Plateforme Auragen
Infra, routine & dev bioinfo

15 mai 2024

Bioinfodiag

Virginie Bernard, Auragen

Le parcours AURAGEN

HYGEN

- Création compte (1ère fois)
- Renseignements cliniques
- Effectuer la demande d'analyse génétique

Validation de la prescription

Patient

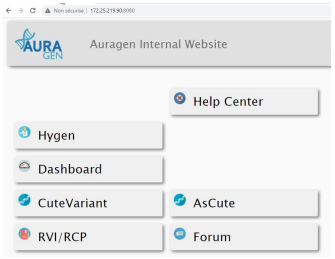
Dépôt de la demande

RCP d'entrée

Validation de l'indication

Envoi des échantillons

- Impression feuille de transfert, étiquettes et ordonnance
- Prélèvement du trio/duo
- Envoi des prélèvements via Biologistic (Connection via compte extranet)



Analyses Bio-informatiques : .vcf annoté

Production des fichiers .vcf

Séquençage

- Réception
- Validation des échantillons
- Préparation des ADN, qualification, préparation de la librairie
- Séquençage WGS 30X

Interprétation des résultats Génération d'un compte rendu

Analyse génétique

Réunion de validation de l'analyse génétique

Revue et discussions des résultats

Récupération des résultats

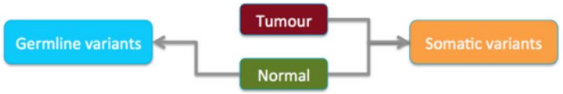
Patient

Restitution du compte rendu disponible sous HYGEN

HYGEN

Analyses KC, MR & OC

Cancer



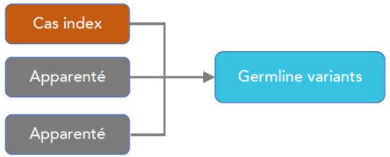
WGS tumeur 60-90x
normal 30-55x

WES tumeur 300x
normal 150x

RNAseq tumeur >64 Mp
(fusions)

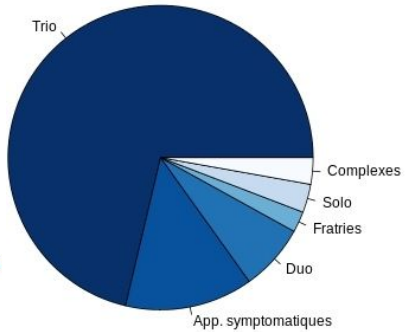
Analyses somatiques

Maladies Rares & Oncogénétique



WGS 30-55x

Un ou plusieurs échantillons
au sein d'une famille



Analyses germinales

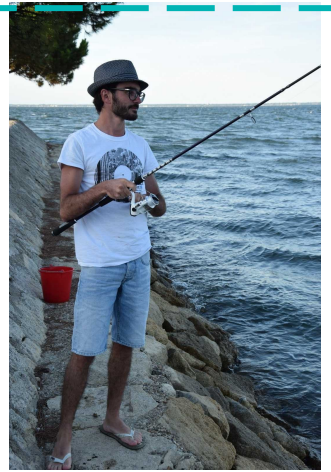
Organisation équipes bioinfo Auragen
CLB (Lyon) -& INRIA, CHUGA (Grenoble)

Alain Viari - responsable SI
Jean François Scariot - ingénieur - infra
Maëlle Martinet Gerphagnon - ingénieur info

Anthony Ferrari - responsable bioinfo KC
Anne Sophie Sertier - bioinfo dev
Elise Dugat ingénieurs - bioinfo prod

Julien Thevenon & Virginie Bernard - responsables bioinfo MR
Quentin Charret & Clément Lionnet - ingénieurs bioinfo prod & dev

\+ internes en génétiques
Alexandre, Laury, Bertrand, Benjamin, Alexis …



Développement et industrialisation

Identification et spécification d'un besoin
→ cahier des charges

1. Revue de littérature
2. Identification d'un gold standard et/ou de jeux de données internes pour l'évaluation
3. Comparaison de méthodes et sélection de la plus appropriée, customisations
4. Création et intégration d'un pipeline en suivant les recommandations de l'INCa
5. Validation du pipeline sur jeux de données internes et gold standard
6. Industrialisation et mise en production versionnée

GENOME IN A BOTTLE

Our mission is to provide the authoritative characterization of human genomes.

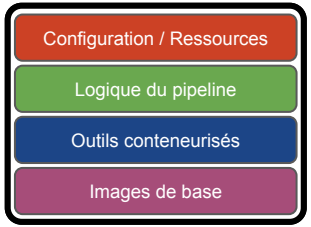
nextflow

docker

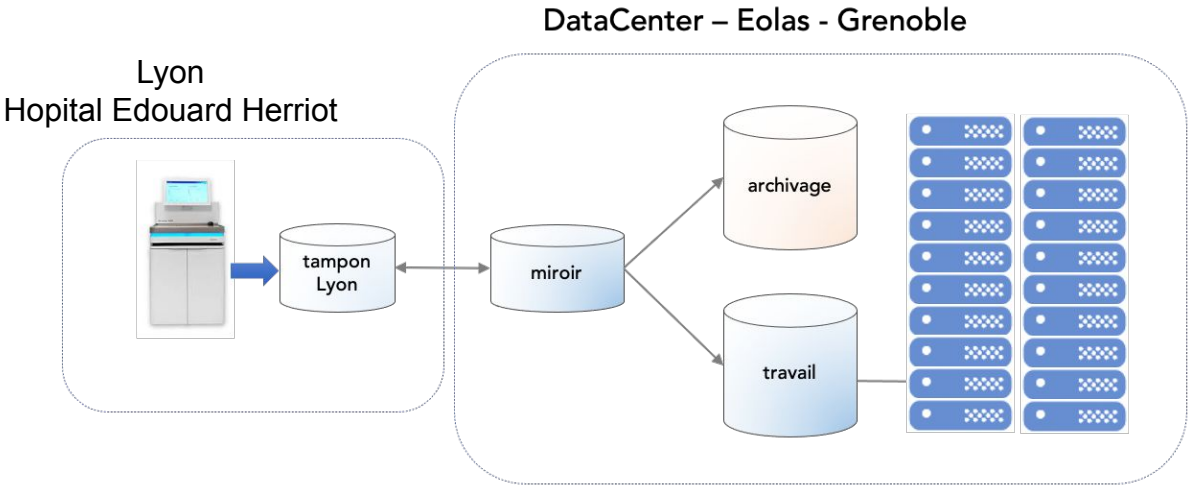
- ✓ Traçabilité
- ✓ Reproductibilité
- ✓ Fiabilité

slurm workload manager

GitLab



Quelques chiffres



- 20 noeuds de calcul
- 1500 CPUs
- 13 To RAM
- 2.5 Po stockage

Configuration HPC

Hébergement HDS

Maintenance

Entre Sept 2019 et Mars 2021 : ~130 ans de calcul

Schéma simplifié de l'infrastructure

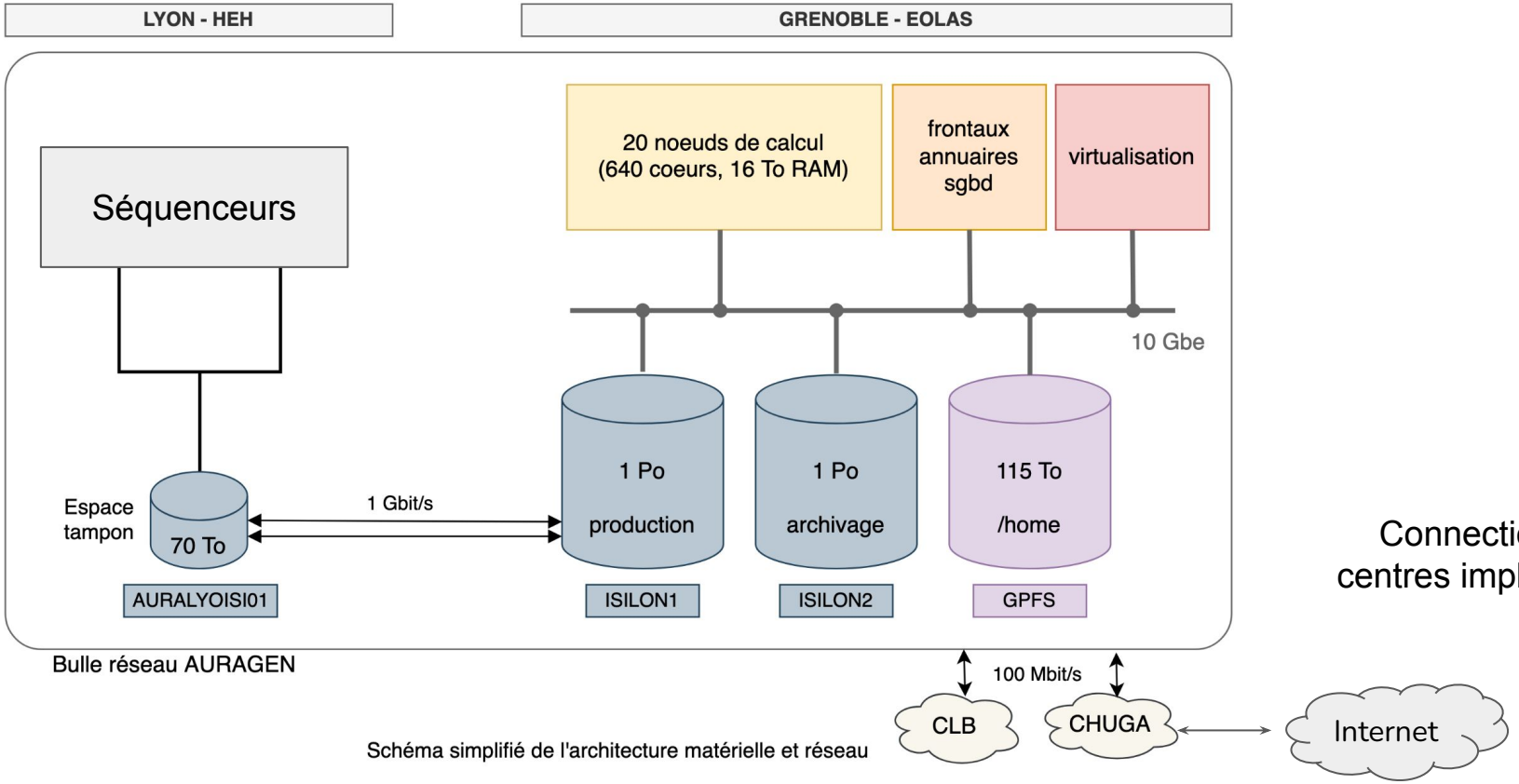
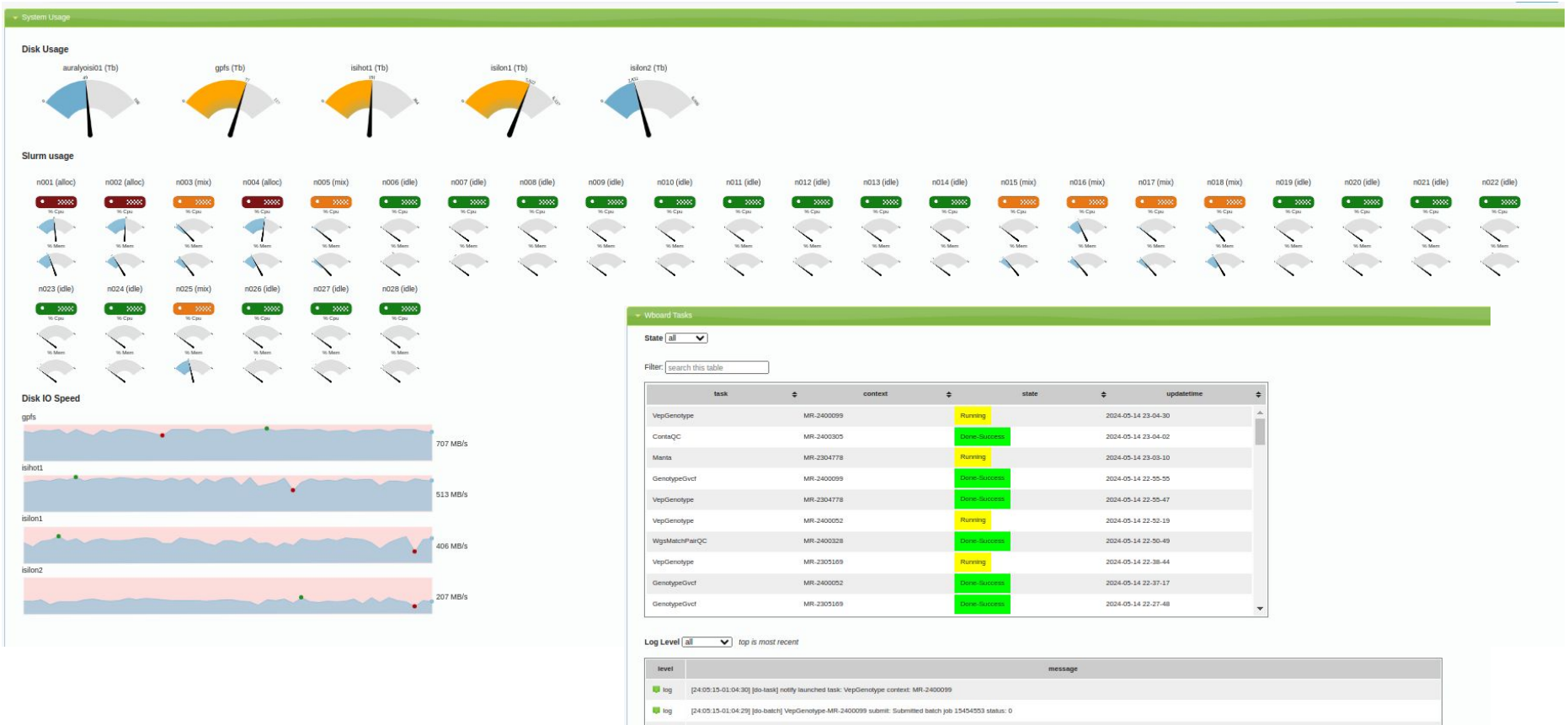


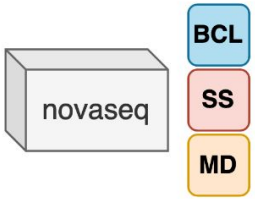
Schéma simplifié de l'architecture matérielle et réseau

Connection avec les centres impliqués dans AURAGEN

Dashboard de suivi



Fichiers source



Fichier brut produit lors du séquençage → conduit à un fichier FASTQ

Sample Sheet → informations librairies qui permet entre autre d'associer un identifiant de prélèvement à ses lectures

Métadonnées → informations cliniques sur le prélèvement, le dossier

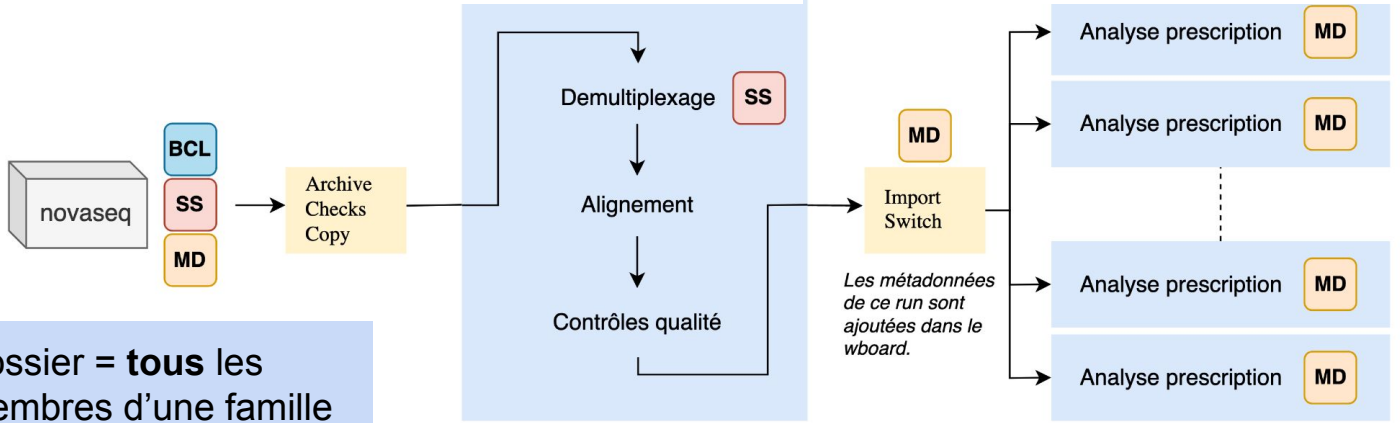
SS : Samplesheet émis par le SGL
MD : Métadonnés associées à la prescription. Transmises par le SGL et l'outil de prescription

Schéma simplifié des processus

Séquençage des échantillons dans le flux de leur réception.
Traitement bioinformatique une fois les dossiers complets avec QC conformes

Mode run
→ quand un séquençage est terminé

Mode patient
→ quand un dossier est complet (**en plusieurs runs parfois**) et à contrôle qualité conforme



Dossier = **tous** les membres d'une famille

Les métadonnées de ce run sont ajoutées dans le wboard.

Tous les échantilons présents dans le run sont traités de façon identique.

Quand tous les échantillons (pouvant provenir de runs différents) associés à une prescription sont finalement disponibles dans le wboard, l'analyse en mode patient peut démarrer.

SS : Samplesheet émis par le SGL
MD : Métadonnés associées à la prescription. Transmises par le SGL et l'outil de prescription

Routine Bioinformatique de la production AURAGEN

Informations sur les librairies

HYGEN

Informations cliniques

samplesheet

metadata cliniques

Fichier de séquences → Démultiplexage → Alignement → Appel de variations → Annotation → Synthèse et interprétation biologique

Génome de référence GRCh38

Données téléchargées
Version des annotations en local (maîtrise des versions)

Réunion de validation et d'interprétation (RVI)
Réunion Clinico-Biologique (RCB)

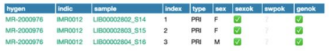
Contrôles qualité à chaque étape

Les contrôles d'identito vigilance requis "conformes" pour une interprétation

Quels contrôles de l'identito vigilance sont mis en place en post séquençage?

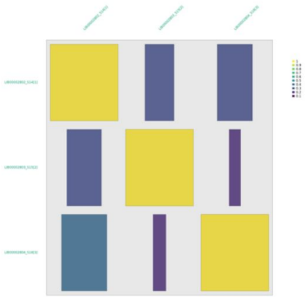
Les informations patient sont elles conformes à la prescription?

Le genre observé est il bien conforme à la prescription ?
→ratio profondeur chrY chrX



Genre conforme

Similarité entre prélèvements conforme à la prescription?
→ 100000 SNP chr1 à 22



Similarité entre prélèvements conforme

Est ce que l'échantillon est contaminé? La contamination fausse-t-elle les résultats ?

→ 2589 positions ayant 3 allèles possibles dans la population
Présence effective de plus de 2 allèles ?
Si contamination, taux impactant les résultats ?

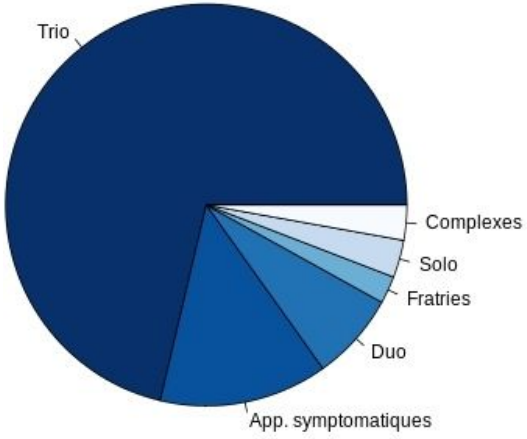
Pas de contamination = conforme
Contamination sans impact sur les variations = conforme

Ce qui est présenté pour interprétation a été validé par tous les contrôles mis en place

Priorisation & interface de tri des variations

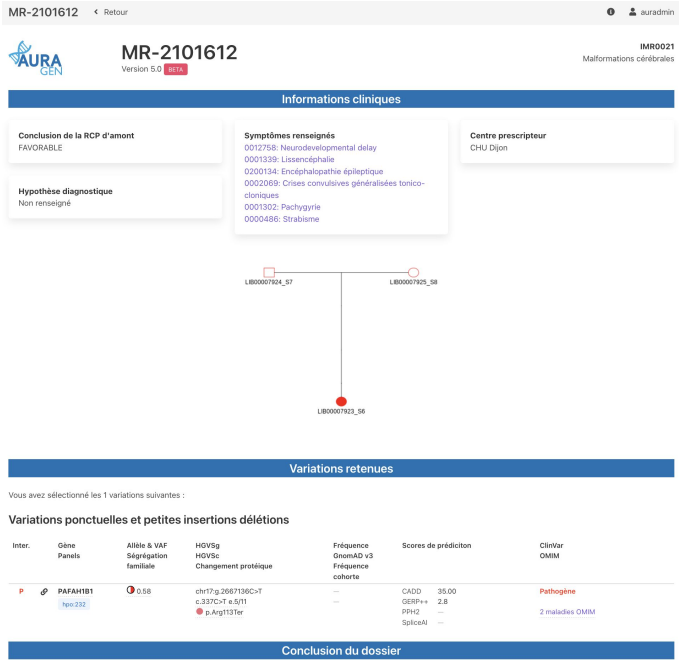
Pipelines adaptés aux structures familiales

→ couverture progressive de chaque structure (en cours)



Rapport de synthèse automatisé interactif

~ 100/120 dossiers par semaine



Interface de tri de variants SeqOne

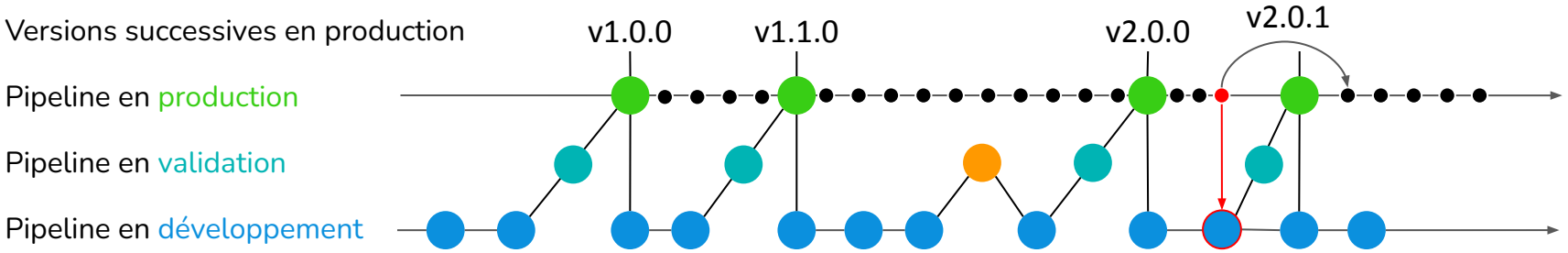
VCF auragen annoté envoyé (SNV & petites indels)

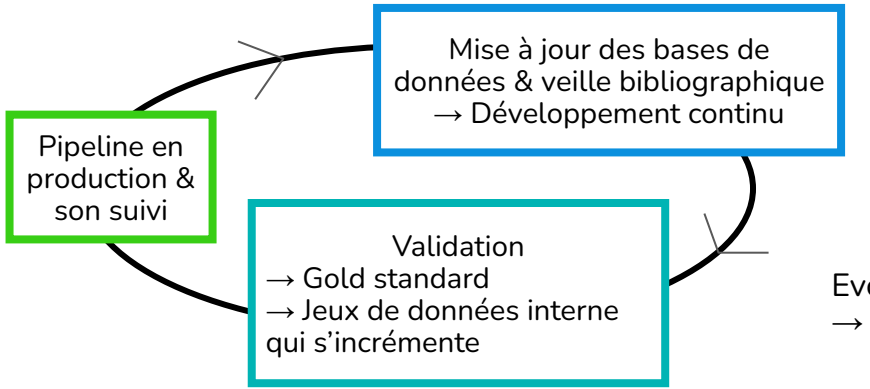
SeqOne “Auragen”

SeqOne GENOMICS

Amélioration continue des rapports

- Analyse d'un dossier en succès
- Analyse d'un dossier en erreur

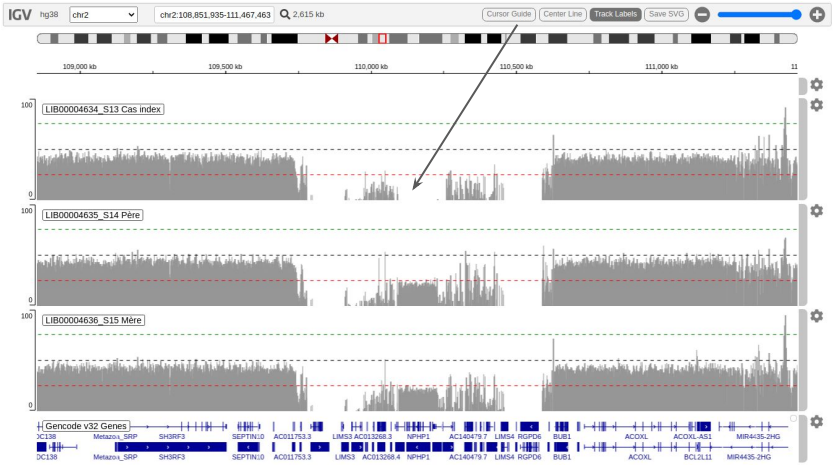




Evolution des pipelines
→ amélioration continue

Gestion des anomalies informatique / bioinformatique **critiques**
→ délai à respecter imposé par les contraintes de la routine

Délétion homozygote, héritée des parents hétérozygotes
chr2 : dossier MR-2000696 → région de faible mappabilité



Cible *codant* à moins de 10 lectures

Les 6 gènes de la filière ayant une couverture < 10 lectures pour une partie des régions codantes +/-50 bases.

Gène	Pourcentage de la cible codant des gènes à 20 lectures	Pourcentage de la cible codant des gènes à 10 lectures
HGSNAT	97.43	99.27
INPP5E	98.56	98.87
NPHP1	0.00	0.00
OPN1LW	66.19	71.68
OPN1MW	6.08	35.28
RPGR	92.01	99.07

Pour ces gènes, nous listons ici tous les défauts impactant la séquence codante. Concernant les régions introniques, nous listons les défauts à moins de 50 bases d'un exon, avec une profondeur entre 0 à 8 lectures sur un intervalle de taille supérieur à 5 bases sont détaillés. Une absence de tableau est possible.

Intervalles de la cible "codant" à moins de 10 lectures et non récurrents

Gène	Nombre de bases < 10 lectures	Nombre de lectures sur l'intervalle (max)	Distance à l'exon (bases)	Intervalle du défaut
NPHP1	137266	[0,9[	0	chr2 :110092811-110230076

<u>ABSENT dans les données brutes:</u>

CNVnator recherche une rupture de profondeur
→ en amont pas de lecture (région de faible mappabilité)
→ del homo pas de lecture
→ en aval pas de lecture (région de faible mappabilité)
= pas de rupture
Vu par interpréteurs via les contrôles qualité ciblés

→ bilans de couverture des gènes d'intérêt

Premier "contrôle positif" pour mettre en place la **nouvelle approche de détection des délétions homozygotes** (cf. slides suivantes)

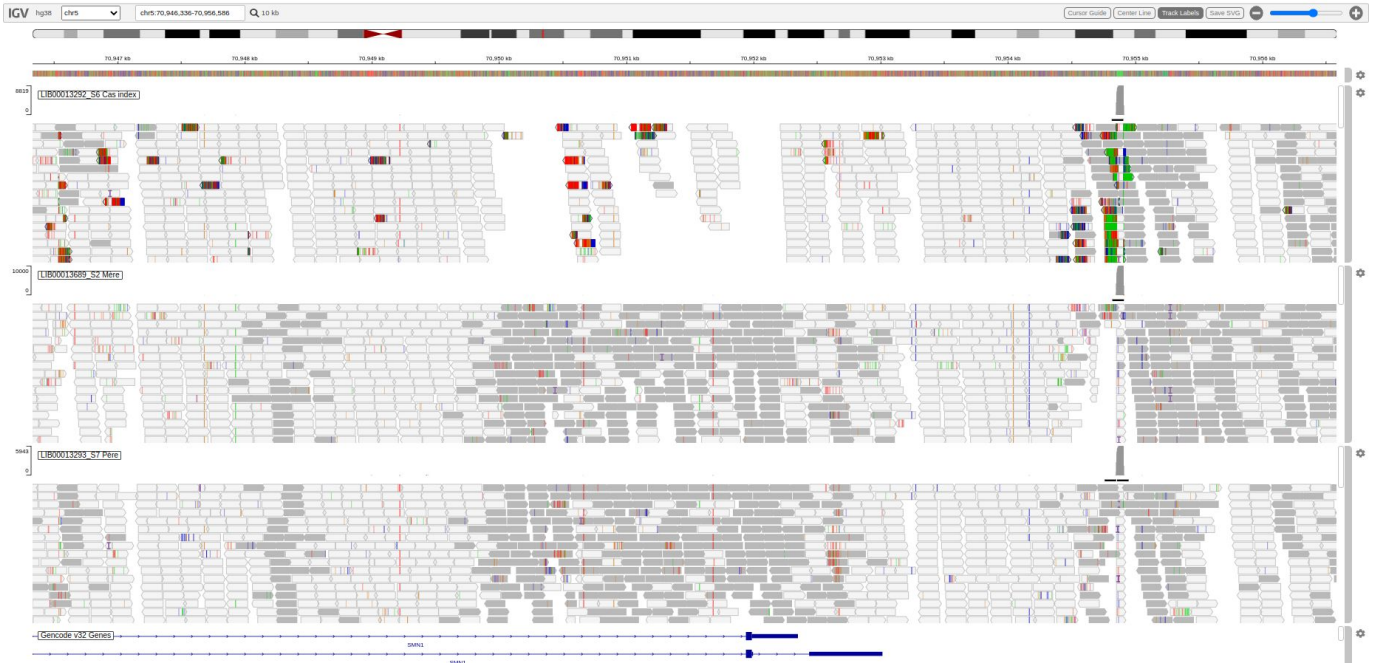
Délétion homozygote & diags rattrapés

→ Mise en place d'un **appel des "délétions homo"** dans les régions à faible mappabilité
Reads blancs : reads de faible mappabilité **SMN1**

CI

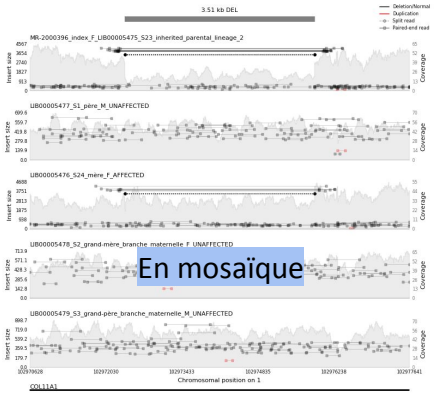
Père

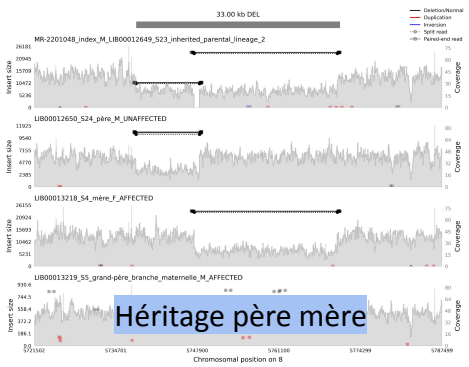
Mère

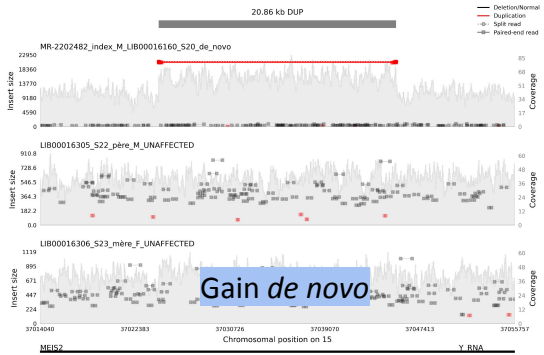


Visualisation des CNV

CI
Père
Mère







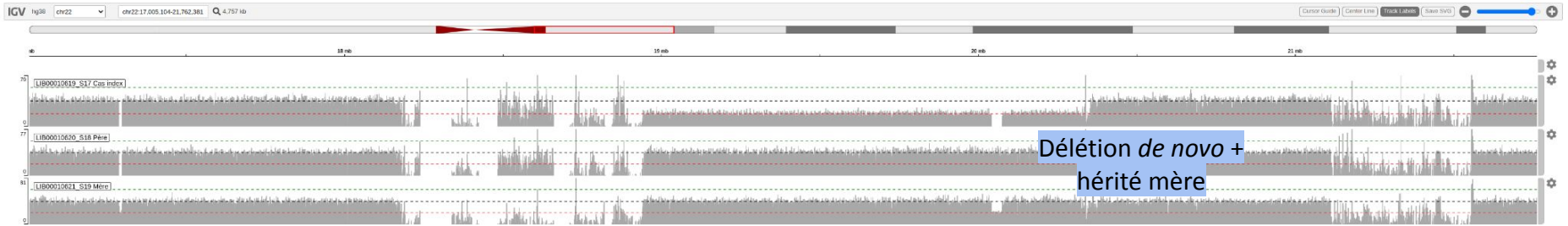


Illustration SNP array like pour le cas index - RoH

Régions d'homozygotie

Les régions d'homozygotie d'une taille supérieure à 2Mb sont représentées dans le tableau ci-dessous.

Intervalle	Taille (en Mb)
chr7:6840356-29640320	22.80
chr7:35946191-56339116	20.39
chr7:65831419-72962554	7.13
chr7:77058056-96745610	19.69

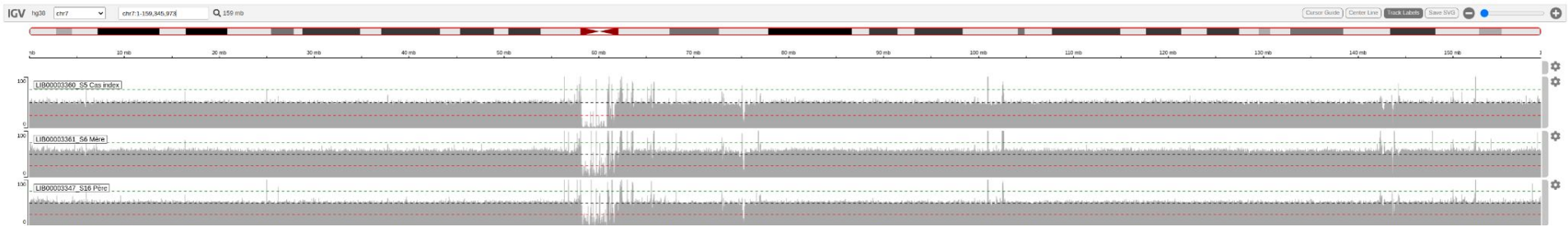


Illustration SNP array like pour le cas index - RoH

Régions d'homozygotie

Les régions d'homozygotie d'une taille supérieure à 2Mb sont représentées dans le tableau ci-dessous.

Intervalle	Taille (en Mb)
chr7:6840356-29640320	22.80
chr7:35946191-56339116	20.39
chr7:65831419-72962554	7.13
chr7:77058056-96745610	19.69

B-allèle frequency

Exploitation de ~1 million de polymorphismes fréquents sur le génome
Critères requis : 20 lectures
BAF entre 0 (ref) et 1 (homozygote), centrée sur 0.5 (hétérozygote)

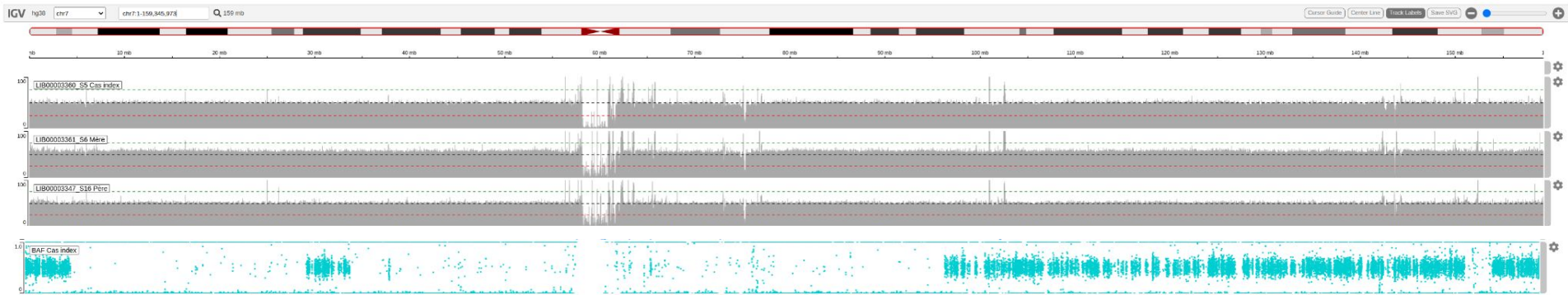
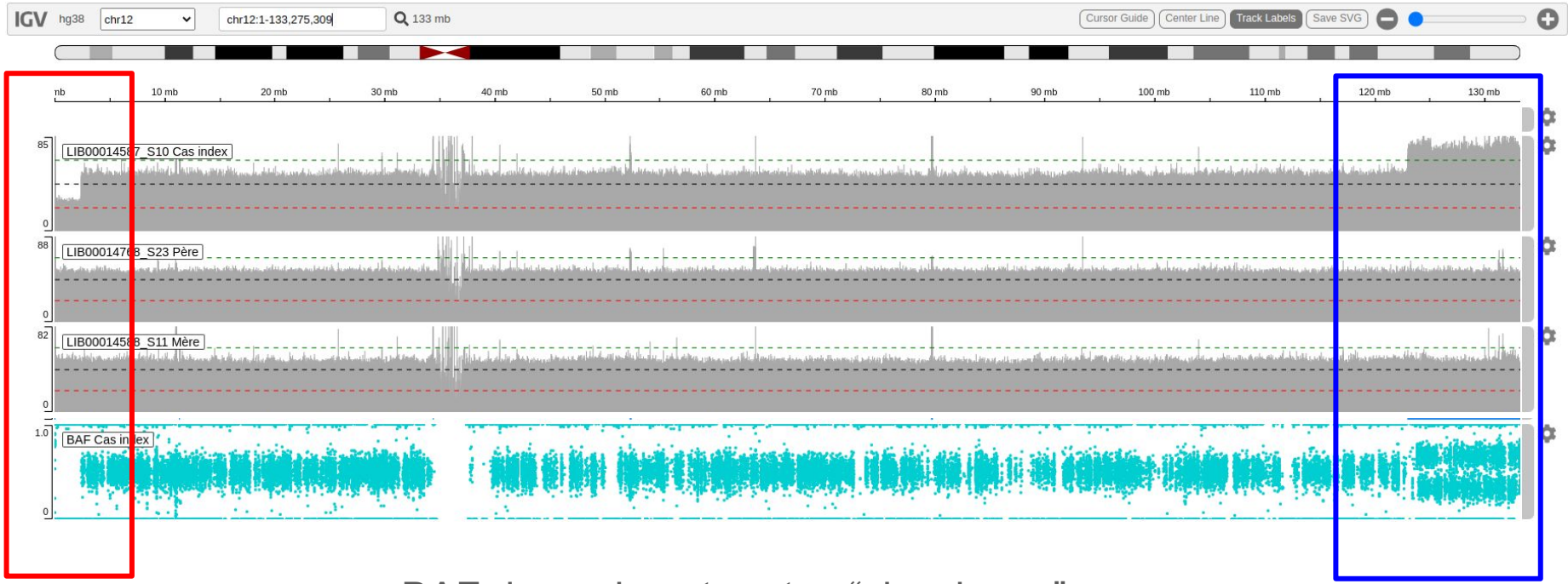


Illustration SNP array like pour le cas index - CNV



BAF des gains et pertes “classiques”

Illustration SNP array like pour le cas index - mosaïque

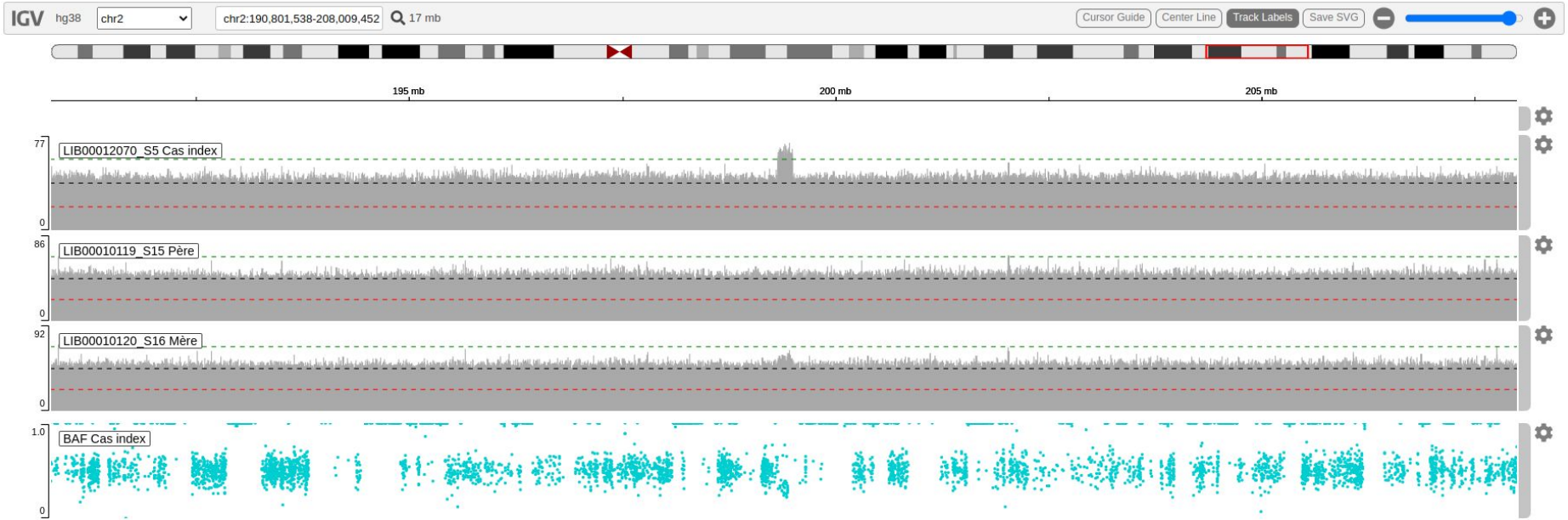


Illustration SNP array like
Pour aller plus loin...

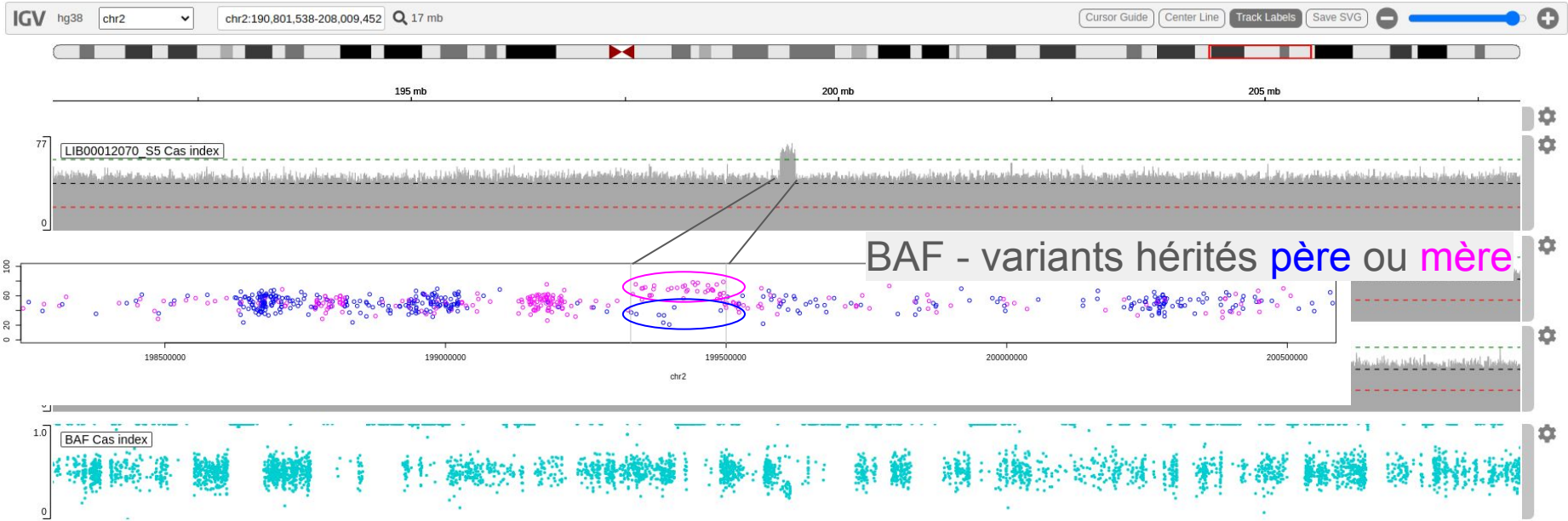


Illustration SNP array like
Pour aller plus loin...

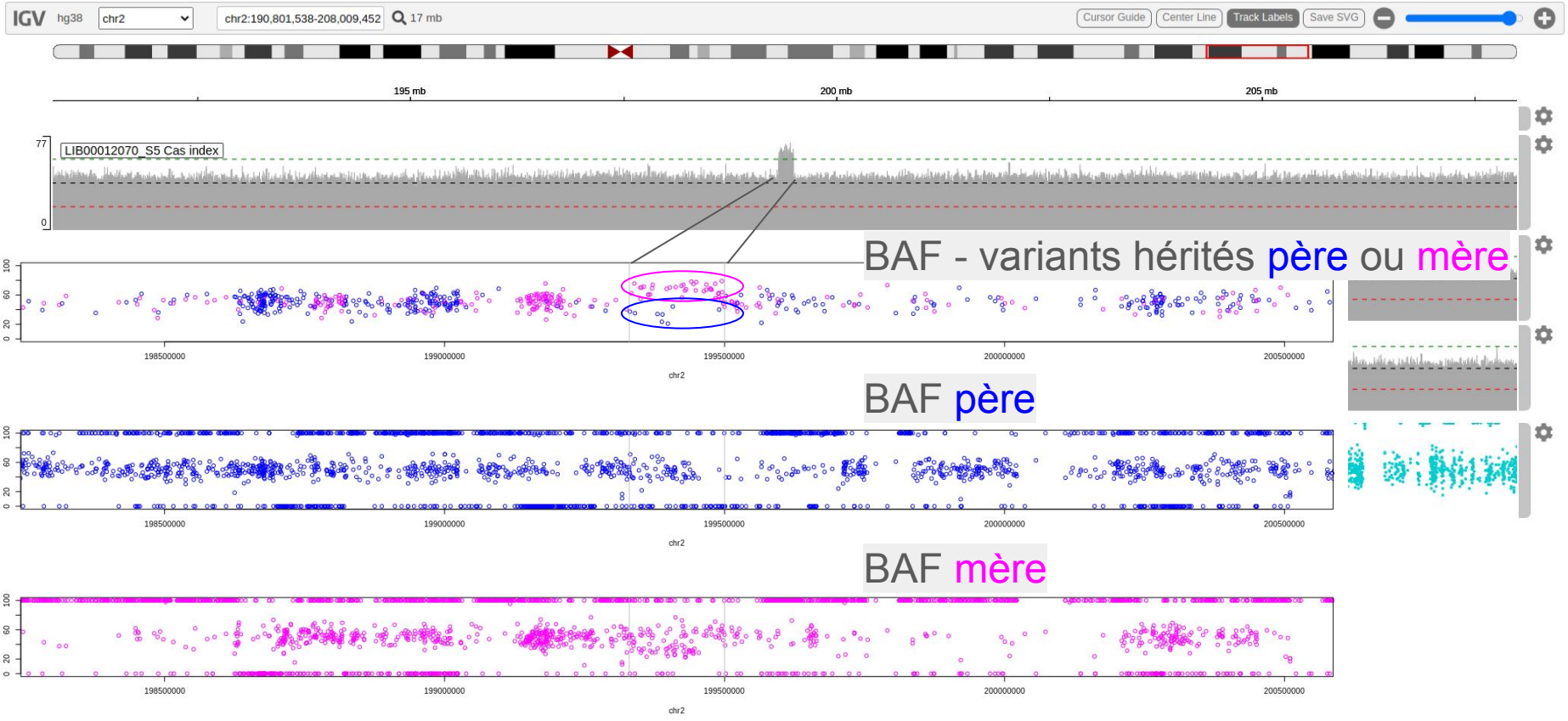
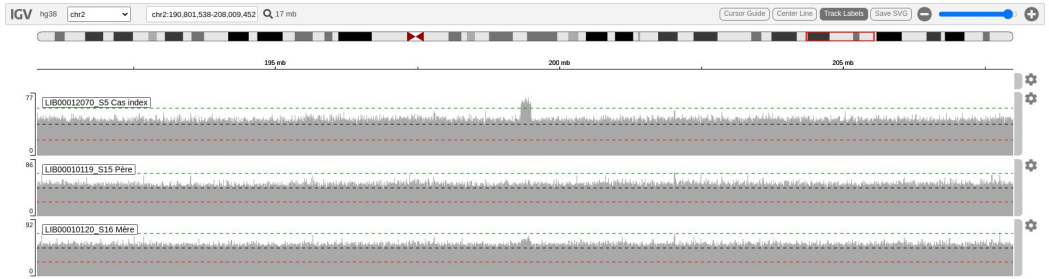
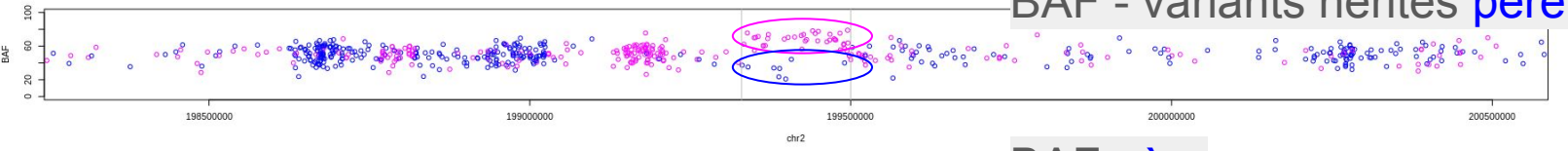


Illustration SNP array like
Pour aller plus loin...

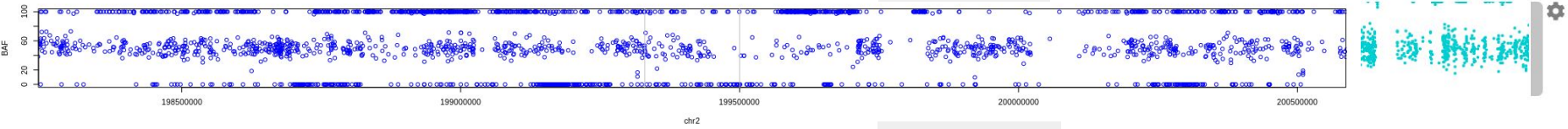


Visible sur le BAM ✓
Renforcé par profils BAF ✓
Confirmé par MLPA ✓

BAF - variants hérités père ou mère



BAF père



BAF mère

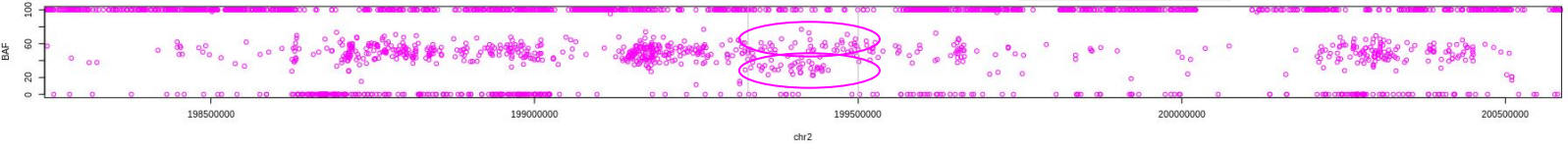
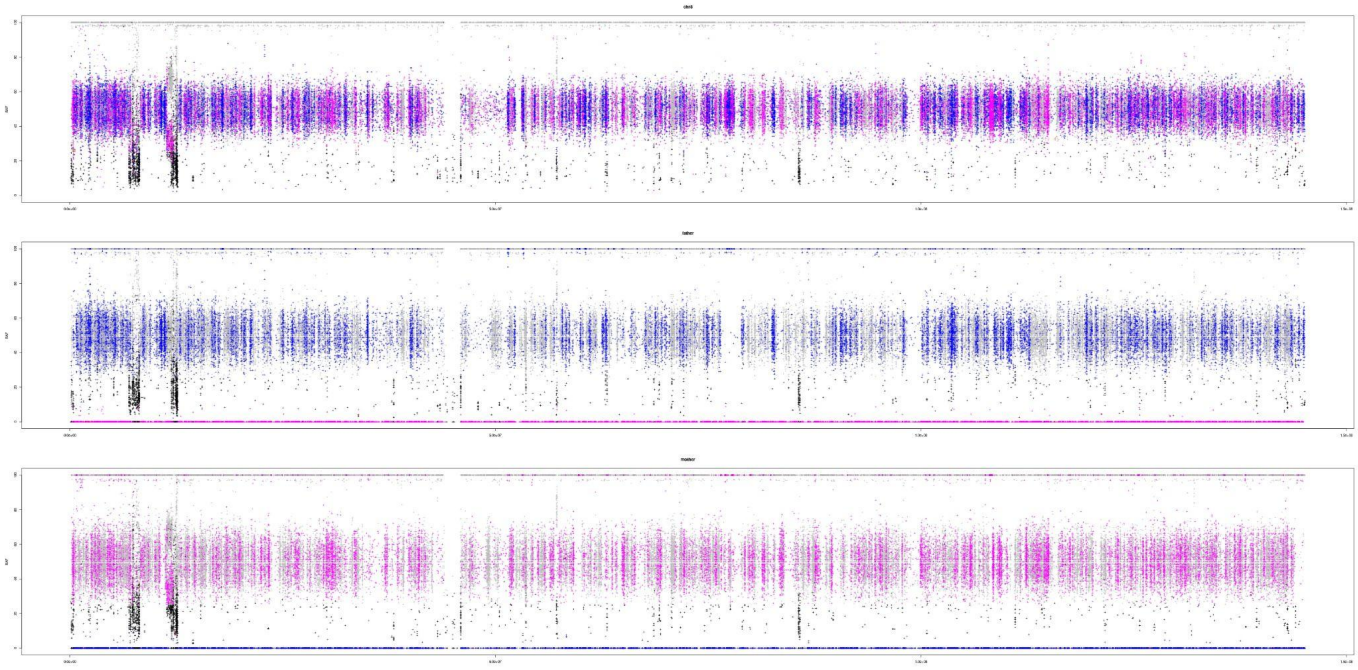


Illustration SNP array like & héritage père mère - profil standard

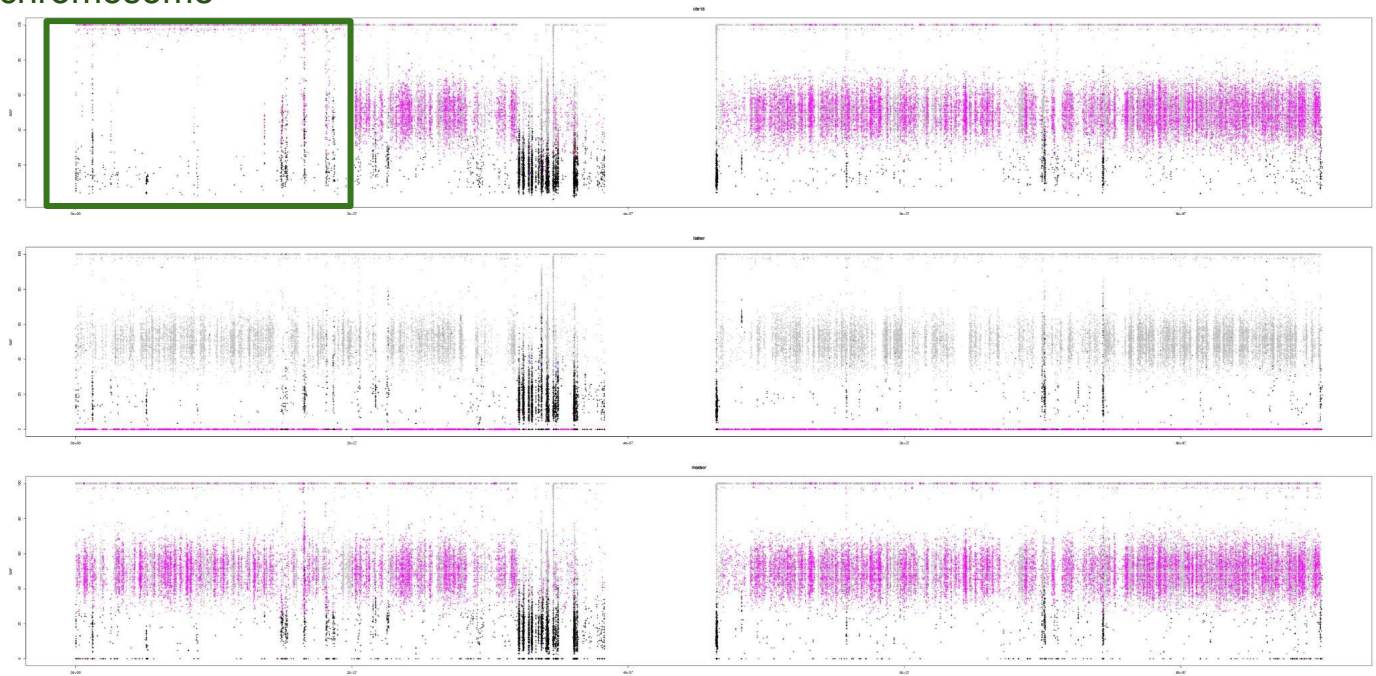


WORK IN PROGRESS

- SNP hérité du père
- SNP hérité de la mère

Illustration SNP array like & héritage père mère - profil UPD

Région d'isodisomie : Origine des SNP : un seul chromosome **maternel**

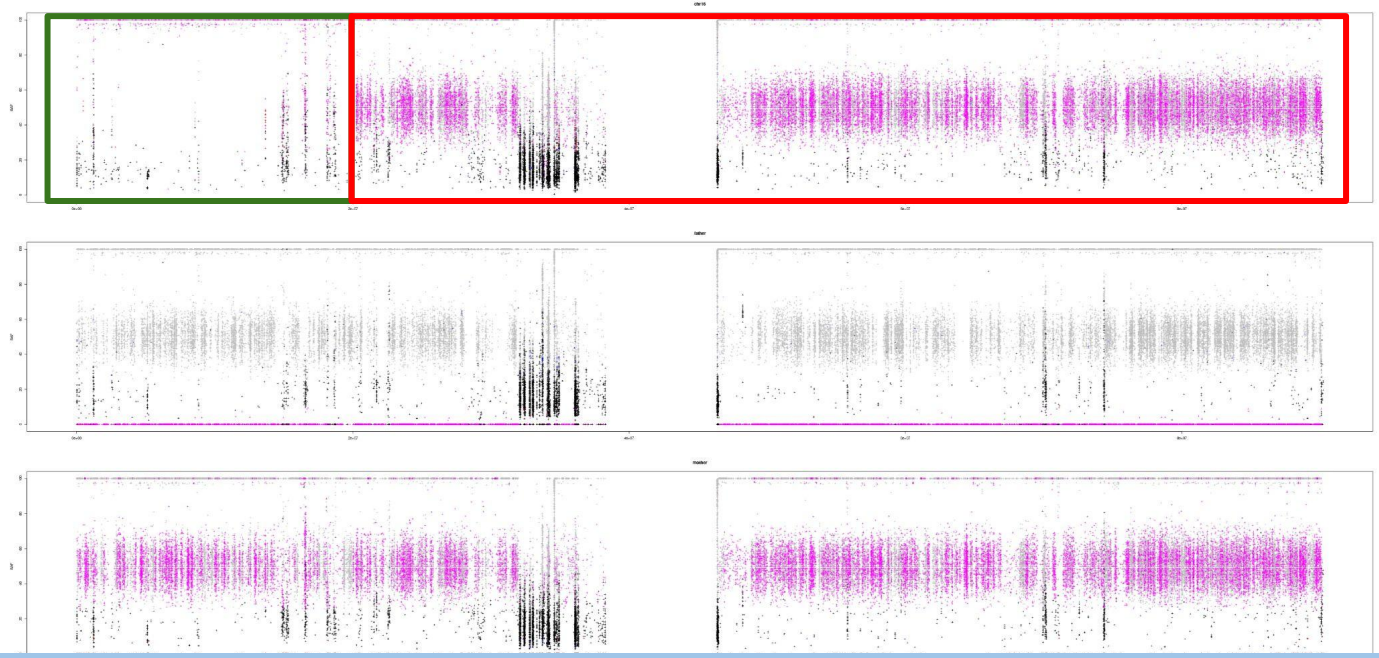


WORK IN PROGRESS

Illustration SNP array like & héritage père mère - profil UPD

Région d'isodisomie : Origine des SNP : un seul chromosome maternel

Région d'hétérodisomie : Origine des SNP : les deux chromosomes **maternels**



WORK IN PROGRESS

=> Disomie uniparentale de tout le chr16 : UPD(16) Avec isodisomie 16pter et hétérodisomie sur le reste du chr

Illustration SNP array like
Quoi prioriser ?

WORK IN PROGRESS

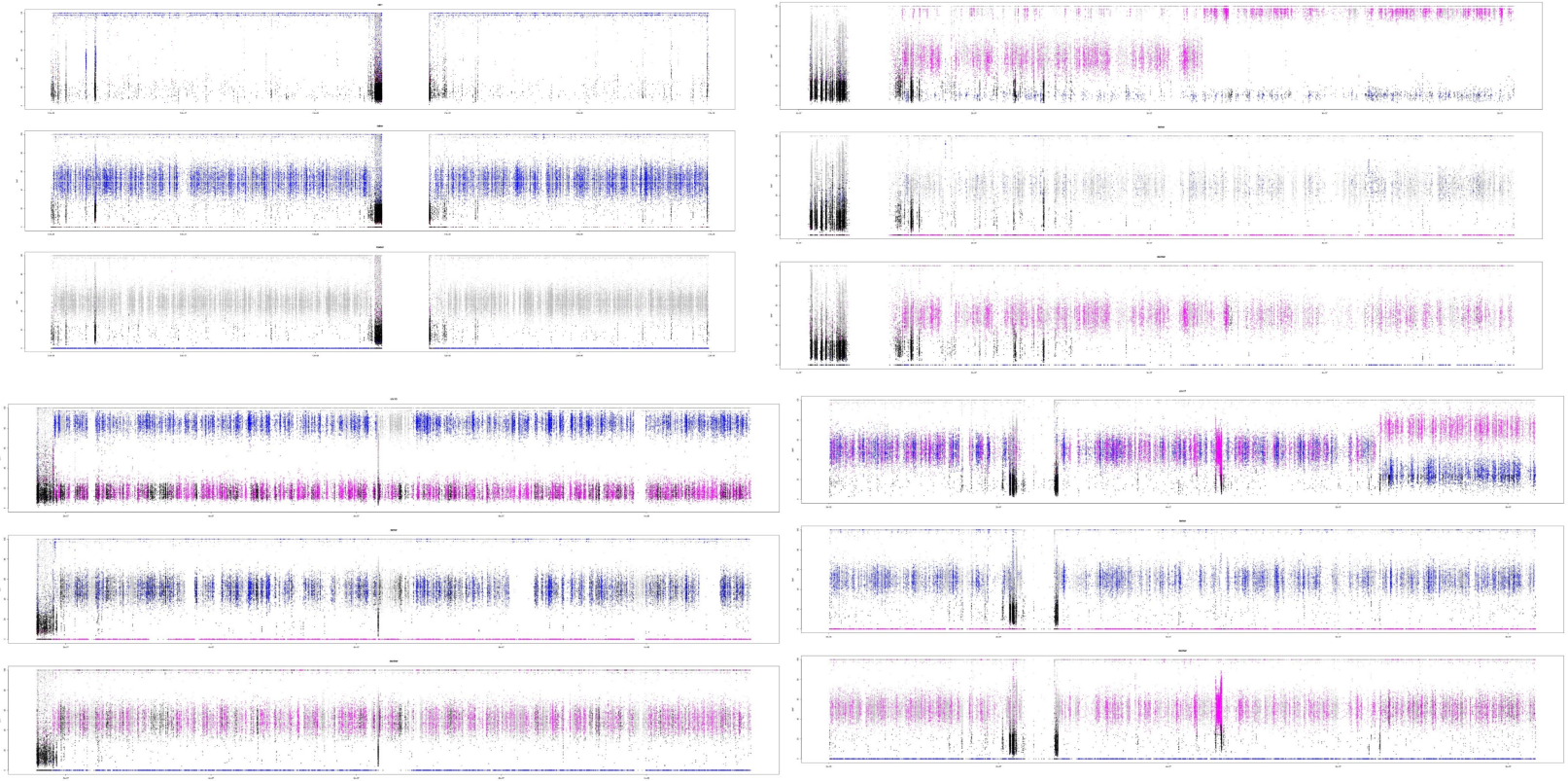
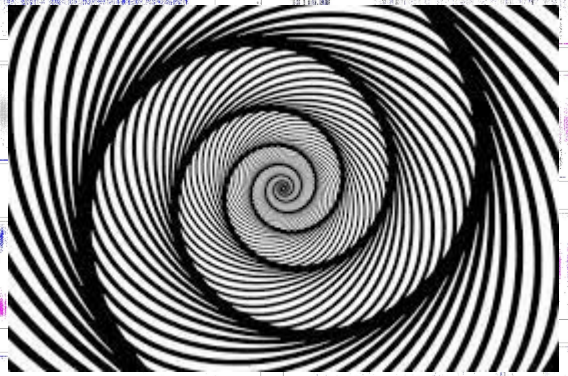


Illustration SNP array like
Quoi prioriser ?

WORK IN PROGRESS



nécessite calling spécifique, filtré sur chromosome soumis à empreinte ?

CCL & perspectives

Mise en place wet & dry Mise en place dry & bio
Importance MAJEURE pour permettre les améliorations

Livraisons pour interprétation dans le flux
→ automatisation de la prod

Confiance!

La suite ?
Auragen : todo liste des améliorations encore grande :)
⇒ Objectif diag!

Collecteur et Analyseur de Données
→ Projets soumis pour les améliorations de demain

WORK IN PROGRESS

CAD
FRANCE MÉDECINE GÉNOMIQUE 2025

Merci!