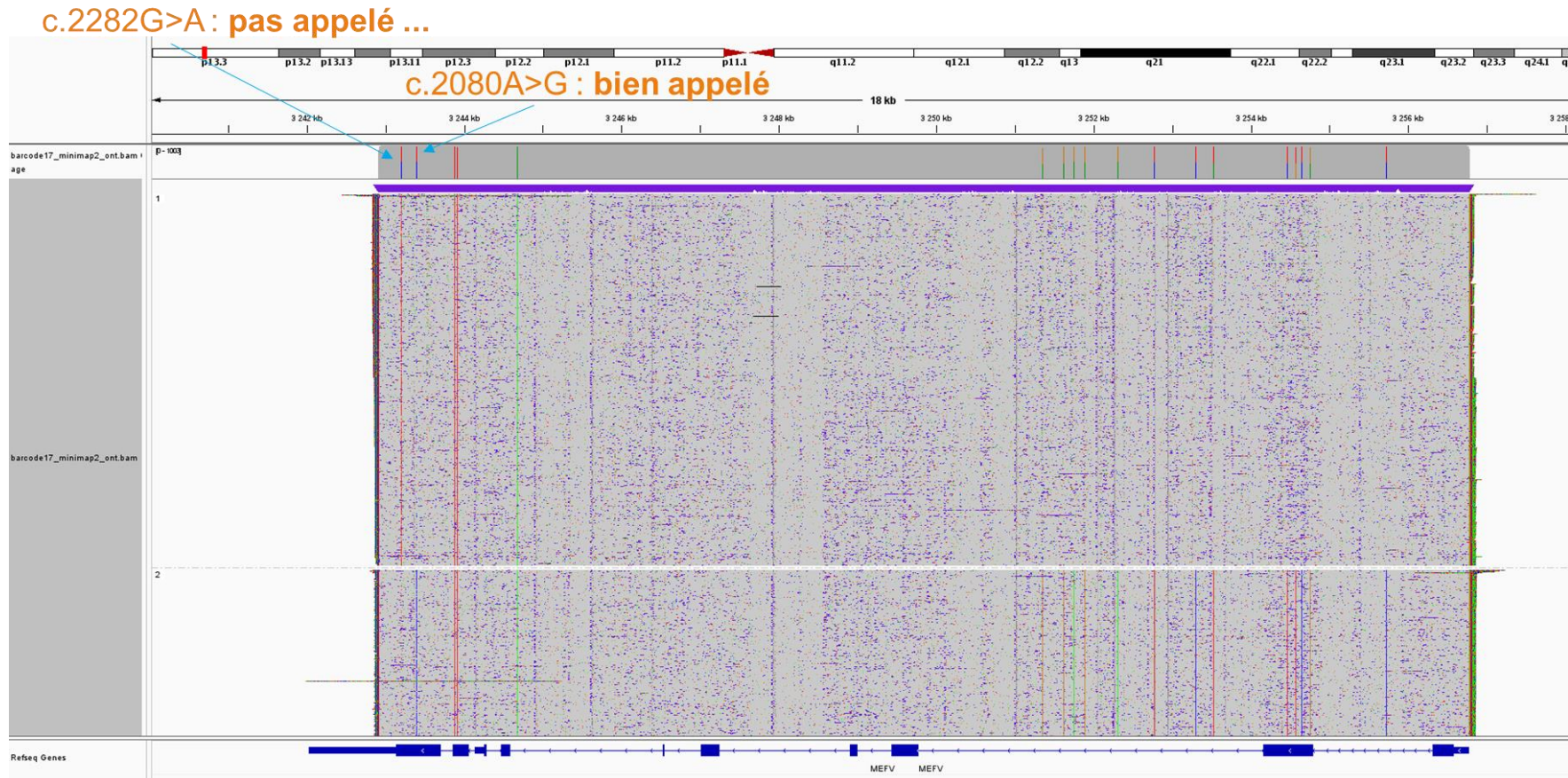
- RUN 39 : CQI en échec
 - Attendu : variants c.2282G>A et c.2080A>G, en trans
 - Obtenu : variant c.2080A>G

Troubleshooting : erreur d'échantillon ? Erreur de manip ? Erreur bio-it ?

[NC : A-05-24-21068](#)

Amélioration continue des pipelines

Analyse du gène *MEFV* en séquençage long reads



Amélioration continue des pipelines

Analyse du gène *MEFV* en séquençage long reads

- RUN 39 : CQI en échec Pipeline Nanopore_TargetedV3
 - **Attendu** : variants c.2282G>A et c.2080A>G, en trans
 - **Obtenu** : variant c.2080A>G

Troubleshooting : erreur d'échantillon ? Erreur de manip ? Erreur bio-it ?

[NC : A-05-24-21068](#)

Pipeline Nanopore_TargetedV4 avec nouvel appel de variant, qui a résolu le pb
+ réanalyse des runs antérieurs