



**HAL**  
open science

# Étude des performances d'une technologie de séquençage haut débit pangénomique pour la détection des copy number variations : comparaison avec l'analyse chromosomique sur puce à ADN

Mathilde Pujalte

## ► To cite this version:

Mathilde Pujalte. Étude des performances d'une technologie de séquençage haut débit pangénomique pour la détection des copy number variations : comparaison avec l'analyse chromosomique sur puce à ADN. Médecine humaine et pathologie. 2020. dumas-02886468

**HAL Id: dumas-02886468**

**<https://dumas.ccsd.cnrs.fr/dumas-02886468>**

Submitted on 1 Jul 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITE DE PICARDIE JULES VERNE  
FACULTE DE MEDECINE D'AMIENS  
ANNEE 2020



THESE 2020-25

THESE POUR LE DIPLOME D'ETAT DE DOCTEUR EN MEDECINE  
SPECIALITE BIOLOGIE MEDICALE

**ETUDE DES PERFORMANCES D'UNE TECHNOLOGIE DE  
SEQUENCAGE HAUT DEBIT PANGENOMIQUE POUR LA  
DETECTION DES COPY NUMBER VARIATIONS : COMPARAISON  
AVEC L'ANALYSE CHROMOSOMIQUE SUR PUCE A ADN**

Présentée et soutenue publiquement  
le vendredi 10 avril 2020  
par  
Mathilde PUJALTE

PRESIDENT DU JURY :  
Monsieur le Professeur Henri COPIN

MEMBRES DU JURY :  
Monsieur le Professeur Gilles DUVERLIE  
Monsieur le Professeur Patrick BERQUIN  
Madame la Maître de Conférence Estelle CADET  
Madame le Docteur Noémie CELTON

DIRECTEUR DE THESE :  
Monsieur le Docteur Guillaume JEDRASZAK



## REMERCIEMENTS

*Monsieur le Professeur Henri COPIN*

*Professeur des Universités – Praticien Hospitalier*

*Histologie-Embryologie-Cytogénétique*

*Chef du Service de Médecine et Biologie de la Reproduction et CECOS de Picardie*

*Pôle « Femme-Couple-Enfant »*

*Responsable de l'UF de Cytogénétique du Laboratoire de Génétique Constitutionnelle*

*Pôle « Biologie, Pharmacie et Santé des Populations »*

*Chevalier dans l'Ordre des Palmes Académiques*

Monsieur le Professeur,

Je vous remercie de me faire l'honneur de présider mon jury de thèse.

Merci pour votre accueil chaleureux et bienveillant au sein du service de cytogénétique ainsi que pour le soutien que vous me témoignez au quotidien.

Je garde en mémoire vos enseignements divers et variés, tant dans les domaines de la génétique que de l'Histoire de France.

Soyez assuré de mon sincère respect et de ma profonde reconnaissance.

***Monsieur le Professeur Gilles DUVERLIE***

*Professeur des Universités-Praticien Hospitalier (Virologie, biologie)*

*Chef du laboratoire de Virologie*

*Pôle « Biologie, Pharmacie et Santé des Populations »*

*Chevalier dans l'Ordre des Palmes Académiques*

Monsieur le Professeur,

Je vous remercie pour l'honneur que vous m'avez fait en acceptant de faire partie de mon jury de thèse. Merci pour votre disponibilité tout au long de mon internat.

Soyez assuré de mon profond respect et de mes plus sincères remerciements.

***Monsieur le Professeur Patrick BERQUIN***

*Professeur des Universités-Praticien Hospitalier (Pédiatrie)*

*Responsable du centre d'activité "Neurologie pédiatrique"*

*Chef du Pôle « Femme-Couple-Enfant »*

Monsieur le Professeur,

Je vous remercie pour l'honneur que vous m'avez fait en acceptant de faire partie de mon jury de thèse. J'espère que ce travail saura retenir votre intérêt.

Soyez assuré de mon profond respect et de mes sincères remerciements.

***Madame le Docteur Estelle CADET***

*Maître de Conférences des Universités – Praticien Hospitalier (Génétique)*

Je te remercie de me faire l'honneur de faire partie de mon jury de thèse.

Merci pour ton accueil au sein du service de génétique moléculaire et pour tes enseignements au cours du semestre.

Sois assurée de mon profond respect et de mes sincères remerciements.

***Madame le Docteur Noémie CELTON***

*Praticien Hospitalier (Biologie Médicale, spécialité Génétique)*

Chère Noémie,

Je te remercie de me faire l'honneur d'être présente dans mon jury aujourd'hui, j'en suis très heureuse. J'espère que cette thèse (malgré le blabla...) sera à la hauteur de tes attentes.

Merci pour ta bienveillance à mon égard, tes conseils avisés et ton soutien infailible.

Tu m'as appris énormément de choses tant professionnellement qu'humainement, j'ai de la chance d'avoir fait mes premiers pas en génétique avec toi.

Sois assurée de mon plus sincère respect et de ma profonde reconnaissance.



***Monsieur le Docteur Guillaume JEDRASZAK***  
*Praticien Hospitalier Universitaire (Génétique Médicale)*

Cher Guillaume,

Je te remercie pour l'honneur que tu m'as fait d'avoir accepté d'encadrer ma thèse, pour tes précieux conseils et ton investissement. J'espère que ce travail est à la hauteur de tes espérances.

Je tenais également à te remercier pour ta grande disponibilité et ta bienveillance à mon égard. J'ai appris énormément de choses grâce à toi et je suis heureuse d'avoir la chance de travailler avec toi dans le futur.

Sois assuré de mon plus sincère respect et de ma profonde reconnaissance.

***L'équipe de Cytogénétique Constitutionnelle et Onco-Hématologique du CHU d'Amiens :***

Merci à vous tous pour cette année et demie passée à vos côtés. Vous m'avez intégré comme l'une des vôtres et je vous en suis extrêmement reconnaissante.

Votre bonne humeur est communicative et c'est grâce à vous que je viens travailler chaque matin avec grand plaisir !

***Florence et Catherine :*** merci pour la bienveillance avec laquelle vous m'avez guidé lors de mes débuts maladroits en cytogénétique. Vous m'avez non seulement appris toutes les techniques conventionnelles mais également toutes les astuces pour reconnaître les chromosomes comme une pro. Merci pour la patience et la gentillesse avec lesquelles vous avez répondu à toutes mes questions et corrigé mes (nombreuses) erreurs de classement.

***Fred, Marie et Nathalie :*** merci de m'avoir appris toutes les bases de la biologie moléculaire, du maniement d'une pipette jusqu'au NGS. Grâce à vous, je n'ai aucune excuse lorsque je rate une technique de bio mol et si un jour mon blanc est contaminé, je sais que vous viendrez me tirer les oreilles ! Un grand merci pour votre bonne humeur au quotidien.

Nathalie, je te suis très reconnaissante de ton investissement dans ce travail ! Malgré les (quelques) embuches de cette dernière série de OneSeq, tout s'est finalement bien passé, merci encore !

***Mélanie, Annelise et Francis :*** merci pour votre accueil chaleureux lors de ma semaine de découverte de l'oncohématologie. Même si je n'ai pas été prise de passion pour les petits chromosomes, vous m'avez presque convaincu de venir par votre enthousiasme tant personnel que professionnel. Votre spontanéité, votre joie de vivre et les phrases (plus ou moins) philosophiques de Francis mettent des paillettes dans la vie du service !

***Sophie Bury :*** merci pour ton dynamisme, ta gentillesse et l'aide précieuse que tu nous apporte au quotidien.

***Dr Marlène Gallet :*** merci pour ta bonne humeur, ta gentillesse à mon égard et ton soutien concernant le master 2 et la thèse. Tu m'as appris plein de choses sur le design des amorces et je t'en suis reconnaissante.

***Dr Yann Ferret :*** je tenais à te remercier pour ta bonne humeur, ton enthousiasme (au travail et à l'extérieur) et tes encouragements. Sache que je me surprends même à te trouver drôle... parfois !

***L'équipe du secrétariat de BDR et de Cytogénétique :***

Audrey, Laure et Nathalie : merci pour votre bonne humeur et votre dynamisme, le secrétariat ne serait pas pareil sans vous.

***L'équipe de Génétique Moléculaire du CHU d'Amiens :***

***Anne-Sophie, Catherine, Didier, Luigina, Patricia :*** merci pour votre accueil au sein du service. Même si le semestre a été chargé, j'ai appris beaucoup de choses, merci de m'avoir montré et expliqué les différentes techniques de génétique moléculaire.

Didier, je tenais à te remercier pour ta pédagogie, ta gentillesse et le temps que tu m'as consacré. Tes diverses explications m'ont beaucoup apporté.

***Dr Virginie Magry :*** tu m'as appris et montré beaucoup de choses sur l'interprétation des variants et je t'en suis reconnaissante.

***A mes co-internes de Biologie Médicale Amiénois :***

Merci pour l'ambiance incroyable entre internes, je n'aurais jamais pensé que l'internat pourrait être aussi génial. Je garde des souvenirs incroyables de mes stages et de mes gardes à vos côtés. Les multiples fous rires, les sushis en garde, le billard du midi, les heures de discussions sur la « banquette des biologistes », Colette Lebrochet, les anniversaires à Buffalo Grill, les courses de trottinettes, les cinémas, les blagues, les inondations en garde, les soirées Miss France, les journées à la mer, les week-ends, St Leu...

Je ne peux pas citer tout le monde mais j'ai une pensée particulière pour mes super co-internes et amis ***Julie, Sophie, Cécile, Elodie, Valentin, Stacy, Arthur, Ben, Geoffrey, Manon, Julien, Cédric, Laurine, Mehdi et Quentin.***

A mes amis ***Julien, Laure A, Laure S, Antoine :*** merci d'avoir été là pendant ces années d'externat. Merci pour les fous-rires, les soirées nems, les voyages, la bonne humeur, les soirées ciné... et tellement plus encore.

***A mes parents, à mes grands-parents de Tarbes et de Montélier et à ma petite sœur Jeanne :***

Merci de m'avoir toujours soutenue et accompagnée. J'espère vous avoir rendus fiers. Je vous embrasse.

***A Nicolas :*** merci pour tout.





# TABLE DES MATIERES

<b>TABLES DES ILLUSTRATIONS</b> .....	17
<b>RESUME</b> .....	21
<b>ABSTRACT</b> .....	23
<b>LISTE DES ABBREVIATIONS</b> .....	25
<b>INTRODUCTION</b> .....	27
<b>I. La déficience intellectuelle</b> .....	27
1) Définitions et diagnostic.....	27
2) Prévalence de la DI.....	28
3) Présentation clinique.....	28
4) Intérêt du diagnostic étiologique de la DI, conseil génétique.....	29
5) Etiologies de la DI .....	30
1. Etiologies environnementales.....	30
a. Causes environnementales prénatales .....	31
b. Causes environnementales périnatales et postnatales.....	31
2. Etiologies génétiques.....	32
a. Etiologies chromosomiques.....	33
b. Etiologies géniques.....	37
c. Modèle du « double hit ».....	39
6) Recommandations des sociétés savantes.....	41
<b>II. Méthodes de détection des anomalies chromosomiques à l'heure du NGS</b> .....	43
1) Caryotype.....	43
2) FISH.....	44
3) ACPA (CGH Array) .....	46
<b>III. Le Séquençage Haut Débit</b> .....	49
1) Introduction .....	49
2) Principe du NGS.....	51

3) Analyse bio-informatique .....	53
4) OneSeq Constitutional Research Panel .....	56
5) Objectif de l'étude .....	58
<b>MATERIEL ET METHODES</b> .....	59
<b>I. Recrutement des patients</b> .....	59
<b>II. Extraction des ADN</b> .....	61
<b>III. Protocole SureSelect XT HS</b> .....	61
<b>IV. Optimisation des paramètres bio-informatiques pour détecter les CNV</b> .....	62
<b>V. Tri des CNV</b> .....	64
<b>RESULTATS</b> .....	65
<b>I. Patients avec un génotype connu et analyse ouverte</b> .....	65
1) Comparaison du OneSeq avec l'ACPA .....	65
2) Détection des régions d'homozygotie avec le OneSeq .....	74
3) Détection des aneuploïdies diagnostiquées par QF-PCR avec le OneSeq .....	76
4) Détection des variants géniques avec le OneSeq .....	78
<b>II. Patients avec un génotype connu et analyse en aveugle par deux opérateurs</b> .....	79
<b>DISCUSSION</b> .....	82
<b>I. Analyse des résultats</b> .....	82
1) Analyse ouverte.....	82
2) Analyse en aveugle.....	88
3) Limite liée au nombre de CNV détectés .....	94
<b>II. Intérêt de la détection simultanée des CNV, SNV et LOH</b> .....	95
1) Rendement diagnostique .....	95
2) Intérêt économique et diminution du délai diagnostique .....	96
<b>III. Avantages de la détection des CNV par utilisation de sondes intergéniques</b> .....	98
<b>CONCLUSION</b> .....	100
<b>REFERENCES BIBLIOGRAPHIQUES</b> .....	103
<b>ANNEXES</b> .....	117

<b>I.</b>	<b>ANNEXE 1 : chevauchement des coordonnées génomiques entre le OneSeq et l'ACPA pour les cas n°5, 8, 9, 13, 14, 16, 18 et 19 .....</b>	<b>117</b>
<b>II.</b>	<b>ANNEXE 2 : liste des gènes impliqués dans la DI en mars 2020 selon SysID et inclus dans l'exome clinique du OneSeq .....</b>	<b>127</b>
<b>III.</b>	<b>ANNEXE 3 : protocole SureSelect XT HS – Agilent Technologies .....</b>	<b>131</b>
<b>IV.</b>	<b>ANNEXE 4 : Serment d'Hippocrate.....</b>	<b>137</b>





# TABLES DES ILLUSTRATIONS

## TABLEAUX

Tableau 1: liste des 44 gènes à explorer au minimum devant une DI sans orientation clinique évidente (ANPGM).....	42
Tableau 2: ratio d'intensité des fluorochromes sur les autosomes en ACPA .....	47
Tableau 3: paramètres bio-informatiques des algorithmes permettant l'identification des CNV en ACPA .....	48
Tableau 4 : comparaison du séquençage Sanger et du NGS .....	50
Tableau 5: étapes successives du traitement bio-informatique des données de NGS .....	53
Tableau 6: patients étudiés et indications ayant conduit à une analyse chromosomique.....	60
Tableau 7: paramètres bio-informatiques du filtre proposé par le fournisseur .....	63
Tableau 8: filtres optimisés pour la détection des anomalies chromosomiques homogènes et en mosaïque .....	64
Tableau 9: comparaison des résultats du OneSeq par rapport à l'ACPA en analyse ouverte ..	68
Tableau 10: analyse des performances du OneSeq pour la détection des LOH.....	75
Tableau 11: trisomie 13 homogène diagnostiquée par QF-PCR.....	76
Tableau 12: analyse en OneSeq de 2 patients présentant des variants géniques.....	78
Tableau 13: patients 23 à 30 analysés en aveugle par deux opérateurs indépendants .....	81
Tableau 14: comparaison du coût estimé d'une analyse de OneSeq par rapport à l'association ACPA-exome clinique .....	97

## FIGURES

Figure 1: nombre de gènes identifiés dans la DI en fonction du mode de transmission depuis 1980, d'après Vissers et al., 2015 .....	38
Figure 2: incidence d'un deuxième CNV d'une taille supérieure à 500 kb, selon le CNV et/ou le type de PIEV concerné, d'après Girirajan et al., 2013.....	40
Figure 3: arbre décisionnel proposé par l'ANPGM pour le diagnostic étiologique de la DI... 41	
Figure 4: caryotype féminin en bandes R.....	44
Figure 5: FISH marquant le centromère du chromosome X et le gène SHOX .....	45
Figure 6: représentation par ratio de fluorescence d'un chromosome 8 en ACPA .....	46
Figure 7: fragment d'intérêt après la ligation.....	51

Figure 8: représentation schématique des notions de profondeur et de couverture de séquençage .....	54
Figure 9: détection des variants et des délétions et duplications chromosomiques par NGS ..	55
Figure 10: design des différentes sondes du OneSeq Constitutional Research Panel.....	56
Figure 11 : délétion 22q11.21 : comparaison des coordonnées génomiques du OneSeq avec l'ACPA, chevauchement de 99% .....	69
Figure 12: duplication 17p11.2 : comparaison des coordonnées génomiques du OneSeq avec l'ACPA, chevauchement de 91% .....	70
Figure 13: délétion 15q11.2 : chevauchement des coordonnées génomiques du OneSeq avec l'ACPA, chevauchement de 88% .....	71
Figure 14: isodicentrique 15q surnuméraire : comparaison des coordonnées génomiques du OneSeq avec l'ACPA, chevauchement de 93% .....	72
Figure 15: chromosome 18 recombinant : comparaison des coordonnées génomiques du OneSeq avec l'ACPA, chevauchement de 94% .....	73
Figure 16: trisomie 13 en mosaïque faible suspectée en ACPA devant une déviation de la ligne de base des log-ratios .....	76
Figure 17: ACPA : représentation des sondes du chromosome 13 .....	77
Figure 18: OneSeq : représentation des sondes du chromosome 13 .....	77
Figure 19: duplication 17p11.2 : sondes dupliquées non détectées .....	83
Figure 20: isodicentrique 15q surnuméraire : sondes dupliquées non détectées.....	84
Figure 21: présence de sondes OneSeq à l'intérieur de la duplication segmentaire et absence de sondes d'ACPA.....	85
Figure 22: cas n°5 et 13, délétions non détectées par le OneSeq mais visibles sur la plateforme Alissa Align & Call.....	85
Figure 23: cas n°5 et 13, délétions détectées par le OneSeq après modification de l'algorithme sur Alissa Align & Call.....	86
Figure 24: détection de la variation c.858C>G dans le gène <i>CHST14</i> .....	87
Figure 25: détection d'une région d'homozygotie en 15q11.2q21.2 comprenant le gène <i>CHST14</i> .....	87
Figure 26: cheminement de la duplication 15q13.3 avec l'arborescence de l'opérateur 1 .....	90
Figure 27: cheminement de la duplication 15q13.3 avec l'arborescence de l'opérateur 2 .....	91
Figure 28: arborescence imaginée pour les CNV après l'analyse en aveugle .....	92
Figure 29: arborescence permettant de filtrer les CNV, SNV et LOH optimisée après l'analyse en aveugle .....	93

Figure 30: représentation de deux TAD et conséquences engendrées par une délétion, une duplication et une inversion sur l'architecture chromatinienne ; d'après Spielmann et al., 2018..... 99



## RESUME

**Introduction :** La Déficience Intellectuelle (DI) est fréquente et hétérogène. Des causes génétiques d'origine génique (20-30%) ou chromosomique (10-15%) sont impliquées dans environ 50% des cas. Le diagnostic étiologique génétique comprend notamment une Analyse Chromosomique sur Puce à ADN (ACPA) puis le séquençage d'un ou plusieurs panels de gènes. L'étude portait sur une technologie innovante permettant la détection simultanée des mutations ponctuelles, d'anomalies chromosomiques déséquilibrées et de régions d'homozygotie. L'objectif était d'évaluer ses performances dans l'identification des délétions et duplications chromosomiques (Copy Number Variations) par comparaison avec l'ACPA.

**Matériel et méthodes :** 30 patients à génotype connu ont été testés avec une capture OneSeq Constitutional Research Panel.

**Résultats :** l'analyse a permis d'obtenir le bon résultat chez 27 des 30 patients testés (33 variants sur 36).

**Discussion :** les CNV non détectés sont liés soit au paramétrage bio-informatique (1 CNV), soit à l'arborescence filtrant les variants (1 CNV d'interprétation débattue), soit au design de l'analyse (1 CNV trop petit). Ils pourraient être identifiés par optimisation de certains paramètres. Une étude médico-économique a montré un coût équivalent à celui de l'association ACPA et séquençage d'un panel de gènes.

**Conclusion :** la sensibilité de cette technologie est très similaire à celle de l'ACPA et pourrait être une alternative à l'association ACPA-séquençage d'un panel de gènes en routine.

Le rendement diagnostique serait identique ce qui devrait permettre de diminuer l'errance diagnostique et le coût analytique liés à la multiplication d'exams génétiques.

**Mots-clefs :** déficience intellectuelle, séquençage haut débit, copy number variation, mutation ponctuelle, single nucleotide polymorphism, région d'homozygotie



## ABSTRACT

**Introduction:** Intellectual Disability (ID) is a common condition and is extremely heterogeneous. Genetic causes are involved in approximately 50% of cases and may be the result from gene mutations (20-30%) or chromosomal aberrations (10-15%). The genetic etiological diagnosis includes a Comparative Genomic Hybridization Array (array-CGH) then, next-generation sequencing (NGS) genes panels. We tested a novel NGS technology enabling combined detection of gene mutations, chromosomal unbalanced aberrations and regions of homozygosity. Our objective was to assess this method performance to detect chromosomal copy number variations (CNV) compared to array-CGH.

**Material and methods:** 30 patients with known genotype were analyzed with the OneSeq Constitutional Research Panel capture design.

**Results:** correct identification was established in 27 of the 30 patients tested (33 variants/36)

**Discussion:** undetected CNV are related either to chosen bioinformatics parameters (1 CNV), to the variant classification tree (1 CNV of questionable clinical significance) or the analysis design (1 very small CNV). They could be identified by optimizing some parameters. A medico-economic study has shown a comparable cost than array-CGH and genes sequencing panel association.

**Conclusion:** this NGS technology sensitivity is very similar to array-CGH for CNV detection and could be an alternative to array-CGH and NGS genes panel association in routine care. The expected diagnosis yield is identical and should reduce time to diagnosis and analytical cost.

**Keywords:** intellectual disability, next generation sequencing, copy number variation, point mutation, single nucleotide polymorphism, region of homozygosity





## LISTE DES ABBREVIATIONS

**AAID:** American Association on Intellectual and Developmental Disabilities

**ACPA :** Analyse Chromosomique sur Puce à ADN (CGH-array)

**AnDDI-Rares :** filière Santé Anomalies du Développement avec ou sans Déficience Intellectuelle de causes Rares

**ANPGM :** Association Nationale des Praticiens de Génétique Moléculaire

**APA:** American Psychiatric Association

**BAF:** B-Allele Frequency

**CIM :** Classification Internationale des Maladies

**CNV :** Copy Number Variations

**ddNTP :** Didésoxyribonucléotides triphosphates

**DGV :** Database of Genomic Variants

**dNTP :** Désoxyribonucléotides triphosphates

**DI :** Déficience Intellectuelle

**DPNI :** Dépistage Prénatal Non Invasif

**Kb :** kilobase (1000 paires de bases d'ADN bicaténaire)

**LOH:** Loss Of Heterozygosity (perte d'hétérozygotie)

**MLPA:** Methylation-Specific Multiplex Ligation-Dependent Probe Amplification

**MS-MLPA:** Multiplex Ligation-Dependent Probe Amplification

**NGS/SHD :** Séquençage Nouvelle Génération/Séquençage Haut Débit

**OMS :** Organisation Mondiale de la Santé

**Pb :** paire de bases

**PCR :** Polymerase Chain Reaction

**PIEV :** remaniement récurrent de susceptibilité aux troubles neuro-développementaux à pénétrance incomplète et/ou expressivité variable

**QF-PCR :** Quantitative Fluorescent PCR

**QI :** Quotient Intellectuel

**SNP:** Single Nucleotide Polymorphism

**SNV:** Single Nucleotide Variation (mutation ponctuelle)

**TAD:** Topologically Associating Domains

**WISC:** Weschsler Intelligence Scale for Children



# INTRODUCTION

## I. La déficience intellectuelle

### 1) Définitions et diagnostic

Trois organisations internationales, l'OMS (CIM-10/CIM-11 en cours de traduction), l'*American Psychiatric Association* (DSM-5) et l'*American Association on Intellectual and Developmental Disabilities* (AAIDD) proposent des définitions de la déficience intellectuelle (DI) ayant en commun les critères suivants [1–3]:

- Un déficit dans les fonctions intellectuelles confirmé cliniquement et par des tests d'intelligence personnalisés et normalisés
- Des limitations significatives du comportement adaptatif
- Une apparition au cours de la période développementale

L'OMS (CIM-10) définit la DI comme un « arrêt ou développement incomplet du fonctionnement mental, caractérisé essentiellement par une altération, durant la période du développement, des facultés qui déterminent le niveau global d'intelligence, c'est-à-dire des fonctions cognitives, du langage, de la motricité et des capacités sociales » [1].

La réalisation d'un bilan neuropsychologique comprenant des tests psychométriques comme le WISC-V pour les enfants de 6 à 16 ans permet d'évaluer le fonctionnement intellectuel par rapport au reste de la population. Si le résultat de l'évaluation du quotient intellectuel (QI) est inférieur à 70, une DI sera évoquée.

En fonction de la valeur mesurée du QI, plusieurs degrés de déficience intellectuelle sont définis par l'OMS (CIM-10) :

- DI légère : QI entre 50 et 69
- DI modérée : QI entre 35 et 49
- DI sévère : QI entre 20 et 34
- DI profonde : QI inférieur à 20

Cette classification est cependant de moins en moins utilisée, le DSM-5 et l'AAIDD définissant les niveaux de gravité de la DI par rapport au retentissement sur le fonctionnement adaptatif et sur l'intensité des besoins de soutien de l'individu.

L'évaluation du QI ne permet pas de caractériser les capacités d'adaptation d'un individu à son environnement et peut être complétée par l'utilisation d'échelles évaluant l'adaptation sociale à un milieu donné. Ces échelles permettent d'évaluer les performances réelles de l'individu dans des situations de la vie quotidienne et sont utiles pour orienter les accompagnements éducatifs les plus adaptés. En France, l'échelle Vineland-II est la plus utilisée [4].

## 2) Prévalence de la DI

Le QI se distribue théoriquement selon une courbe de Gauss dans laquelle 2,3% de la population présenterait un QI inférieur à 70 [5]. L'association d'un QI inférieur à 70 et de limitations du comportement adaptatif est présente chez 0,97% de la population [6]. Ainsi, la prévalence de la DI varie en fonction des critères considérés entre 1% et 2,3%.

Selon une méta-analyse réalisée en 2010, la prévalence de la DI serait de 10,37 pour 1000 soit environ 1% de la population mondiale [7].

L'OMS recommande de distinguer la DI légère (QI entre 50 et 69) de la DI sévère (QI inférieur à 50) dans l'estimation de la fréquence de la DI. La DI légère serait 2 à 6 fois plus fréquente que la DI sévère. Dans les pays développés, la prévalence de la DI sévère est estimée entre 3 et 4 enfants pour 1000 ; celle de la DI légère varie entre 10 et 18 enfants pour 1000 [4].

La DI est également plus fréquente chez les garçons avec un *sex-ratio* entre 1,2 et 1,6 [8].

Cette fréquence s'explique en grande partie par les nombreux gènes impliqués dans la DI liée au chromosome X qui atteint principalement les garçons [9].

Ainsi, la DI est une problématique importante de santé publique du fait de sa fréquence et de la nécessité d'une prise en charge spécialisée à vie pour ces patients.

## 3) Présentation clinique

Les DI ont une présentation clinique hétérogène puisqu'elles peuvent être isolées ou syndromiques.

Les DI syndromiques peuvent être identifiées par le phénotype qu'elles induisent, qui sera similaire d'un patient à l'autre. Le diagnostic étiologique peut alors être orienté par le

clinicien. Les syndromes prédominants comportant une DI sont la trisomie 21, le syndrome de DiGeorge, le syndrome de l’X fragile, le syndrome de Williams-Beuren et le syndrome de Prader-Willi [10].

Cependant, les DI entraînant des syndromes reconnaissables restent minoritaires puisqu’elles représentent seulement 2% des causes identifiées de DI [4]. Les DI isolées sont plus fréquentes et, en l’absence d’orientation clinique, elles requièrent des analyses génétiques pangénomiques pour parvenir à un diagnostic.

La DI est fréquemment associée à des comorbidités pouvant modifier le pronostic ou la prise en charge.

Les malformations congénitales sont présentes chez 30% des enfants avec une DI contre environ 5% des enfants non déficients [11]. La fréquence des malformations congénitales varie en fonction de la sévérité de la DI : environ 28% chez les patients avec un QI entre 40 et 69 contre 54% chez ceux ayant un QI inférieur à 40. Le risque de malformation congénitale augmente donc avec la sévérité de la DI [11]. Les malformations associées les plus fréquentes sont celles du système nerveux central (6%), du système musculosquelettique (3%) et du système uro-génital (3%) [4]. La fréquence des malformations du système nerveux central s’explique probablement par leur relation avec l’étiologie de la DI.

Les troubles du spectre autistique sont les troubles neuro-développementaux les plus souvent associés à la DI avec une prévalence d’environ 8 à 10% [4,12,13].

Parmi les autres comorbidités fréquemment associées à la DI, il faut citer notamment l’épilepsie, la paralysie cérébrale, le trouble de déficit de l’attention avec ou sans hyperactivité et les troubles anxieux [13,14].

#### **4) Intérêt du diagnostic étiologique de la DI, conseil génétique**

L’intérêt du diagnostic étiologique en cas de DI est multiple. Il permet notamment de préciser le pronostic, de mettre en place un suivi médical approprié et d’accéder aux protocoles thérapeutiques. De plus, la fiabilité du conseil génétique dépend grandement du diagnostic étiologique.

Un conseil génétique approprié permet d’évaluer le risque de récurrence chez un enfant à naître pour les parents [4]. La mise en évidence d’une anomalie génétique accidentelle sera plutôt rassurante malgré le risque de mosaïque germinale tandis qu’une anomalie héritée

pourra permettre au couple de recourir au diagnostic prénatal voire au diagnostic pré-implantatoire.

Un conseil génétique de la famille élargie pourra également s'avérer utile et permettra de dépister les personnes à risque de transmettre l'anomalie génétique et donc de leur donner accès au diagnostic prénatal.

Il est également important de préciser que le diagnostic étiologique donne une réponse aux interrogations des familles et peut avoir en ce sens un bénéfice potentiel.

Enfin, avec l'exploration de la physiopathologie de la DI et l'augmentation des connaissances, des potentielles cibles thérapeutiques pourraient être envisagées à l'avenir.

Le diagnostic étiologique de la DI est donc essentiel à la prise en charge de ces patients. Cependant, la majorité d'entre eux restent sans diagnostic moléculaire en raison de l'extrême hétérogénéité clinique et étiologique de la DI.

## **5) Etiologies de la DI**

La DI est liée à un événement perturbant le développement cérébral en prénatal, périnatal et/ou en postnatal.

Les étiologies impliquées dans la DI sont nombreuses et hétérogènes. Elles peuvent être environnementales, génétiques ou plurifactorielles.

Dans 35 à 40% des cas, et malgré un dépistage approprié, l'origine de la DI reste inexplicée. Le diagnostic étiologique est établi dans 40 à 60% des cas. Cette variation dépend de la population étudiée, de l'analyse génétique réalisée et de la sévérité de la DI. En effet, le rendement diagnostique est d'environ 75% dans les DI sévères contre seulement 20% dans les DI légères [4].

L'hétérogénéité étiologique est responsable de difficultés diagnostiques pouvant entraîner une errance diagnostique et un retard dans la prise en charge adaptée des patients.

### **1. Etiologies environnementales**

Les causes environnementales représentent environ 15% des étiologies des DI [4].

Elles sont différenciées en fonction de la période de survenue : prénatales, périnatales ou postnatales.

### **a. Causes environnementales prénatales**

Les causes environnementales prénatales de la DI sont notamment :

- Des infections fœtales comme la toxoplasmose, la rubéole ou le cytomégalovirus qui augmentent significativement le risque de troubles neuro-développementaux postnataux [15,16].
- Des intoxications durant la grossesse (iatrogénie médicamenteuse, tabac, drogues, neurotoxiques, alcool...) pouvant altérer le développement du système nerveux central [4].
- Une grande prématurité peut avoir comme complication une déficience intellectuelle avec une fréquence estimée de 5 à 15% [4,17].
- Une hypothyroïdie maternelle durant la grossesse expose à un QI global inférieur de 7 points par rapport à une population témoin [18].
- Une hyperphénylalaninémie maternelle peut avoir un effet tératogène pour l'embryon et le fœtus [19].

### **b. Causes environnementales périnatales et postnatales**

Les causes environnementales périnatales de DI comprennent notamment les complications obstétricales comme les anoxies et les hémorragies cérébrales pouvant causer des lésions cérébrales irréversibles [20–22].

Les étiologies postnatales de DI sont nombreuses, il est possible de citer :

- Les maladies infectieuses de la petite enfance comme les méningites ou les encéphalites [23,24].
- Les traumatismes crâniens sévères peuvent entraîner des séquelles cognitives et une diminution du QI de 15 points en moyenne [25,26].
- Les intoxications aux métaux lourds neurotoxiques comme le plomb ou le mercure. Il a été montré que l'exposition au plomb entraîne une diminution du QI d'environ 5 points pour chaque augmentation de 10 µg/dL de la plombémie [27].
- Les facteurs psychosociaux et économiques sont importants, notamment dans la DI légère. Un niveau élevé d'éducation parentale serait un facteur protecteur de la DI [28–30].



## 2. Etiologies génétiques

A l'heure actuelle, une cause génétique est identifiée dans 25 à 50% des cas de DI. Les causes génétiques les plus fréquentes sont la trisomie 21 et le syndrome de l'X fragile.

En dehors de ces deux étiologies, chaque cause génétique de DI peut être considérée comme une maladie rare car chaque DI d'origine génétique ne représente qu'une très petite proportion des étiologies.

Environ 1/3 des 20 000 gènes du génome humain sont exprimés au niveau cérébral et participent à son développement et à son fonctionnement [4]. L'altération d'un ou plusieurs de ces gènes peut entraîner des anomalies du développement cérébral. Ils sont responsables, en majorité, de DI non syndromiques ne permettant pas au clinicien d'orienter le diagnostic étiologique.

A noter également que de nombreux gènes responsables de DI n'ont probablement pas été identifiés et peuvent entraîner un bilan étiologique génétique faussement négatif. Il est donc important de ré-analyser les résultats négatifs à intervalle régulier pour tenir compte de l'avancée des connaissances en génétique.

Enfin, tous les modes de transmission sont décrits dans la DI et compliquent le diagnostic étiologique. Il faut citer notamment :

- La transmission mendélienne (maladie génétique monogénique) : autosomique ou liée au chromosome X de façon dominante ou récessive.
- La transmission non mendélienne c'est-à-dire sans altération de la séquence du génome nucléaire : hérédité mitochondriale à transmission maternelle exclusive, anomalie de l'empreinte, disomie uniparentale et hérédité multifactorielle dont la pathogénicité dépend de la combinaison d'allèles normaux de certains gènes dans un environnement particulier.

Ces différents éléments expliquent l'hétérogénéité clinique et étiologique génétique majeure de la DI et la grande difficulté rencontrée pour établir un diagnostic [4].

Les étiologies génétiques peuvent être séparées en 2 grands groupes : les étiologies chromosomiques et les étiologies géniques.

### **a. Etiologies chromosomiques**

Les anomalies chromosomiques représentent environ 10 à 15% des DI [4,31].

Il est possible de les diviser en deux groupes : les anomalies de nombre et les anomalies de structure [31].

#### **Les anomalies de nombre :**

Les aneuploïdies correspondent à la perte ou au gain d'un ou plusieurs chromosomes. Elles sont fréquentes dans l'espèce humaine et concernent environ 0,3% des nouveau-nés [32].

Parmi les trisomies viables, la plus connue est la trisomie 21 ou syndrome de Down, responsable d'une DI syndromique [4].

Les trisomies 13 et 18 sont responsables de malformations congénitales sévères associées à un retard de développement profond avec DI [33]. Ces deux trisomies sont le plus souvent rapidement létales avec un âge médian du décès à 10 jours de vie [34].

Les trisomies 13, 18 et 21 sont causées en particulier par des non-disjonctions méiotiques dont le risque augmente avec l'âge maternel [35]. Leur prévalence totale est donc en augmentation [36,37]. Cependant, la prévalence à la naissance reste stable du fait du dépistage prénatal et de la possibilité d'interruption médicale de grossesse. Celle-ci est d'environ 11,2/10 000 naissances pour la trisomie 21 ; elle est beaucoup plus faible pour les trisomies 13 et 18 : 0,48/10 000 naissances et 1,04/10 000 naissances respectivement [37].

Plus rarement, les trisomies 13, 18 et 21 sont liées à la formation de gamètes déséquilibrés durant la ségrégation méiotique et résultent d'une anomalie de structure équilibrée parentale.

Parmi les aneuploïdies responsables de DI syndromiques, il convient également de citer les dysgonosomies avec présence d'un ou plusieurs chromosomes X surnuméraires comme par exemple la tétrasomie X, la pentasomie X ou encore le syndrome 49,XXXXY. La sévérité de la DI augmente avec chaque chromosome X surnuméraire [38]. Il s'agit d'anomalies très rares le plus souvent liées à des non-disjonctions en méiose I et II maternelle [39].

De rares cas de trisomies 8 et 9 en mosaïque ont également été décrits et sont responsables de DI syndromiques [40,41].

## **Les anomalies de structure :**

Les anomalies de structure résultent de cassures chromosomiques suivies par une reconstitution anormale.

Elles sont de deux types :

- Équilibrées c'est-à-dire sans perte ou gain de matériel chromosomique (inversion, translocation réciproque ou robertsonienne, insertion)
- Déséquilibrées c'est-à-dire avec perte ou gain de matériel chromosomique (délétion, duplication, dérivé de translocation, chromosome en anneau, isochromosome...)

**Les anomalies de structure équilibrées** sont habituellement sans conséquence phénotypique. Elles peuvent cependant induire une pathologie du point de cassure notamment par interruption d'un gène, par effet de position ou par perturbation de l'empreinte parentale. Ces mécanismes ont déjà été décrits chez des patients présentant des DI syndromiques [42–45].

Il est estimé dans une étude qu'un phénotype anormal est retrouvé dans environ 6,7% des translocations *de novo* et dans 9,4% des inversions *de novo* [46]. Ce risque augmente avec le nombre de points de cassure [47].

A l'heure actuelle, il est difficile d'estimer l'implication de ces remaniements équilibrés dans la DI car la détection des points de cassure n'est pas communément réalisée. Cependant, il a été montré que le séquençage de génome entier permettait la détection des anomalies de structure équilibrées et devrait rendre possible, à l'avenir, de les rechercher en routine diagnostique [48–50].

**Les anomalies de structures déséquilibrées** sont responsables de phénotypes très divers en fonction notamment du type de déséquilibre (délétion ou duplication), de la taille du segment délété ou dupliqué et du ou des chromosomes impliqués. Les anomalies de structure déséquilibrées sont responsables de nombreuses DI syndromiques ou isolées.

Il est possible de citer notamment :

Les syndromes microdélétionnels et microduplicationnels : au début des années 2000, l'essor de l'Analyse Chromosomique sur Puce à ADN (ACPA) a permis la détection de remaniements submicroscopiques appelés Copy Number Variations (CNV). Il s'agit de variations génétiques correspondant à des gains (duplications, triplications) ou des pertes (délétions) de matériel génétique. Les CNV sont à l'origine d'une grande variabilité génomique entre les individus et constituent environ 4,8 à 9,7% du génome humain [51].

Ils peuvent être sans conséquence phénotypique ou responsables de pathologies [52,53].

Il est possible de distinguer les remaniements récurrents et non récurrents.

La recherche des CNV a permis de détecter des remaniements déséquilibrés spécifiques de syndromes décrits cliniquement mais jusqu'alors inexplicables au caryotype.

De nombreuses anomalies récurrentes médiées par des duplications segmentaires ont ainsi été découvertes. Ces duplications segmentaires possèdent une forte homologie de séquence propice aux recombinaisons homologues non-alléliques. Elles entraînent, par mésappariement et crossing-over inégal, une délétion ou duplication de taille identique d'un patient à l'autre car les points de cassures se situent dans les duplications segmentaires [4,54].

Les gènes présents dans le remaniement sont ainsi soumis à un dosage génique anormal responsable du phénotype. Les syndromes microduplicationnels récurrents ont souvent un phénotype plus variable et moins sévère que les syndromes microdélétionnels correspondants [4]. Les syndromes récurrents responsables de DI syndromiques sont nombreux, le plus fréquent étant le syndrome de DiGeorge.

Les remaniements non récurrents peuvent être responsables de DI isolées ou syndromiques par interruption d'un gène impliqué dans le neuro-développement ou sensible au dosage génique (haplo-insuffisance ou hyper-expression) [55]. Par exemple, la duplication du gène *MECP2* est responsable, par disomie fonctionnelle, d'une DI syndromique sévère [56].

D'une façon générale, une corrélation positive entre la taille du CNV, le nombre de gènes qu'il contient et la sévérité du phénotype a été décrite [57].

Les chromosomes en anneau : ils résultent de la circularisation d'un chromosome après double délétion terminale du bras court et du bras long et fusion des extrémités [58]. Dans les monosomies X avec présence d'une mosaïque 45,X/46X,r(X), un anneau du chromosome X de petite taille avec absence du gène *XIST* sera responsable d'un phénotype plus sévère avec DI par disomie fonctionnelle liée à l'absence d'inactivation de l'anneau du chromosome X [59,60].

Il est également possible de citer l'anneau du chromosome 20, le plus souvent en mosaïque, responsable d'un syndrome associant épilepsie réfractaire, DI et troubles du comportement [61].

Les marqueurs chromosomiques surnuméraires : il s'agit par définition d'un chromosome additionnel de structure anormale non identifiable avec le banding au caryotype. Une étude a montré que des marqueurs chromosomiques étaient retrouvés avec une fréquence de 0,288% dans la DI contre seulement 0,044% en population générale [62].

Ils posent des difficultés diagnostiques en raison des phénotypes variés engendrés, de leur petite taille et de leur présence fréquente en mosaïque [63].

Les marqueurs chromosomiques surnuméraires du chromosome 15 sont les plus fréquents [62]. Ceux qui sont d'origine maternelle et comprennent le gène *SNRPN* sont responsables de 80% des cas de syndrome de duplication 15q entraînant une DI syndromique sévère [64].

#### Régions d'homozygotie et pathologies liées à une anomalie de l'empreinte parentale :

Les régions d'homozygotie concerneraient 5 à 6% des patients [65,66]. Elles correspondent à des zones de taille variable dans lesquelles les deux allèles sont identiques [67]. En fonction du contexte, elles peuvent indiquer une homozygotie ancestrale, une consanguinité ou une disomie uniparentale.

La présence de plusieurs régions d'homozygotie réparties sur différents chromosomes sera plutôt en faveur d'une consanguinité ou d'une homozygotie ancestrale. En revanche, une région d'homozygotie localisée sur un seul chromosome fera évoquer une disomie uniparentale c'est-à-dire l'absence de contribution biparentale à un locus donné [68]. Si l'individu hérite d'une paire de chromosomes homologues issus du même parent, il s'agit d'une hétérodisomie ; s'il hérite de deux copies du même chromosome parental, il s'agit d'une isodisomie.

Dans de nombreux cas, la signification clinique des régions d'homozygotie est indéterminée. Elles peuvent être impliquées en pathologie par révélation d'une maladie autosomique récessive ou par absence de contribution biparentale sur des chromosomes soumis à empreinte parentale (chromosomes 6, 7, 11, 14, 15...).

En effet, les régions génomiques soumises à empreinte parentale sont exprimées de façon mono-allélique en fonction de leur origine parentale, maternelle ou paternelle. Toute anomalie génétique empêchant la transcription de cet allèle entraînera une suppression totale de l'expression du gène concerné.

Parmi les anomalies de l'empreinte responsable de DI, il convient de citer les syndromes de Prader-Willi et d'Angelman dont l'incidence est estimée à 1/15 000 à 25 000 naissances [69]. Le syndrome de Prader-Willi est un syndrome des gènes contigus lié à l'absence d'expression de l'allèle paternel [4]. Le phénotype comprend notamment petite taille, hyperphagie avec obésité, troubles du comportement et DI modérée.

Le syndrome d'Angelman est causé par l'absence d'expression de l'allèle maternel du gène *UBE3A*. Le phénotype comprend une DI sévère avec absence de langage, une épilepsie et une microcéphalie post-natale [70].

Les syndromes de Prader-Willi et d'Angelman sont ainsi deux DI syndromiques avec un phénotype très différent causés par une anomalie d'empreinte de la région 15q11-q13 [70]. Ils peuvent être causés par une délétion de l'allèle normalement exprimé, par une disomie uniparentale, par une mutation du centre de l'empreinte ou par une mutation du gène *UBE3A* dans le cas du syndrome d'Angelman.

### **b. Etiologies géniques**

L'accessibilité du séquençage haut débit (NGS) en routine diagnostique permet aujourd'hui l'identification de nouveaux gènes responsables de DI par identification de mutations ponctuelles (substitutions nucléotidiques) et d'insertions ou délétions nucléotidiques (indels).

Les étiologies monogéniques sont caractérisées par une hétérogénéité génétique très importante [71]. La découverte de gènes impliqués dans la DI ne cesse d'augmenter (Figure 1) ce qui s'explique par l'émergence du séquençage de l'exome [71].

Compte tenu du nombre extrêmement important de gènes impliqués dans le développement et le fonctionnement cérébral, de la faible récurrence des gènes impliqués et des nombreuses mutations survenant *de novo* [4], dresser une liste exhaustive des étiologies monogéniques de la DI est extrêmement difficile.

D'une façon générale, l'altération de gènes impliqués dans la neurogenèse et la migration neuronale entraînerait des DI syndromiques alors que les gènes impliqués dans la régulation synaptique seraient plutôt responsables de DI isolées [72,73].

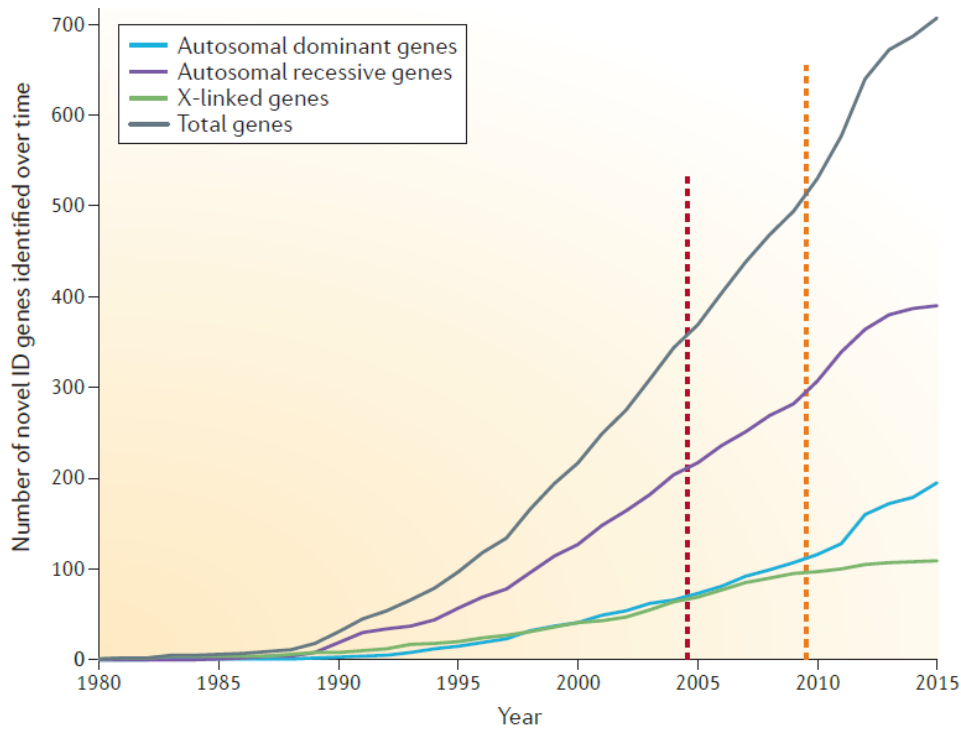


Figure 1: nombre de gènes identifiés dans la DI en fonction du mode de transmission depuis 1980, d'après Vissers et al., 2015 [71] (pointillé rouge : introduction de l'ACPA, pointillé orange : introduction du NGS)

Les DI autosomiques dominantes représenteraient plus de 20% des cas de DI [4]. Dans les pays occidentaux, elles surviennent le plus souvent de façon sporadique par mutation *de novo* [74,75]. Leur prévalence est estimée entre 0,22 et 0,42% des naissances vivantes [76].

En mars 2020, la base de données Systems Biology Approaches to ID (SysID, <http://sysid.cmbi.umcn.nl/>) liste 872 gènes possiblement impliqués dans les DI autosomiques dominantes. Parmi les gènes les plus fréquemment mutés, il est possible de citer *ARID1B*, *DYRK1A*, *SMARCA2* ou encore *MED13L* [76,77].

Les DI autosomiques récessives seraient responsables de 10 à 20% des cas de DI dans les pays occidentaux [78]. Dans les populations dans lesquelles la consanguinité est commune, elles sont probablement la cause la plus fréquente de DI [78]. En Europe, les DI autosomiques récessives apparaissent le plus souvent comme étant sporadiques et seraient dues en grande partie à des hétérozygoties composites [79].

A ce jour, la base de données SysID liste 1514 gènes possiblement impliqués dans les DI autosomiques récessives ce qui concorde avec les données de la littérature en faveur d'une hétérogénéité génique importante [80].

Près de 15% des gènes actuellement impliqués dans la DI sont portés par le chromosome X [81]. La dernière publication proposant une actualisation des gènes impliqués dans la DI liée à l’X comprend 141 gènes qui peuvent être responsables de DI isolées ou syndromiques [81]. La plupart d’entre eux ne sont responsables que d’un petit nombre de cas et d’un phénotype aspécifique [4].

L’étiologie monogénique la plus fréquemment impliquée dans la DI est le syndrome de l’X fragile [31]. La prévalence du syndrome de l’X fragile est estimée à environ 1,4/10 000 garçons [82]. Ce syndrome est causé par une inactivation transcriptionnelle du gène *FMRI* localisé en Xq27.3 [4]. Elle est liée à une expansion de triplets CGG dans la région 5’ non traduite de *FMRI*. Quand le nombre de triplets dépasse 200, on parle de mutation complète. Celle-ci entraîne une hyperméthylation de *FMRI* et une répression de son expression équivalente à une perte de fonction [83]. De rares cas de mutations ponctuelles et de délétions nucléotidiques du gène *FMRI* ont également été décrites comme responsables du syndrome de l’X fragile [84,85]. Les individus atteints présentent une DI syndromique avec notamment troubles du spectre autistique et dysmorphie [86].

### c. Modèle du « double hit »

Les remaniements récurrents de susceptibilité aux troubles neuro-développementaux à pénétrance incomplète et/ou expressivité variable (PIEV) ont été décrits chez des patients présentant des phénotypes neuro-développementaux divers de sévérité variable.

Il a été montré que dans la très grande majorité des cas, ces remaniements étaient hérités d’un parent apparemment sain [87,88].

L’observation chez environ 25% des patients d’un second remaniement de grande taille a permis de décrire le modèle du « double hit » expliquant la variabilité phénotypique importante chez deux individus porteurs d’un même remaniement récurrent.



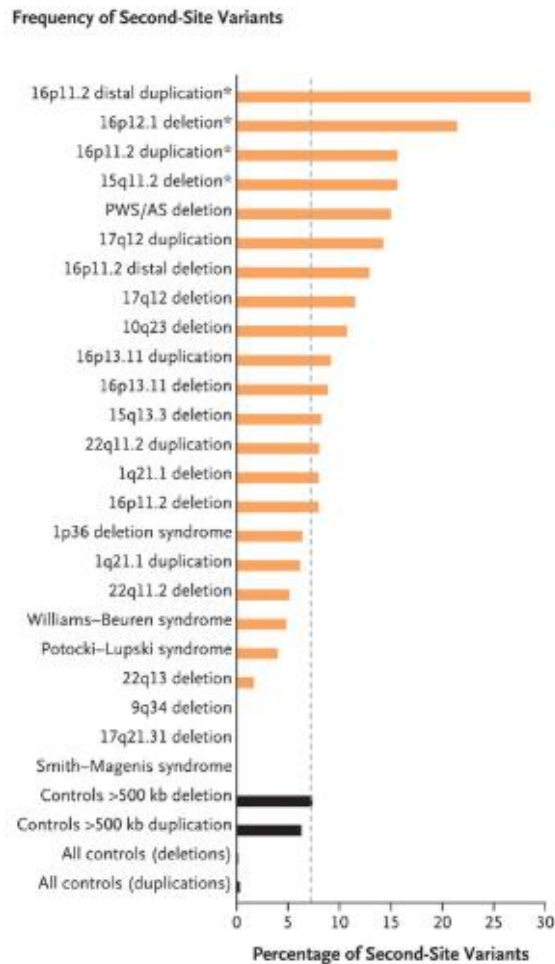


Figure 2: incidence d'un deuxième CNV d'une taille supérieure à 500 kb, selon le CNV et/ou le type de PIEV concerné, d'après Girirajan et al., 2013 [87]

Ainsi, le second remaniement, en modifiant le dosage génique de plusieurs gènes supplémentaires, serait responsable du basculement vers un phénotype sévère [88,89].

Il a également été montré que la présence de deux remaniements de signification inconnue d'une taille supérieure à 500 kb est 8 fois plus fréquente chez les individus présentant un retard de développement ou une DI ce qui semble conforter ce modèle de « double hit » [87,88](Figure 2). De plus, la présence de plusieurs grands remaniements est corrélée positivement avec la sévérité de la DI et la présence de malformations congénitales [57].

Ce modèle peut également s'appliquer par la survenue non pas d'un deuxième remaniement chromosomique mais d'une mutation ponctuelle sur le chromosome non remanié.

La délétion 16p13.11 est présente en population générale et parfois responsable d'un phénotype sévère comprenant notamment microcéphalie, épilepsie et DI. Une étude a montré la présence chez ces patients d'une mutation ponctuelle du gène *NDE1* sur le chromosome

homologue non remanié. La présence d'un remaniement récurrent associé à une mutation ponctuelle révèle ainsi un phénotype autosomique récessif [90].

Une autre étude a mis en évidence que la variabilité phénotypique du syndrome de DiGeorge serait en lien avec des mutations additionnelles du gène *SNAP29* sur le chromosome homologue non délété [91].

Ainsi, il semble pertinent de poursuivre les investigations diagnostiques par un panel de gènes ou un exome devant un patient présentant un remaniement récurrent de susceptibilité aux troubles neuro-développementaux et un phénotype sévère.

## 6) Recommandations des sociétés savantes

En 2016, l'Association Nationale des Praticiens de Génétique Moléculaire (ANPGM) et les filières « DéfiScience » et « AnDDI-Rares » (Santé Anomalies du Développement avec ou sans Déficience Intellectuelle de causes Rares) ont proposé un arbre décisionnel pour établir le diagnostic étiologique des DI sans orientation clinique évidente. Celui-ci repose sur une stratégie à deux niveaux (Figure 3).

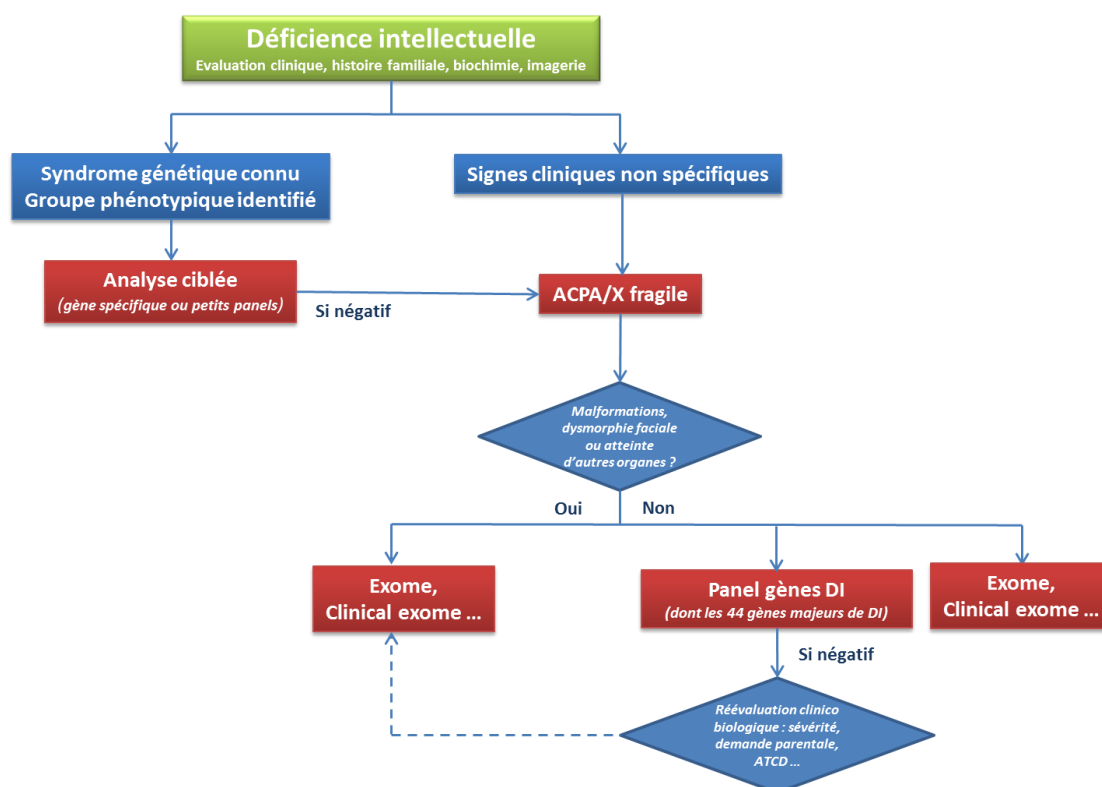


Figure 3: arbre décisionnel proposé par l'ANPGM pour le diagnostic étiologique de la DI [92]

Dans un premier temps, après évaluation clinique et en l'absence d'orientation évidente vers une cause chromosomique ou génique, une ACPA sera réalisée. Cette analyse permet de détecter les anomalies « quantitatives » du génome (délétions, duplications). Une recherche d'un syndrome de l'X fragile est également recommandée.

Dans un second temps, en cas de négativité de ces analyses de premier niveau, une exploration génique par NGS sera réalisée. Il s'agit ici de rechercher les anomalies « qualitatives » du génome (mutations ponctuelles...). Le rendement diagnostique de cette analyse dépend du nombre de gènes qui seront étudiés.

L'ANPGM, les filières « DéfiScience » et « AnDDI-Rares » ont défini un panel comprenant les 44 gènes les plus souvent impliqués dans la DI (Tableau 1).

DI44
<p><i>ANKRD11, ARID1B, ATRX, CASK, CTNNB1, CUL4B, DLG3, DYRK1A, EP300, FOXP1, GATAD2B, GRIA3, GRIN1, GRIN2A, GRIN2B, IL1RAPL1, IQSEC2, KDM5C, KMT2A (MLL), MECP2, MED13L, NAA10, PQBP1, PTCHD1, SATB2, SCN2A, SCN8A, SETBP1, SHANK3, SLC16A2, SLC2A1, SLC6A8, SLC9A6, SMARCA2, SMC1A, STXBPI, SYNGAP1, TBR1, TCF4, UPF3B, FOXG, CDKL5, RAI1, WDR45</i></p>

*Tableau 1: liste des 44 gènes à explorer au minimum devant une DI sans orientation clinique évidente (ANPGM)[92]*

De plus grands panels (200 à 500 gènes) sont apparus récemment et permettent, en association avec l'ACPA, un rendement diagnostique plus important.

Le développement du NGS permet aujourd'hui l'étude de l'exome c'est-à-dire l'analyse des exons, les parties codantes des gènes qui influencent la formation des protéines dans l'organisme. Bien que ceux-ci ne représentent que 1 à 2% du génome humain, leur étude permet une bonne rentabilité diagnostique puisqu'ils contiennent la plupart des mutations à l'origine des maladies génétiques [93–95].

Deux types d'exomes sont possibles :

- L'exome clinique permettant d'étudier environ 5000 gènes responsables de pathologies humaines.
- L'exome complet permettant de séquencer les exons des 20 000 gènes codants présents dans l'ADN, c'est-à-dire le séquençage d'environ 30 millions de paires de bases.

## **II. Méthodes de détection des anomalies chromosomiques à l'heure du NGS**

### **1) Caryotype**

Le caryotype permet, après culture des lymphocytes sanguins, de visualiser les chromosomes (Figure 4). Il s'agit d'une exploration globale du génome permettant d'analyser les chromosomes de façon quantitative (nombre) et qualitative (morphologie).

Le caryotype reste aujourd'hui la seule technique de routine capable d'identifier les anomalies dites équilibrées, sans perte ni gain de matériel génétique.

A partir de sang périphérique prélevé sur tube hépariné, les lymphocytes sanguins vont être mis en culture dans un milieu nutritif contenant notamment un mitogène, la phytohémagglutinine. Après 72h de culture, un poison du fuseau est ajouté afin de bloquer les cellules en métaphase puis, à la suite d'un choc hypotonique et d'une fixation, les cellules sont étalées sur une lame de verre humide.

Survient ensuite l'étape de dénaturation permettant l'obtention du banding spécifique à chaque paire chromosomique. Cette dénaturation peut être enzymatique pour obtenir des bandes G, ou thermique pour obtenir des bandes R. L'association de ces deux types de bandes permet de bien visualiser les extrémités chromosomiques avec les bandes R et de distinguer plus clairement les bandes à l'aide des bandes G.

Le rendement diagnostique du caryotype dans les anomalies du développement est faible puisqu'il ne permet pas de détecter les anomalies ayant une taille inférieure à 5 mégabases.

A l'heure actuelle, le caryotype est surtout indiqué dans les troubles de la reproduction pour lesquels les anomalies équilibrées sont prédominantes. Un caryotype sera également réalisé après diagnostic d'une aneuploïdie en biologie moléculaire pour orienter le conseil génétique.

Il reste cependant une technique de choix pour explorer la mécanique chromosomique des anomalies détectées en biologie moléculaire (ACPA, NGS) à condition qu'elles aient une taille supérieure à 5 mégabases.

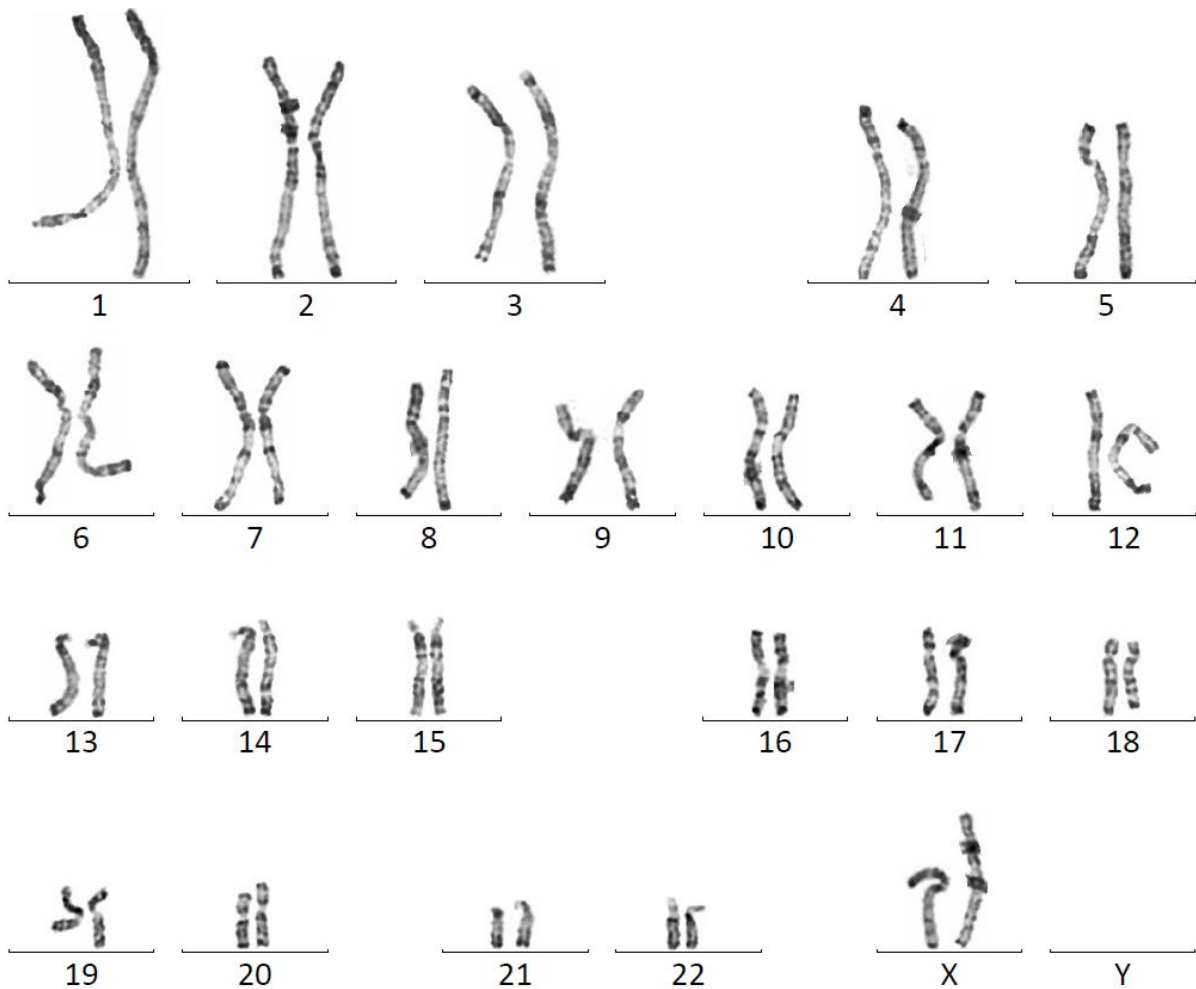


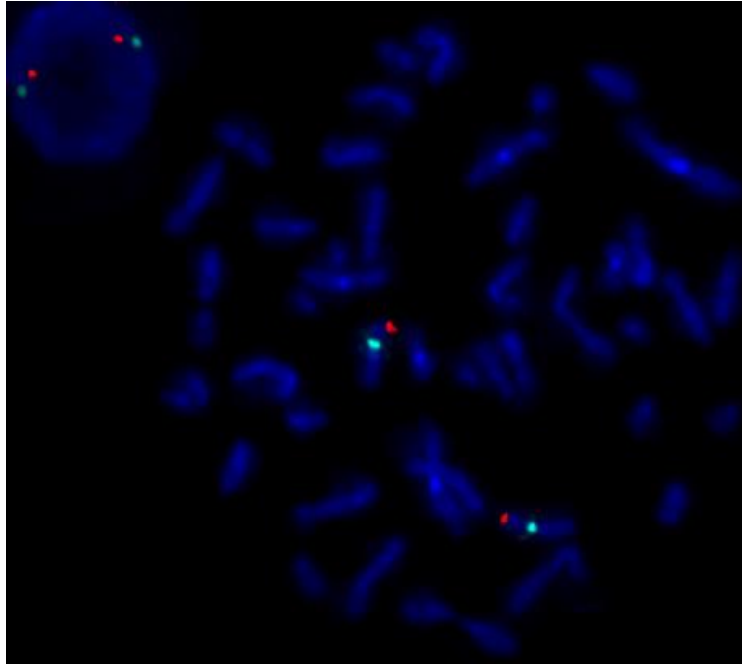
Figure 4: caryotype féminin en bandes R

## 2) FISH

La Fluorescence *In Situ* par Hybridation (FISH) permet d'identifier des anomalies chromosomiques équilibrées ou déséquilibrées avec une bien meilleure résolution que le caryotype. Elle nécessite cependant une orientation clinique puisqu'il s'agit d'une exploration ciblée du génome.

Le principe de la FISH est basé sur des homologies de séquence d'acides nucléiques permettant l'identification de la séquence à étudier. Pour cela, une sonde nucléaire marquée par un fluorochrome et spécifique de cette séquence d'intérêt va être utilisée.

La FISH peut être réalisée sur des chromosomes métaphasiques ou sur des cellules interphasiques (noyaux).



*Figure 5: FISH marquant le centromère du chromosome X (spots aqua) et le gène SHOX (spots rouges)*

Les cellules sont d'abord étalées sur une lame de verre puis la sonde est déposée directement sur cette lame. Une étape de dénaturation thermique de l'ADN cible et de la sonde entre 70 et 85°C va séparer les deux brins d'ADN par rupture des liaisons hydrogènes.

Elle est suivie par une étape d'hybridation à 37°C durant laquelle les brins d'ADN se réassocient de façon spécifique permettant ainsi à la sonde de venir s'hybrider à l'ADN cible par complémentarité de séquence. Après une étape de lavage éliminant les sondes hybridées de façon non spécifique, l'ADN est contremarqué directement sur la lame avec du 4',6-diamidino-2-phenylindole (DAPI). Celui-ci va colorer en bleu et de façon aspécifique tous les chromosomes et noyaux pour permettre de les repérer. La lecture peut alors être réalisée avec un microscope à fluorescence.

Avec le développement de l'ACPA, l'intérêt de la FISH dans les anomalies du développement est aujourd'hui restreint. Cette technique pourra être utilisée devant une forte suspicion clinique d'un syndrome génétique permettant une recherche ciblée en FISH, pour réaliser une enquête familiale ou en seconde ligne après l'ACPA pour préciser un potentiel remaniement de structure à l'origine d'un CNV.

### 3) ACPA (CGH Array)

L'Analyse Chromosomique sur Puce à ADN (ACPA) est une exploration chromosomique pangénomique de choix pour étudier les pathologies associées aux anomalies chromosomiques constitutionnelles. Cette technique permet d'identifier des anomalies chromosomiques déséquilibrées autrefois identifiées au caryotype avec une bien meilleure résolution permettant ainsi l'identification d'anomalies cryptiques.

L'ACPA est basée sur le principe d'hybridation génomique comparative (CGH) : elle repose sur la comparaison du nombre de molécules d'une même quantité d'ADN entre un patient et un témoin. Ceux-ci sont marqués par deux fluorochromes différents (cyanines) avant d'être déposés sur une lame de verre sur laquelle des oligonucléotides simples brins d'environ 60 à 80 bases et spécifiques de régions génomiques sont fixés (array). Après dépôt sur la lame, les deux ADN vont entrer en compétition pour s'hybrider sur les oligonucléotides présents.

Après hybridation, les signaux générés par les fluorochromes sont numérisés et un ratio d'intensité de fluorescence est calculé au niveau de chaque sonde.

Les résultats sont donnés sous forme graphique où un « point » correspond à la valeur du ratio d'intensité de fluorescence au niveau d'une sonde (Figure 6). L'ensemble des « points » est placé sur un idéogramme pour chaque chromosome. L'existence d'un déséquilibre génomique sera objectivée par une déviation du ratio d'intensité de fluorescence au niveau des sondes fixées sur la lame.

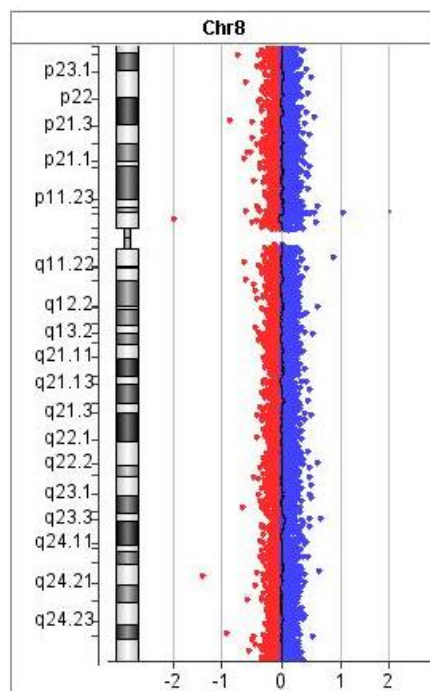


Figure 6: représentation par ratio de fluorescence d'un chromosome 8 en ACPA

Le ratio théorique pour chaque région génomique étudiée est égal à 0. En cas de délétion, le ratio sera de -1, en cas de duplication le ratio sera de 0,58 (Tableau 2).

<b>Ratio d'intensité des fluorochromes sur les autosomes</b>	
<b>Ratio théorique</b>	0
<b>Duplication</b>	0.58
<b>Triplication</b>	$\infty$
<b>Délétion hétérozygote</b>	-1
<b>Délétion homozygote</b>	$-\infty$

*Tableau 2: ratio d'intensité des fluorochromes sur les autosomes en ACPA*

Un traitement statistique des données est réalisé grâce à au logiciel CytoGenomics (Agilent Technologies). Ce traitement est basé sur un algorithme (Aberration Detection Method 2 ou ADM-2) qui permet l'identification des CNV via 3 paramètres principaux (Tableau 3) :

- **Aberration score treshold** : il s'agit du score statistique minimum permettant de considérer un intervalle génomique significatif par rapport aux log-ratios moyens pondérés de l'échantillon et de la référence
- **Log-ratio** : il s'agit de la valeur absolue minimale des log-ratios d'un intervalle génomique permettant d'identifier un CNV
- **Minimum number of probes** : il s'agit du nombre minimal de sondes déviées pour identifier un CNV

La résolution de l'ACPA dépend du nombre d'oligonucléotides fixés sur la lame. Dans notre laboratoire, les lames utilisées permettent une résolution théorique de 30 kb.



Paramètres bio-informatiques	Filtres de routine en ACPA	
	Post-natal Seuil de détection moyen : 60 kb	Pré-natal Seuil de détection moyen : 600 kb
Aberration score treshold	6	6
Log-ratio	0,25	0,17
Minimum number of probes	3	20

Tableau 3: paramètres bio-informatiques des algorithmes permettant l'identification des CNV en ACPA

L'ACPA permet d'identifier une cause moléculaire dans 10 à 15% des déficiences intellectuelles [96,97].

Cependant, l'ACPA ne détecte pas les anomalies chromosomiques équilibrées, les remaniements chromosomiques déséquilibrés en mosaïque dans moins de 20 à 30% des cellules et les régions d'homozygotie [98].

Il est estimé que des remaniements équilibrés impliqués dans le phénotype par interruption d'un gène ou par dérégulation de l'expression d'un gène peuvent être retrouvés sur le caryotype dans 0,78 à 0,8% des cas après une ACPA rendue sans anomalie [99,100].

Les anomalies chromosomiques en mosaïque concerneraient 0,55% des patients présentant des troubles neuro-développementaux [101]. La réalisation d'un caryotype après une ACPA sans anomalie pourrait permettre de détecter une anomalie en mosaïque dans 0,16% des cas [99].

Enfin, les régions d'homozygotie ne sont pas détectées en ACPA. Devant un phénotype évocateur avec une ACPA sans anomalie, il paraît donc important de rechercher ces régions d'homozygotie par étude de la méthylation et/ou SNP array [98]. Il peut ainsi être nécessaire, en fonction du contexte, de réaliser des analyses génétiques complémentaires.

### **III. Le Séquençage Haut Débit**

#### **1) Introduction**

Après la découverte de la structure en double hélice de l'ADN en 1953 [102] et l'établissement du premier caryotype humain en 1956 [103], la génétique moléculaire a connu une véritable révolution avec l'invention du séquençage Sanger en 1977 et de la Polymerase Chain Reaction (PCR) en 1983 [104,105].

The Human Genome Project est un projet collaboratif international qui a permis, entre 1990 et 2003, de séquencer pour la première fois l'intégralité du génome humain en utilisant le principe du séquençage Sanger [106]. Il s'agissait alors d'un véritable défi technologique dont le coût a été estimé à plus de 3 milliards de dollars.

Le Séquençage Nouvelle Génération (NGS) ou Séquençage Haut Débit (SHD) est une technologie développée depuis le début des années 2000 qui permet aujourd'hui un véritable essor du séquençage des génomes humains individuels. Il est ainsi devenu accessible de séquencer un génome humain en 3 jours et pour moins de 1000 dollars.

Cet exploit technologique est rendu possible par le séquençage simultané de millions de fragments d'ADN surpassant ainsi largement les capacités du séquençage Sanger (Tableau 4). Aujourd'hui, le NGS est l'outil de choix pour le diagnostic étiologique génétique de multiples pathologies et permet, en routine, le séquençage de panels de gènes et d'exomes.

Le séquençage Sanger, qui permet une analyse ciblée des gènes d'intérêt, conserve un rôle important dans le cadre des enquêtes familiales et des confirmations des variations détectées par séquençage haut débit.

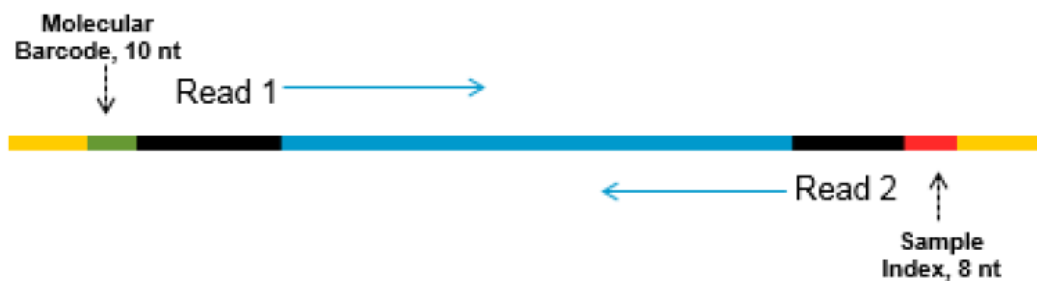
Séquençage Sanger	NGS
<ul style="list-style-type: none"> <li>• Méthode dite par terminaison de chaîne</li> <li>• Synthèse d'un brin d'ADN complémentaire par une ADN polymérase en présence de désoxyribonucléotides (dNTP) et de didésoxyribonucléotides (ddNTP) fluorescents</li> <li>• ddNTP terminateurs : leur incorporation aléatoire entraîne la terminaison de la synthèse du brin complémentaire</li> <li>• Un ddNTP sera statistiquement incorporé à toutes les positions possibles ce qui permet d'obtenir des fragments interrompus de tailles différentes</li> <li>• Migration sur un séquenceur capillaire</li> <li>• Reconstitution de la séquence nucléotidique du brin matrice</li> <li>• Faible quantité de données générées</li> <li>• Pas de découverte incidente</li> </ul>	<ul style="list-style-type: none"> <li>• Sélection des fragments d'intérêt par capture (hybridation)</li> <li>• Amplification clonale sur une lame de verre permettant la séparation spatiale des différents clones</li> <li>• Séquençage simultané par recopiage du brin existant et détection en temps réel de l'émission de fluorescence</li> <li>• dNTP terminateurs réversibles : leur incorporation empêche la fixation d'une autre base ; après clivage enzymatique, l'incorporation d'un autre nucléotide est possible</li> <li>• Un cycle : incorporation d'un dNTP, identification, clivage</li> <li>• Grande quantité de données générées</li> <li>• Etape bio-informatique indispensable : alignement des reads sur le génome de référence, appel de variants</li> <li>• Risque de découverte incidente</li> </ul>

*Tableau 4 : comparaison du séquençage Sanger et du NGS*

## 2) Principe du NGS

Plusieurs méthodes de séquençage existent mais nous traiterons ici uniquement du séquençage par synthèse (technologie Illumina) qui est la méthode utilisée dans notre laboratoire.

Après fragmentation enzymatique de l'ADN du patient pour obtenir des fragments d'ADN d'environ 300 à 400 paires de bases, la ligation d'adaptateurs à l'extrémité des fragments d'ADN va permettre l'hybridation au support de séquençage, l'hybridation des primers de séquençage et l'ajout d'un index qui est un « code-barre » spécifique permettant la reconnaissance de chaque échantillon (Figure 7).



*Figure 7: fragment d'intérêt après la ligation (en jaune, adaptateurs qui permettront l'hybridation au support de séquençage, en rouge l'index spécifique pour chaque patient et en noir les primers de séquençage)*

Les fragments d'ADN d'intérêt sont sélectionnés par capture c'est-à-dire par hybridation avec des milliers d'oligonucléotides biotinylés qui vont assurer la couverture de toutes les régions d'intérêt. Après hybridation, les complexes ADN-oligonucléotides biotinylés vont être récupérés grâce à des billes magnétiques recouvertes de streptavidine, une molécule fortement affine à la biotine. On obtient alors après dénaturation une librairie d'ADN simple-brin prête à séquencer.

Une PCR en phase solide de la librairie ainsi constituée est réalisée avant le séquençage proprement dit. Le support de séquençage est un support en verre nommé « flow cell ». Il comporte 4 lignes sur lesquelles des oligonucléotides complémentaires des fragments d'ADN d'intérêt sont fixés de manière covalente. Cette PCR en phase solide va permettre une amplification monoclonale et localisée qui va créer jusqu'à 5000 molécules identiques de chaque fragment d'intérêt sur un périmètre de l'ordre de 2 microns. Pour cela, l'ADN dénaturé va venir s'hybrider avec l'un des oligonucléotides présents sur le support de séquençage. Les molécules non hybridées vont être lavées puis un nouveau brin est synthétisé par une polymérase.

Une nouvelle dénaturation élimine alors le brin d'origine tandis que le brin nouvellement synthétisé est attaché à la flow cell. Une nouvelle amplification est réalisée durant laquelle l'extrémité libre du brin synthétisé va venir se fixer sur une amorce adjacente complémentaire pour former un pont (bridge PCR). La polymérase synthétise le brin complémentaire formant ainsi un pont d'ADN double-brin.

Une nouvelle dénaturation forme deux copies de la même molécule initiale qui vont elles-mêmes servir de substrat à une nouvelle amplification par formation d'un nouveau pont.

Plusieurs cycles d'amplification sont ainsi réalisés pour permettre l'obtention d'un « cluster » c'est-à-dire une région d'ADN monoclonal et localisé à la surface de la flow cell. Cette amplification est réalisée en parallèle pour tous les fragments d'intérêt et aboutit à la formation de millions de clusters. Les séquences reverse sont éliminées et lavées et l'extrémité 3'OH est bloquée afin d'obtenir un signal pur.

Après le clustering, le séquençage par synthèse commence. Cette technologie utilise des désoxyribonucléotides triphosphates (dNTP) couplés à des fluorochromes et à un terminateur en 3'OH.

Pour le séquenceur MiSeq™ d'Illumina, 4 fluorochromes différents sont couplés aux 4 bases de l'ADN. Pour le séquenceur NextSeq™ d'Illumina, en revanche, seulement 2 fluorochromes sont utilisés. Deux bases sont liées à ces fluorochromes, la troisième correspond à une combinaison de ces deux fluorochromes et la dernière base n'est pas liée à un fluorochrome. L'intensité du signal permet de les différencier.

Lors de chaque cycle de séquençage, les dNTP modifiés sont ajoutés. Le nucléotide incorporé sert de terminaison pour la polymérisation grâce à son terminateur en 3'OH qui empêche une autre base de s'incorporer de manière réversible.


Ainsi, après chaque incorporation d'un dNTP, un laser permet l'excitation du fluorochrome et donc l'identification de la base par émission d'une fluorescence spécifique à chaque nucléotide. Une photographie de l'ensemble des clusters est prise puis le fluorochrome et le terminateur en 3'OH sont clivés de façon enzymatique pour permettre, après un lavage de la flow cell, l'incorporation d'un autre nucléotide.

Un cycle de séquençage correspond donc à l'incorporation d'une base modifiée, à son identification puis au clivage du fluorochrome et du terminateur en 3'OH. Les signaux obtenus sont convertis informatiquement en séquences de bases appelés « reads » et associés à un score de qualité. A la fin du premier read, les brins nouvellement synthétisés sont lavés. L'index est séquençé permettant l'association informatique de chaque cluster au bon patient.

Le séquençage du deuxième read débute et permet le séquençage du brin reverse. L'extrémité 3'OH est libérée permettant son attachement à la flow cell sur une amorce adjacente puis la polymérase synthétise le brin complémentaire. Les ponts ainsi formés sont linéarisés puis le brin forward précédemment séquençé est clivé permettant le séquençage du brin reverse. Ceci permet un séquençage « paired-end » c'est-à-dire le séquençage des deux extrémités du fragment puis l'alignement des reads forward et reverse pour former une paire de reads.

### 3) Analyse bio-informatique

L'analyse bio-informatique est une étape très importante du NGS. Elle comprend plusieurs étapes successives (Tableau 5).



Bio-informatique	Fichier	Logiciel utilisé
Démultiplexage	BCL	Bcl2fastq
Qualité	FASTQ	Alissa Align & Call
Alignement (BWA MEM)	BAM	
Appel de variants	VCF	
Annotation		Alissa Interpret

Tableau 5: étapes successives du traitement bio-informatique des données de NGS

Les données brutes du séquençage sont obtenues en format binaire ayant une extension .BCL (Base Call Files). Ces fichiers contiennent les séquences de bases des reads ainsi que le score de qualité associé. Ils nécessitent une conversion pour pouvoir être utilisés. Le logiciel « bcl2fastq » permet le démultiplexage des échantillons c'est-à-dire l'attribution de chaque read au bon patient. Les adaptateurs et les erreurs de séquençage (biais d'amplification, mismatches, séquences anormalement mappées) vont être éliminés. On obtient ainsi 8 fichiers ayant une extension .FASTQ pour chaque patient.

Ces fichiers sont ensuite téléversés sur la plateforme Alissa Align & Call (Agilent Technologies).

Les reads vont être alignés sur le génome de référence (GRCh37/hg19) c'est-à-dire positionnés dans le bon ordre et au bon endroit pour reconstituer le génome du patient. L'alignement permet d'obtenir un fichier ayant une extension .BAM (Binary Alignment Map) contenant pour chaque read sa séquence nucléotidique, les scores de qualité de ses bases et les coordonnées génomiques correspondant à son alignement.

Deux notions très importantes permettent d'évaluer l'efficacité du séquençage (Figure 8) :

- La profondeur correspond au nombre de lecture de chaque base exprimée en X.
- La couverture correspond aux nombres de bases séquencées des régions d'intérêt par rapport au nombre de bases totales de ces régions. Ce résultat est exprimé en pourcentage pour une profondeur donnée.

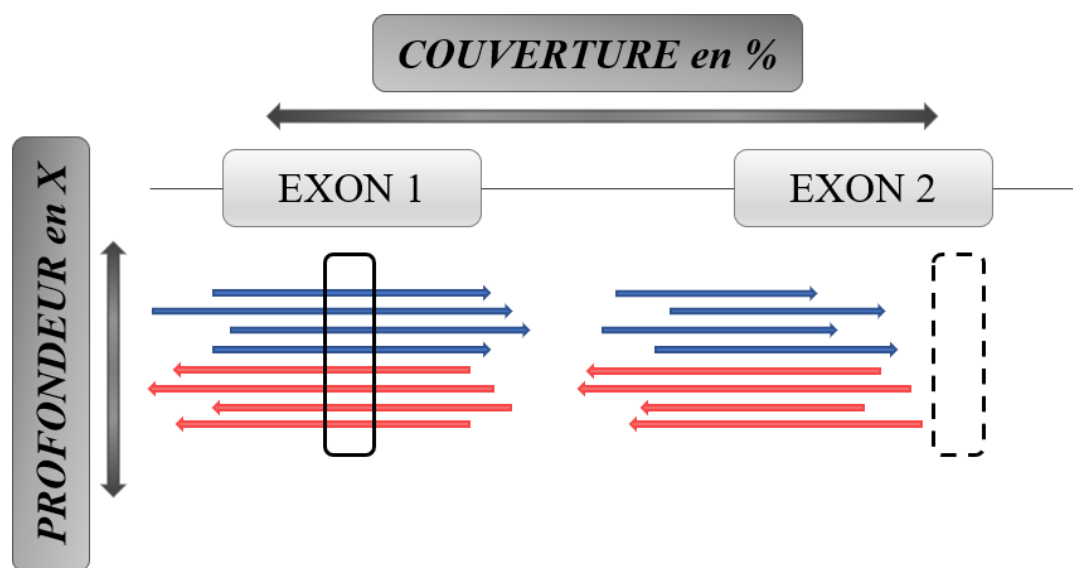


Figure 8: représentation schématique des notions de profondeur et de couverture de séquençage. Chaque flèche bleue représente une lecture du brin forward, chaque flèche rouge représente une lecture du brin reverse.

L'exon 1 est ici couvert à 100% (encadré noir : couverture de 100% à une profondeur de 8X)

L'exon 2 n'est pas couvert à 100% (encadré en pointillé : zone terminale non couverte)

L'étape d'appel des variants permet de détecter les variations de séquence grâce à des algorithmes spécifiques. Les bases de l'ADN du patient sont comparées une par une au génome de référence pour mettre en évidence une variation de séquence (substitution d'une base, délétion ou insertion d'une ou plusieurs bases ; Figure 9). Un fichier ayant une extension .VCF (Variant Call Format) est alors obtenu.

La détection des délétions et duplications par NGS est réalisée par analyse de la profondeur de séquençage. Pour détecter ces CNV, en théorie, le nombre de reads alignés sur une région génomique est proportionnel au nombre de fois où cette région est représentée. En cas de délétion, il y aura deux fois moins de reads alignés sur cette région alors qu'en cas de

duplication, il y en aura deux fois plus (Figure 9). Le séquençage n'étant pas uniforme sur l'ensemble du génome, les données de séquençage sont normalisées par rapport à une référence afin de pouvoir identifier avec fiabilité les CNV.

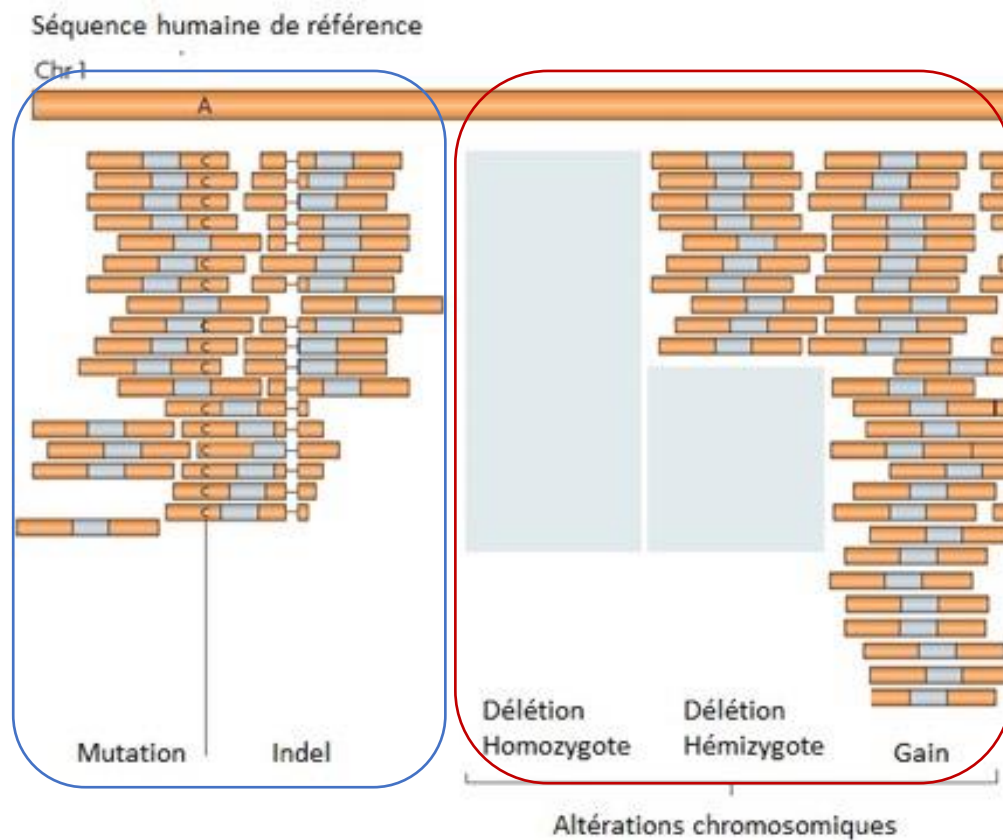


Figure 9: détection des variants (encadré bleu) et des délétions et duplications chromosomiques (encadré rouge) par NGS, d'après Meyerson et al., 2010 [107]

Enfin, l'étape analytique proprement dite débute avec l'annotation des variations détectées sur la plateforme Alissa Interpret (Agilent Technologies).

Le biologiste, en fonction de différents paramètres tels que la fréquence en population générale, le gène impliqué, la localisation du variant, le type de mutation, les bases de données de patients, l'impact sur la structure protéique, les scores des logiciels de prédiction de pathogénicité et la littérature existante, va venir attribuer un score allant de 1 à 5 au variant en fonction de la pathogénicité présumée (1 : bénin, 2 : probablement bénin, 3 : signification inconnue, 4 : probablement pathogène, 5 : pathogène)



#### 4) OneSeq Constitutional Research Panel

Compte tenu de l'hétérogénéité clinique et étiologique de la DI, nous nous sommes demandé si une technique permettait, à l'heure actuelle, de détecter les anomalies chromosomiques et géniques.

Nous nous sommes ainsi intéressés à une technologie de NGS basée sur le kit de capture OneSeq Constitutional Research Panel du fournisseur Agilent Technologies.

Cette technologie innovante est une nouvelle approche pour identifier les causes moléculaires chromosomiques et géniques de DI. Elle permet, au cours d'une technique unique, de rechercher les variants nucléotidiques, de détecter les délétions et duplications chromosomiques et d'identifier les régions d'homozygotie.

Pour cela, cette technologie combine trois séries de sondes permettant de détecter les mutations ponctuelles ou Single Nucleotide Variation (SNV), les délétions ou duplications chromosomiques (CNV) et les régions d'homozygotie (Loss Of Heterozygosity - LOH) (Figure 10).



Figure 10: design des différentes sondes du OneSeq Constitutional Research Panel

En vert : sondes ciblant les exons de 5227 gènes impliqués en pathologie (SNV)

En bleu : sondes « backbone » réparties sur tout le génome (CNV)

En rouge : sondes ciblant les régions génomiques avec SNP fréquents (LOH)

#### **Les sondes de capture exoniques :**

Ces sondes de capture sont disséminées dans les séquences codantes de 5227 gènes connus en pathologie et sont représentées en vert sur la figure 10. Il s'agit d'un exome dit « clinique ».

Ces sondes permettent la détection des mutations et des insertions/délétions à l'échelle du gène (SNV). Ces sondes ciblent notamment 1641 gènes potentiellement impliqués dans la DI selon la base de données SysID (liste en Annexe II).

### **Les sondes « backbone » intergénomiques pangénomiques :**

Ces sondes sont réparties uniformément et à intervalles réguliers sur le génome. Elles sont représentées en bleu sur la figure 10. Elles permettent de détecter les CNV par étude de la profondeur de séquençage avec une résolution moyenne de 300 kb. La résolution sur les gènes impliqués en pathologie (bases de données ClinGen/ISCA) est de 25 à 50 kb.

L'analyse chromosomique est réalisée par comparaison avec un échantillon de référence, sur le même principe que l'ACPA. Ces sondes ont été désignées de façon à cibler les régions génomiques permettant une profondeur de lecture uniforme garantissant la fiabilité des CNV détectés.

### **Les sondes ciblant les Single Nucleotide Polymorphism :**

Ces sondes ciblent les régions génomiques dans lesquelles la fréquence des Single Nucleotide Polymorphism (SNP) est élevée et sont représentées en rouge sur la figure 10. Elles permettent la détection des régions d'homozygotie. Ces sondes vont mettre en évidence les disomies uniparentales et les régions d'homozygotie en relation avec des situations de consanguinité.

La détection des régions d'homozygotie est fondée sur une mesure normalisée nommée « B-Allele Frequency » (BAF) ou « fréquence de l'allèle B » qui permet de visualiser le ratio de deux allèles, A et B. Une valeur de BAF à 1 ou 0 indique l'absence complète de l'un ou l'autre de ces allèles et donc une homozygotie. Une BAF à 0,5 indique une présence égale de chaque allèle ce qui correspond à un statut d'hétérozygote sur la région génomique observée.

Ainsi, la BAF permet de détecter les régions d'homozygotie à condition qu'il s'agisse d'une isodisomie, c'est-à-dire d'une transmission de deux copies du même chromosome parental et donc portant les mêmes SNP. Une hétérodisomie ne sera pas détectée car elle résulte de la transmission d'une paire de chromosomes homologues issus du même parent et ayant des SNP différents. Elle sera donc faussement considérée comme une région d'hétérozygotie sauf si l'analyse génétique des parents est possible [108].

## 5) Objectif de l'étude

La technologie OneSeq pourrait être une alternative à l'association de l'ACPA et de l'étude d'un panel de gènes en première intention pour le diagnostic des déficiences intellectuelles.

Cette technologie permet en effet de combiner la détection des anomalies génomiques quantitatives et qualitatives qui sont des causes fréquentes de DI et qui sont parfois associées. Leur détection simultanée permettrait ainsi de diminuer le délai et le coût analytique.

Nous avons souhaité évaluer les performances de cette technologie NGS pangénomique pour la détection des anomalies chromosomiques par comparaison avec l'ACPA.

Compte tenu de l'hétérogénéité génétique de la DI, il est important d'analyser des anomalies génétiques classiques mais également atypiques. Pour cela, nous avons étudié 22 patients présentant diverses anomalies du développement avec un génotype connu pour évaluer les performances de cette technologie. Nous avons ensuite analysé 8 patients en aveugle pour évaluer son applicabilité en routine diagnostique.

# MATERIEL ET METHODES

## I. Recrutement des patients

Nous avons recruté 22 patients présentant diverses anomalies du développement en pré et post-natal pour lesquels une analyse chromosomique avait été réalisée (ACPA, QF-PCR, analyses des marqueurs microsatellites et/ou caryotype)(Tableau 6).

Chez 15 patients sur 22, une ACPA a été réalisée préalablement à notre étude.

Parmi eux, certains présentent des anomalies chromosomiques fréquentes (délétion 22q11.21, dérivé de translocation...) de taille relativement importante, et d'autres des anomalies de signification inconnue, en mosaïque ou de petite taille.

Parmi les 7 autres patients recrutés, 2 présentent des variants géniques, 2 patients ont une aneuploïdie qui a été diagnostiquée par Quantitative Fluorescent PCR et 3 patients ont des régions d'homozygotie confirmées ou suspectées.

L'analyse de ces échantillons permet de tester les performances de cette nouvelle technologie dans l'identification de différents types d'anomalies chromosomiques et géniques dont les régions d'homozygotie.

Les 8 patients analysés en aveugle ont été sélectionnés par un opérateur indépendant et numérotés de 23 à 30. Ils ont ensuite été analysés indépendamment sur la plateforme Alissa Interpret (Agilent Technologies) par deux opérateurs différents. Les variants identifiés ont ensuite été comparés pour chaque patient aux variants identifiés en ACPA.

<b>Cas n°1</b>	16 ans	Syndrome de Di George avec phénotype sévère : calcifications cérébrales, malformation vertébrale, double scoliose	<b>Cas n°12</b>	3 ans	Troubles de l'équilibre, déficience intellectuelle
<b>Cas n°2</b>	33 ans	Epilepsie	<b>Cas n°13</b>	15 ans	Ichtyose sévère, obésité, retard pubertaire
<b>Cas n°3</b>	16 ans	Dysmorphie, trouble du spectre autistique, petite taille	<b>Cas n°14</b>	22 SA	Marqueurs sériques du deuxième trimestre à 1/347 - échec de 2 DPNI
<b>Cas n°4</b>	4 mois	Retour veineux pulmonaire anormal	<b>Cas n°15</b>	13 SA	Holoproencéphalie
<b>Cas n°5</b>	9 ans	Déficience intellectuelle, troubles du spectre autistique	<b>Cas n°16</b>	24 SA	PAPP-A et HCG abaissés
<b>Cas n°6</b>	1 an	Hypotonie néonatale, atrésie choanale, cardiopathie, suspicion de syndrome CHARGE	<b>Cas n°17</b>	23 ans	Déficience intellectuelle, dysmorphie, scoliose
<b>Cas n°7</b>	4 ans	Déficience intellectuelle	<b>Cas n°18</b>	1 an et 3 mois	Syndrome de West
<b>Cas n°8</b>	7 ans	Retard de développement, dysmorphie	<b>Cas n°19</b>	1 an et 9 mois	Déficience intellectuelle, hypotonie, dysmorphie faciale, RCIU, cardiopathie, signe de Peters
<b>Cas n°9</b>	11 ans	Retard des acquisitions, troubles du comportement	<b>Cas n°20</b>	2 mois	Hypotonie axiale majeure, troubles alimentaires, dysmorphie, syndrome de Prader-Willi
<b>Cas n°10</b>	9 jours	Hypotonie globale, hypoglycémies profondes, hyperinsulinisme, thrombopénie, infection materno- fœtale	<b>Cas n°21</b>	1 an et 9 mois	Epilepsie, déficience intellectuelle, retard des acquisitions, suspicion de syndrome d'Angelman
<b>Cas n°11</b>	12 ans	Encéphalopathie épileptique	<b>Cas n°22</b>	24 SA	Mains crispées, pieds bots

*Tableau 6: patients étudiés et indications ayant conduit à une analyse chromosomique*

## **II. Extraction des ADN**

L'ADN des patients a été extrait soit à partir de sang périphérique sur tube EDTA, soit à partir d'un prélèvement de liquide amniotique (patients 14, 16 et 22) ou de villosités choriales (patient 15) après signature d'un consentement éclairé permettant la recherche d'anomalies chromosomiques.

Les prélèvements ont été extraits de façon automatisée avec l'extracteur d'ADN QIASymphony SP<sup>TM</sup> (Qiagen) en utilisant le DNA Midi Kit pour les prélèvements sanguins et le DNA Mini Kit pour les prélèvements prénataux. Après une étape de lyse cellulaire, l'ADN présent dans l'échantillon s'est lié à des particules de silice. Après lavage et rinçage à l'éthanol, les acides nucléiques étaient élués dans une solution aqueuse.

La quantité d'ADN extrait a été obtenue par une méthode de fluorométrie en utilisant le Qubit<sup>TM</sup> (ThermoFisher Scientific).

## **III. Protocole SureSelect XT HS**

Le protocole choisi pour réaliser notre étude est le SureSelect XT HS d'Agilent Technologies. Ce protocole permet de préparer 8 échantillons pour le séquençage au cours de 2 jours de technique. Le protocole détaillé utilisé est présenté en Annexe III.

Les bibliothèques ont été préparées avec le OneSeq Constitutional Research Panel d'Agilent Technologies en suivant les recommandations du fournisseur.

La bibliothèque finale était ensuite chargée sur le séquenceur NextSeq 550 avec une cassette High Output qui permet d'obtenir environ 100 gigabytes de données et de générer 400 millions de clusters. Le séquençage durait environ 24 heures.

A la fin du séquençage, les échantillons étaient démultiplexés avec le logiciel « bcl2fastq » puis les fichiers .FASTQ étaient téléversés sur la plateforme Alissa Align & Call (Agilent Technologies) pour l'alignement et l'appel des variants.

## **IV. Optimisation des paramètres bio-informatiques pour détecter les CNV**

L'analyse bio-informatique est indispensable pour traiter les données générées et pour optimiser la détection des anomalies génétiques.

Certaines étapes de l'analyse bio-informatique, comme le démultiplexage et l'alignement, sont dépendantes d'algorithmes ne pouvant pas être modifiés à notre niveau (Tableau 5).

En revanche, l'étape d'appel des variants peut être optimisée par modification de certains paramètres bio-informatiques permettant une détection plus ou moins stringente des anomalies chromosomiques ou géniques.

Dans ce travail, nous nous sommes intéressés à l'optimisation des paramètres bio-informatiques liés à la recherche des anomalies chromosomiques (CNV).

Les paramètres liés à la recherche des anomalies géniques (SNV) restent ceux préconisés par le fournisseur. En effet, ces paramètres ont déjà été jugés performants dans notre laboratoire lors de la mise en place de l'exome clinique.

Concernant l'optimisation des paramètres bio-informatiques permettant la recherche des CNV, nous avons décidé de créer deux filtres bio-informatiques différents : l'un permettant la recherche des anomalies chromosomiques homogènes et l'autre dédié à la recherche des anomalies chromosomiques en mosaïque. En effet, les anomalies en mosaïque sont plus difficiles à diagnostiquer en biologie moléculaire que les anomalies homogènes.

Par exemple, les mosaïques inférieures à 20% ne sont pas détectables en ACPA dans notre laboratoire. Pour détecter des mosaïques entre 20 et 30%, un filtre spécial avec des paramètres bio-informatiques plus permissifs est utilisé en ACPA dans certains contextes précis. En effet, ces paramètres ne sont pas utilisables en routine en raison d'une permissivité trop importante entraînant la détection d'un grand nombre de CNV artéfactuels.

La sensibilité attendue du NGS pour la détection des mosaïques est similaire à celle de l'ACPA. Ainsi, nous avons optimisé un filtre dédié à la détection des mosaïques pour le OneSeq.

L'optimisation de la détection des CNV a été réalisée en modulant 4 paramètres bio-informatiques principaux modifiant l'appel des variants (Tableau 7) :

- **Aberration score treshold** : il s'agit du score statistique minimum permettant à l'algorithme de considérer un intervalle génomique significatif par rapport au bruit de fond présent dans les log-ratios de l'échantillon et de la référence
- **Log-ratio** : il s'agit de la valeur absolue minimale des log-ratios pour identifier un CNV, comme pour l'ACPA
- **Minimum number of points** : il s'agit du nombre minimal de points déviés de façon significative pour identifier un CNV
- **Minimum read depth filter on reference for calculating ratios** : il s'agit de la profondeur minimale nécessaire sur la référence pour permettre le calcul des log-ratios utilisés pour identifier les CNV. Ce paramètre va notamment être important pour identifier les délétions homozygotes.

Un CNV est donc identifié lorsque 3 points consécutifs sont déviés par rapport au seuil défini (log-ratios) et lorsque le score statistique (aberration score treshold) est supérieur au seuil de significativité.

<b>Paramètres bio-informatiques principaux</b>	<b>Valeurs possibles</b>	<b>Filtre par défaut proposé par Agilent</b>
<b>Aberration score treshold</b>	0 - 50	6
<b>Log-ratio</b>	0 - 1	0,25
<b>Minimum number of points</b>	1 - 10	5
<b>Minimum read depth filter on reference</b>	1 - 100	40

*Tableau 7: paramètres bio-informatiques du filtre proposé par le fournisseur*

Pour identifier tous les variants testés sans augmenter de façon importante le nombre d'artéfacts, nous avons testé 17 filtres en modifiant à chaque fois un ou deux paramètres afin d'utiliser pour chaque variable le paramètre le plus adapté.

Ainsi, les deux filtres choisis ont les caractéristiques présentées dans le Tableau 8.



<b>Paramètres bio-informatiques principaux</b>	<b>Filtre de routine</b> (Détection de la plupart des anomalies chromosomiques déséquilibrées)	<b>Filtre mosaïque</b> (Détection des anomalies en mosaïque)
<b>Aberration score treshold</b>	6	5
<b>Log-ratio</b>	0,2	0,17
<b>Minimum number of points</b>	3	10
<b>Minimum read depth filter on reference</b>	30	40

*Tableau 8: filtres optimisés pour la détection des anomalies chromosomiques homogènes et en mosaïque*

La vérification de la performance de ces deux filtres a été réalisée par l'analyse de 22 patients à génotype connu puis par analyse de 8 patients en aveugle.

## **V. Tri des CNV**

Pour l'analyse en aveugle, les CNV, SNV et LOH ont été triés grâce à des arborescences comprenant différents filtres successifs.

De façon schématique, les CNV sont filtrés en fonction de différents critères (présence en population générale, classification dans notre base de données interne et dans les bases de données nationales et internationales, taille des CNV, gènes inclus...).

Les LOH sont séparés des CNV puis filtrés en fonction de leur présence dans des régions chromosomiques soumises à empreinte ou non.

Les SNV sont également triés en fonction de plusieurs critères (fréquence en population générale, présence dans notre base de données interne et dans les bases internationales, impact de la variation sur la protéine, indications des logiciels de prédiction, phénotype du patient...)

# RESULTATS

## I. Patients avec un génotype connu et analyse ouverte

### 1) Comparaison du OneSeq avec l'ACPA

Le OneSeq a permis d'identifier les variants d'intérêt chez 13 patients sur 15, soit 87% des anomalies diagnostiquées par ACPA (Tableau 9).

Le chevauchement moyen des coordonnées génomiques entre le OneSeq et l'ACPA est de 93%. Le chevauchement médian est de 94% avec un minimum à 80% et un maximum à 99%. En moyenne, 201 CNV ont été détectés par patient avec le OneSeq contre 10 à 15 pour l'ACPA.

Les représentations des chevauchements des coordonnées génomiques ont été réalisées avec l'University of California Santa Cruz Genome Browser (<http://genome.ucsc.edu/>), génome de référence GRCh37/hg19. Celles-ci permettent de comparer les intervalles génomiques délétés et dupliqués obtenus en ACPA et en OneSeq. Le calcul des chevauchements a été réalisé sur la plateforme Cartagenia de Agilent Technologies.

Les résultats des cas n°1, 2, 4, 11 et 12 sont représentés ci-après (Figures 11, 12, 13, 14 et 15)

Les résultats des cas n°5, 8, 9, 13, 14, 16, 18 et 19 sont présentés en Annexe I.

Patient	Technique	Coordonnées génomiques (GRCh37/hg19)	Log-ratio	Taille (kb)	Chevauchement coordonnées génomiques OneSeq/ACPA
Cas n°1	ACPA	22q11.21(18894835_21464119)x1	-0,83	2 569,3 kb	99%
	ONESEQ	22q11.21(18874807_21468808)x1	-0,85	2 594 kb	
Cas n°2	ACPA	17p11.2(17630064_22154574)x3	0,29	4 524,5 kb	91%
	ONESEQ	17p11.2(18321433_22198630)x3	0,59	3 877,2 kb	
Cas n°3	ACPA	12p12.1(23793692_23955771)x3	0,59	162,1 kb	Anomalie non détectée
	ONESEQ	Duplication non détectée	/	/	
Cas n°4	ACPA	15q11.2(22765628_23208901)x1	-0,71	443,3 kb	88%
	ONESEQ	15q11.2(22650839_23207016)x1	-1,52	556,2 kb	
Cas n°5	ACPA	8p22(15488829_15721013)x0	-2,14	232,2 kb	84%
	ONESEQ	8p22(15425544_15743696)x0	-2,22	318,1 kb	
Cas n°6	ACPA	19p13.2(11159389_11199893)x1	-0,88	40,5 kb	Anomalie non détectée
	ONESEQ	Délétion non détectée	/	/	
Cas n°8	ACPA	5q12.1(59660840_59895067)x3	0,5	234,2 kb	80%
	ONESEQ	5q12.1(59629955_59974168)x3	0,72	344,2 kb	

<b>Cas n°9</b>	ACPA	1q44(245427067_249212668)x3 21q22.3(45324065_48090317)x1	0,51 -0,80	3 785,6 kb 2 766,3 kb	<b>99%</b>
	ONESEQ	1q44(245470070_249174770)x3 21q22.3(45342290_48102424)x1	0,63 -0,85	3 704,7 kb 2 726,5 kb	
<b>Cas n°11</b>	ACPA	15q11.2q13.2(20102541_30334911)x4 15q13.2q13.3(30652489_32510863)x3	0,89 0,50	10 232,3 kb 1 858,4 kb	<b>93%</b>
	ONESEQ	15q11.2q13.2(20710631_30654161)x4 15q13.2q13.3(30662389_32676282)x3	0,88 0,63	9 943,5 kb 2 013,8 kb	
<b>Cas n°12</b>	ACPA	18p11.32p11.22(118760_10809101)x3 18q23(73361697_78010032)x1	0,53 -0,91	10 690,3 kb 4 648,3 kb	<b>94%</b>
	ONESEQ	18p11.32p11.22(41148_10785145)x3 18q23(74352551_78013015)x1	0,60 -0,90	10 743,9 kb 3 660,4 kb	
<b>Cas n°13</b>	ACPA	Xp22.31(6489877_8131810)x0	-2,92	1 641,9 kb	<b>96%</b>
	ONESEQ	Xp22.31(6432834-8178738)x0	-7,95	1 745,9 kb	
<b>Cas n°14</b>	ACPA	12p13.33p11.1(163593_34398316)x4	0,75	34 234,7 kb	<b>98%</b>
	ONESEQ	12p13.33p11.1(190375_33287942)x4	0,94	33 097,5 kb	
<b>Cas n°16</b>	ACPA	Xp22.33p11.21(2701273_57704961)x1~2 Xq11.1q12(61998756_65953299)x1~2 Xq21.1q28(76024760_155097214)x1~2	-0,36 -0,3 -0,56	55 003,7 kb 3 954,5 kb 79 072,5 kb	<b>94%</b>
	ONESEQ	Xp22.32q13.1(4277390_68127730)x1~2 Xq13.1(68916740_69918545)x1~2 Xq13.2(72150111_73135173)x1~2 Xq13.2q28(73353354_154822924)x1~2	-0,64 -0,35 -0,36 -0,97	63 850,3 kb 1 001,8 kb 985 kb 81 469,5 kb	
<b>Cas n°18</b>	ACPA	Yp11.32q12(11091_59326866)x2	0,60	59 315,8 kb	<b>99%</b>
	ONESEQ	Yp11.32q12(10328_58986673)x2	0,98	58 258,4 kb	

<b>Cas n°19</b>	ACPA	Xp22.33q28(61091_155232907)x4~5	1,05	151 291 kb	<b>99%</b>
	ONESEQ	Xp22.33q28(60328_155257985)x4~5	1,23	150 394 kb	

*Tableau 9: comparaison des résultats du OneSeq par rapport à l'ACPA en analyse ouverte*

## Cas n°1 : délétion 22q11.21, chevauchement de 99%

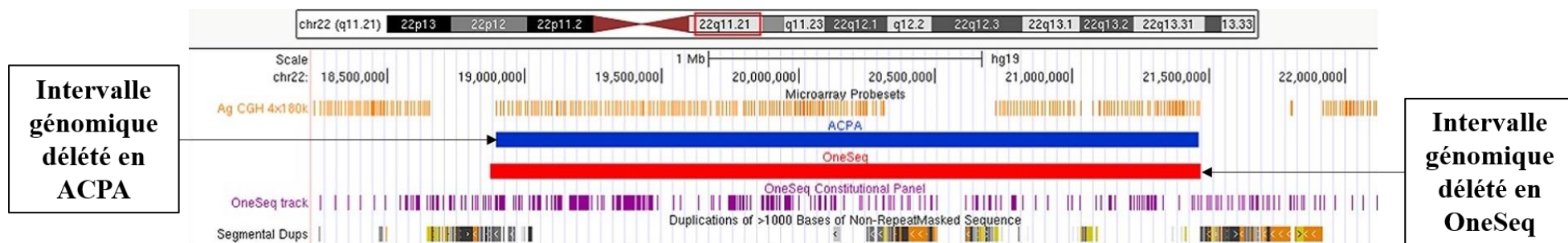


Figure 11 : délétion 22q11.21 : comparaison des coordonnées génomiques du OneSeq avec l'ACPA, chevauchement de 99% (en orange : sondes de l'ACPA, en violet : sondes du OneSeq)

## Cas n°2 : duplication 17p11.2, chevauchement de 91%

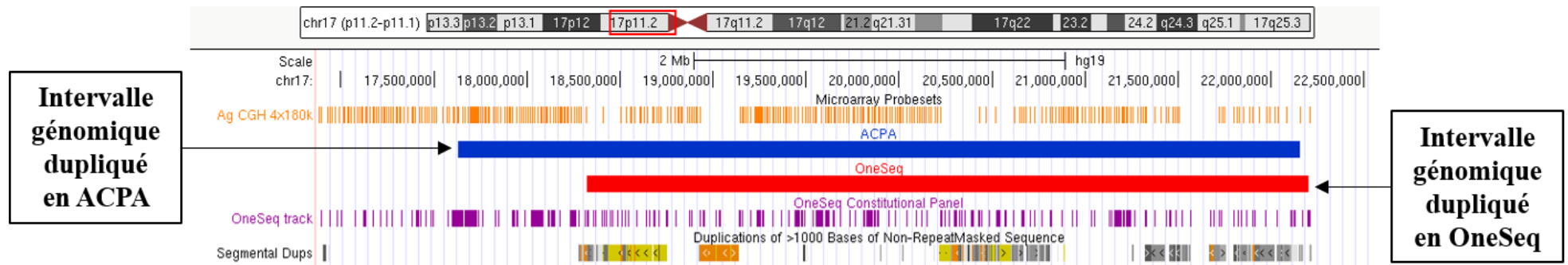


Figure 12: duplication 17p11.2 : comparaison des coordonnées génomiques du OneSeq avec l'ACPA, chevauchement de 91% (en orange : sondes de l'ACPA, en violet : sondes du OneSeq)

## Cas n°4 : délétion 15q11.2, chevauchement de 88%

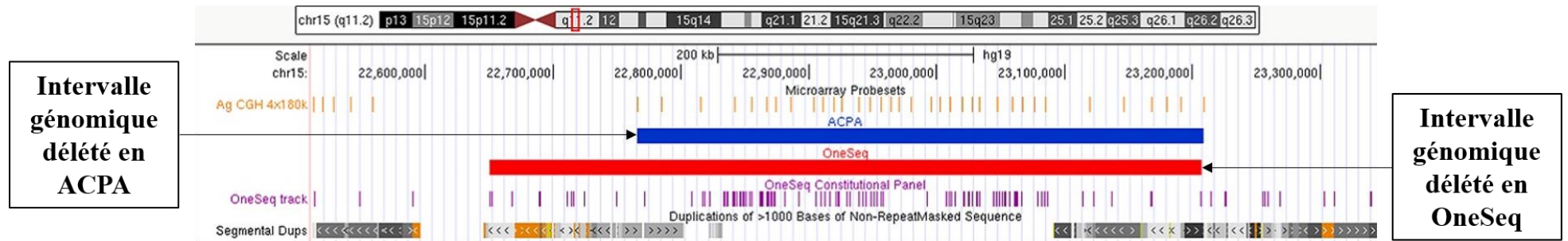


Figure 13: délétion 15q11.2 : chevauchement des coordonnées génomiques du OneSeq avec l'ACPA, chevauchement de 88% (en orange : sondes de l'ACPA, en violet : sondes du OneSeq)



## Cas n°11 : isodicentrique 15q surnuméraire, chevauchement de 93%

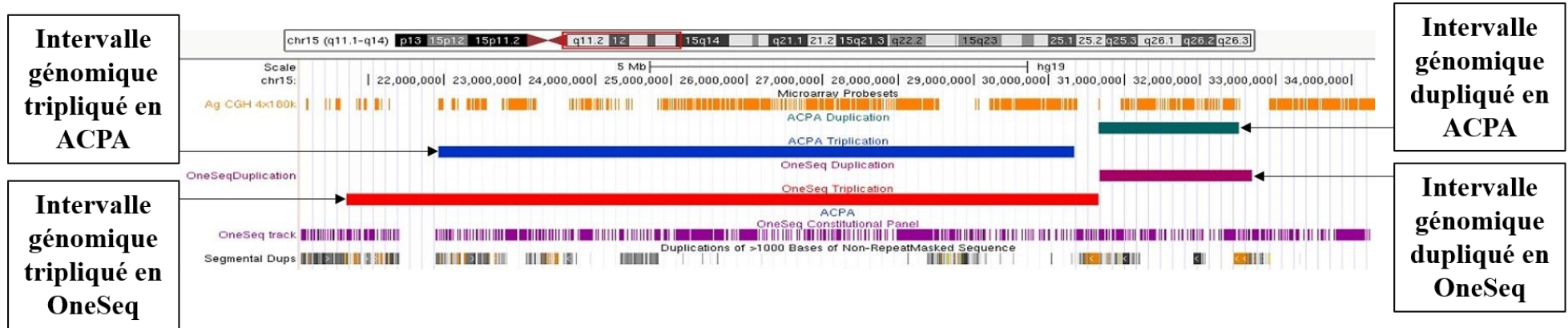


Figure 14: isodicentrique 15q surnuméraire : comparaison des coordonnées génomiques du OneSeq avec l'ACPA, chevauchement de 93% (en orange : sondes de l'ACPA, en violet : sondes du OneSeq)

## Cas n°12 : chromosome 18 recombinant, chevauchement de 94%

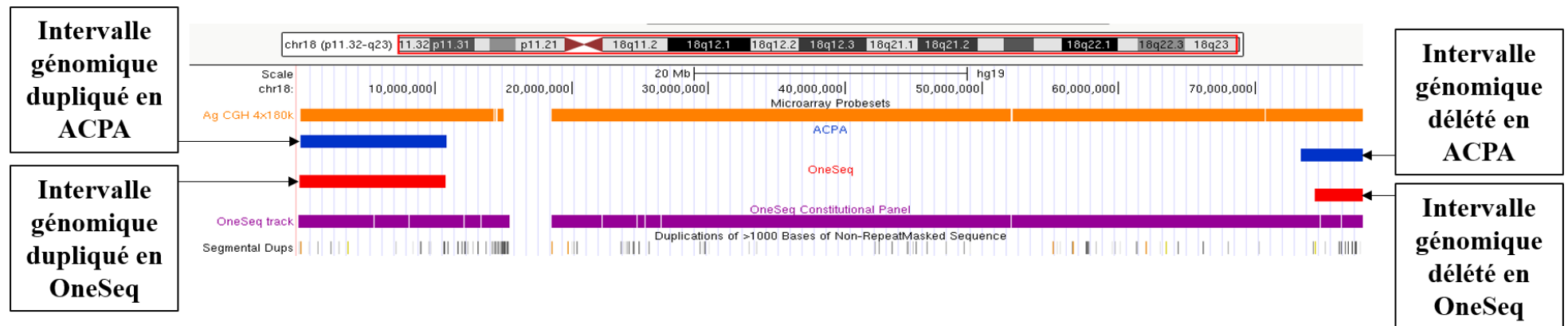


Figure 15: chromosome 18 recombinant : comparaison des coordonnées génomiques du OneSeq avec l'ACPA, chevauchement de 94% (en orange : sondes de l'ACPA, en violet : sondes du OneSeq)

## 2) Détection des régions d'homozygotie avec le OneSeq

Deux disomies uniparentales diagnostiquées par marqueurs microsatellites ont été choisies pour tester les performances du OneSeq pour la détection des régions d'homozygotie (Tableau 10).

Dans notre étude, nous avons choisi 2 patients ayant une disomie uniparentale détectée par marqueurs microsatellites du chromosome 15.

Chez le patient n°7, le OneSeq a identifié une région d'homozygotie de la région critique du syndrome d'Angelman qui témoigne d'une isodisomie paternelle.

Chez le patient n°20 atteint d'un syndrome de Prader-Willi, le OneSeq n'a pas détecté d'homozygotie de la région 15q11.2. Ce résultat est en accord avec celui des marqueurs microsatellites du chromosome 15 qui confirment une hétérodisomie maternelle et donc indétectable en OneSeq.

Un troisième patient (n°21) ayant une suspicion clinique de syndrome d'Angelman avec ACPA sans anomalie et profil de méthylation normal a également été analysé en OneSeq. Aucune délétion ou LOH de la région critique du syndrome d'Angelman et aucun variant du gène *UBE3A* n'ont pu être mis en évidence. Ces résultats semblent remettre en cause la suspicion clinique.

Patient	Technique	Résultat	Taille (kb)	Diagnostic
Cas n°7	ACPA	ACPA rendue sans anomalie	/	Non
	Marqueurs microsatellites du 15	3 marqueurs microsatellites significatifs	/	Syndrome d'Angelman par Isodisomie Uniparentale Paternelle du chromosome 15
	ONESEQ	LOH 15q11.2q13.2(22868300_30366246) 15q13.3(31255902_33359574)	7 497,9 kb 2 103,6 kb	Syndrome d'Angelman par Isodisomie Paternelle du chromosome 15
Cas n°20	ACPA	Non réalisée	/	/
	Marqueurs microsatellites du 15	3 marqueurs microsatellites significatifs	/	Syndrome de Prader-Willi par Hétérodisomie Uniparentale Maternelle du chromosome 15
	ONESEQ	Absence de LOH de la région 15q	/	Non
Cas n°21	ACPA	ACPA rendue sans anomalie	/	Non
	ONESEQ	Absence de LOH Absence de variant du gène <i>UBE3A</i>	/	Suspicion de Syndrome d'Angelman non confirmée

Tableau 10: analyse des performances du OneSeq pour la détection des LOH

### 3) Détection des aneuploïdies diagnostiquées par QF-PCR avec le OneSeq

Le patient n°15 présentait une trisomie 13 homogène diagnostiquée par Quantitative Fluorescent PCR (QF-PCR). Celle-ci a été identifiée avec une concordance de 100% par le OneSeq (Tableau 11).

Patient	Technique	Coordonnées génomiques	Log-ratio	Taille (kb)	Diagnostic
Cas n°15	QF-PCR	13q14.11, 13q12.13, 13q21.33, 13q31.1, 13q13.3	NA	NA	Trisomie 13
	ONESEQ	13q11q34(19168098_115064594)x3	0,49	95 896,4 kb	

Tableau 11: trisomie 13 homogène diagnostiquée par QF-PCR

Le patient n°10 présentait une trisomie 13 en mosaïque à 23%. Celle-ci avait été suspectée par QF-PCR puis diagnostiquée par FISH avec évaluation du pourcentage de la mosaïque.

Une ACPA a également été réalisée et ne permettait pas la détection de cette anomalie en mosaïque avec les paramètres utilisés en routine diagnostique. Cependant, l'anomalie pouvait être suspectée en raison de la déviation du log-ratio des sondes situées sur le chromosome 13 (Figure 16).

En revanche, en diminuant le log-ratio pour l'adapter à la recherche des mosaïques, cette anomalie était alors détectée.

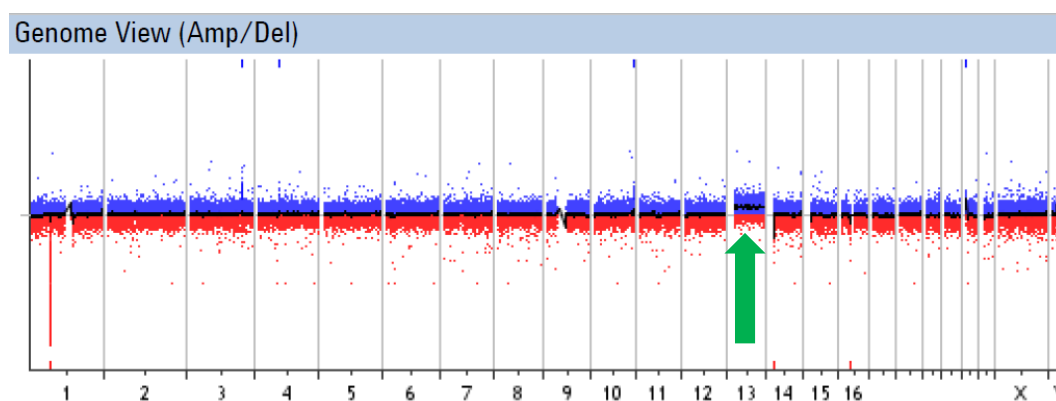


Figure 16: trisomie 13 en mosaïque faible suspectée en ACPA devant une déviation de la ligne de base des log-ratios (flèche verte)

Dans le cas de cette trisomie 13 en mosaïque, il est retrouvé en ACPA comme en OneSeq une hybridation en excès de l'ADN du patient sur la puce à la place de la référence.

Les log-ratios des sondes présentes sur le chromosome 13 sont proches de 0,20 ce qui est caractéristique d'une anomalie en mosaïque.

La sensibilité de l'ACPA et du OneSeq pour la détection des mosaïques semble donc similaire (Figures 17 et 18).

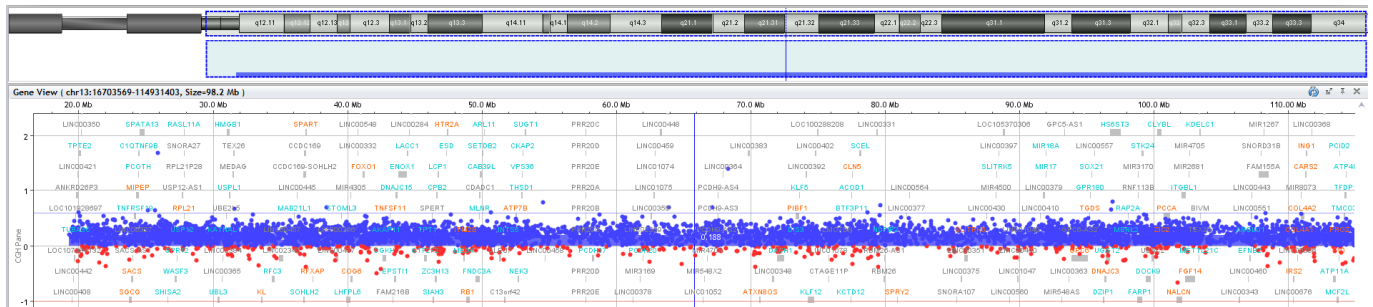


Figure 17: ACPA : représentation des sondes du chromosome 13 ; les sondes dupliquées sont indiquées en bleu (Image issue du logiciel CytoGenomics - Agilent Technologies)

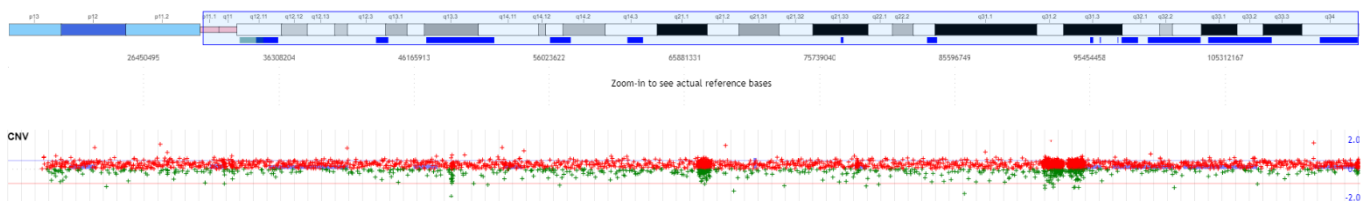


Figure 18: OneSeq : représentation des sondes du chromosome 13 ; les sondes dupliquées sont indiquées en rouge (Image issue de la plateforme Alissa Align & Call - Agilent Technologies)

Le filtre de routine du OneSeq permettait de suspecter la trisomie 13 par détection de plusieurs régions génomiques dupliquées sur ce chromosome mais de façon moins importante que le filtre paramétré pour détecter les mosaïques.

#### 4) Détection des variants géniques avec le OneSeq

Nous avons également souhaité vérifier que le OneSeq détectait correctement les variations géniques. L'objectif principal de l'étude étant de comparer les performances du OneSeq par rapport à l'ACPA pour la détection des CNV, nous n'avons pu tester que deux patients présentant des mutations connues.

Les données qualités du séquençage sont comparables à celles générées par l'exome clinique réalisé dans le cadre de la DI au laboratoire et pour lequel la profondeur moyenne est de 167X et la couverture moyenne à 20X de 95,32%. En OneSeq, sur l'ensemble des patients, la profondeur moyenne était de 309X et la couverture moyenne à 20X de 96,43%.

Ces deux techniques détectent un nombre équivalent de variants.

Pour les deux échantillons présentant un variant génique, les variants pathogènes recherchés ont été détectés par le OneSeq avec une profondeur correcte (Tableau 12).

Patient	Profondeur	Mutation	Protéine	Gène
Cas n°17	48	c.5958C>T	p.Gly1986=	<i>KMT2A</i>
Cas n°22	457	c.858C>G	p.Cys286Trp	<i>CHST14</i>

Tableau 12: analyse en OneSeq de 2 patients présentant des variants géniques

## **II. Patients avec un génotype connu et analyse en aveugle par deux opérateurs**

Les anomalies chromosomiques des 8 patients testés en aveugle ont été analysées par deux opérateurs indépendants. Chaque opérateur a créé sa propre arborescence pour filtrer les CNV, SNV et LOH.

Les deux opérateurs ont identifié correctement les variants recherchés chez 7 des 8 patients.

Pour le patient n°29 qui présente 2 CNV, les deux opérateurs ont détecté uniquement la délétion en 1q21.1q21.2. La duplication en 15q13.3 n'a pas été identifiée (Tableau 13).

Pour ces 8 patients, le chevauchement moyen des coordonnées génomiques entre le OneSeq et l'ACPA est de 91%. Le chevauchement médian est de 89,5% avec un minimum à 83% et un maximum à 99%.



Patient	Technique	Coordonnées génomiques	Log-ratio	Taille (kb)	Chevauchement coordonnées génomiques OneSeq/ACPA	Identification par l'opérateur 1	Identification par l'opérateur 2
<b>Cas n°23</b>	ACPA	1q43q44(243597509_243750450)x1	-0,92	152,9 kb	<b>87%</b>		
	ONESEQ	1q43q44(243570365_243768994)x1	-0,73	198,6 kb		<b>Oui</b>	<b>Oui</b>
<b>Cas n°24</b>	ACPA	17p11.2(17485651_17878387)x1	-0,88	392,7 kb	<b>92%</b>		
	ONESEQ	17p11.2(17493939_17932866)x1	-0,86	438,9 kb		<b>Oui</b>	<b>Oui</b>
<b>Cas n°25</b>	ACPA	7q11.23(72726578_74139390)x1	-0,65	1 412,8 kb	<b>98%</b>		
	ONESEQ	7q11.23(72744583_74172680)x1	-0,99	1 428,1 kb		<b>Oui</b>	<b>Oui</b>
<b>Cas n°26</b>	ACPA	Xp22.33p22.31(61091_6755201)x3 Xq26.2q28(132842738_155190083)x1	0,43 -0,71	6 694,1 kb 22 347,3 kb	<b>99%</b>		
	ONESEQ	Xp22.33p22.31(75466_6825671)x3 Xq26.2q28(132845291_155257985)x1	0,68 -0,99	6 750,2 kb 22 412,7 kb		<b>Oui</b>	<b>Oui</b>

<b>Cas n°27</b>	ACPA	1q21.1(145388355_145833054)x3 1q21.1q21.2(146506310_147824207)x1	0,52 -0,83	444,7 kb 1 317,9 kb	<b>87%</b>		
	ONESEQ	1q21.1(145378288_145894757)x3 1q21.1q21.2(146399192_147506558)x1	0,57 -0,74	516,5 kb 1 107,4 kb			
<b>Cas n°28</b>	ACPA	8p23.3p22(176814_15235814)x3 18q21.32q23(57529103_78010032)x1	0,52 -0,88	15 059 kb 20 480,9 kb	<b>99%</b>		
	ONESEQ	8p23.1p22(281161_15147668)x3 18q21.32q23(57581487_77770406)x1	0,54 -1,01	14 866,5 kb 20 188,9 kb			
<b>Cas n°29</b>	ACPA	1q21.1q21.2(146506310_147745610)x1 15q13.3(32021733_32510863)x3	-0,78 0,55	1 239,3 kb 489,1 kb	<b>85%</b>		
	ONESEQ	1q21.1q21.2(146246695_147506558)x1 <b>15q13.3(31992834_32452767)x3</b>	-0,88 <b>0,54</b>	1 259,9 kb <b>459,9 kb</b>			
<b>Cas n°30</b>	ACPA	16p11.2(28824794_29042118)x3	0,48	217,3 kb	<b>83%</b>		
	ONESEQ	16p11.2(28782401_29085946)x3	0,39	303,5 kb			

Tableau 13: patients 23 à 30 analysés en aveugle par deux opérateurs indépendants

# DISCUSSION

## I. Analyse des résultats

### 1) Analyse ouverte

Sur les patients testés en analyse ouverte, le OneSeq montre une détection des anomalies très proche de l'ACPA. En effet, 13 anomalies sur 15 ont été correctement identifiées avec un chevauchement médian des coordonnées génomiques avec l'ACPA de 94%.

Deux anomalies non détectées par le filtre optimisé du OneSeq auraient modifié le diagnostic puisqu'un résultat faussement négatif aurait été rendu :

- **Cas n°3 : duplication intra-génique de SOX5 en 12p12.1 d'environ 162 kb :**

5 sondes sont dupliquées avec un log-ratio moyen de 0,54.

Toutes les sondes dupliquées ont un log-ratio supérieur à 0,25 mais cette duplication est indétectable avec le filtre de routine malgré des changements de référence.

Cette duplication est identifiée (chevauchement de 98% avec l'ACPA) en modifiant l'aberration score treshold de ce filtre à 5. L'absence de détection de cette duplication est donc liée à ce paramètre car l'algorithme ne considère pas cette duplication significative.

Ce CNV est donc non détectable en raison des paramètres bio-informatiques utilisés.

- **Cas n°6 : délétion terminale de SMARCA4 en 19p13.2 d'environ 40 kb :**

2 sondes sont délétées avec un log-ratio moyen de -0,92.

Il est possible de détecter cette délétion en modifiant le paramétrage bio-informatique pour qu'un intervalle génomique soit considéré significatif dès lors que deux sondes successives sont déviantes. Cependant, nous avons choisi de ne pas identifier les CNV pour lesquels moins de 3 sondes sont déviées. En effet, cela n'est pas envisageable car il en résulterait un trop grand nombre de CNV (artéfacts et faux positifs).

A l'heure actuelle, les CNV sont détectés en utilisant uniquement les sondes « backbone » intergénomiques pangénomiques. Lors d'une future mise à jour d'Alissa Align & Call, les sondes exoniques seront également exploitées pour identifier les CNV *via* l'étude de la profondeur de



- **Cas n°11 : chevauchement de 93%**

Pour le cas n°11, les chevauchements des régions tripliquées et dupliquées sont très proches. Cependant, la région tripliquée est de taille plus importante en ACPA et cette dissimilitude ne peut pas être expliquée par une différence liée aux sondes (Figure 14). En observant le log-ratio des sondes du OneSeq dans cette région, la duplication identifiée en ACPA est pourtant visible (Figure 20, encadré rouge).

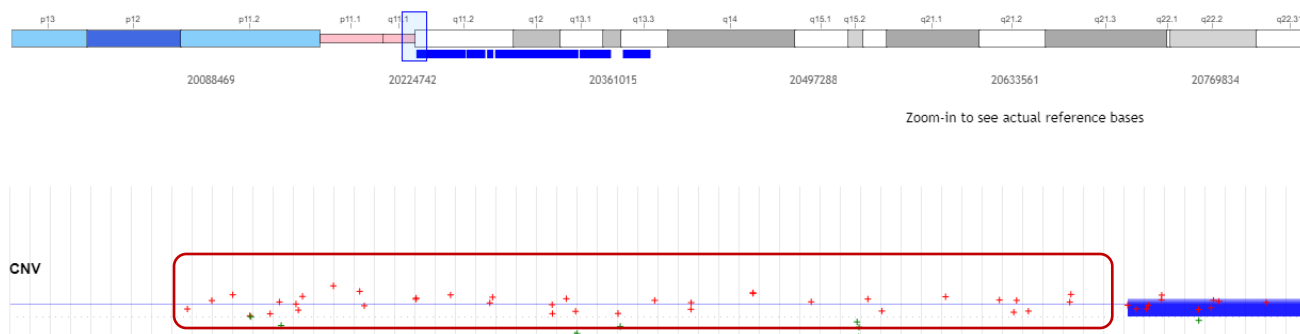


Figure 20: isodicyentrique 15q surnuméraire : sondes dupliquées non détectées (encadré rouge)

Même après modification des paramètres bio-informatiques cités précédemment et en changeant la référence, cette duplication n'est pas identifiée par l'algorithme de détection des CNV. Il est possible que les paramètres permettant son identification ne soient pas accessibles *via* l'optimisation de l'appel des variants et que la proximité du centromère en soit responsable. Cependant, cette différence n'aurait pas modifié le diagnostic puisque que le OneSeq détecte une duplication de plus de 12 000 kb. De plus, en raison de la présence d'une duplication très proche du centromère et de la fréquence des isodicyentriques du chromosome 15, cette anomalie aurait entraîné la réalisation d'un caryotype. Celui-ci aurait alors identifié cet isochromosome.

- **Cas n°4 : chevauchement de 88%**

Pour le cas n°4, la délétion détectée en OneSeq est plus grande que celle identifiée en ACPA. Ce résultat s'explique par le design du panel : les sondes intergéniques du OneSeq ont une répartition différente de celles de l'ACPA.

Il s'agit dans ce cas d'une délétion récurrente 15q11.2 médiée par des duplications segmentaires. Contrairement aux sondes d'ACPA, des sondes du OneSeq sont présentes à l'intérieur de la duplication segmentaire (Figure 21). Ainsi, malgré le taux apparemment faible de chevauchement, le résultat peut être considéré comme identique.

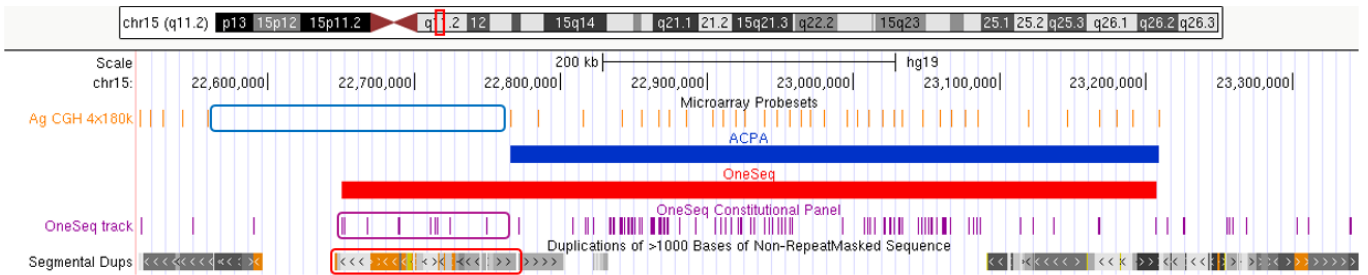


Figure 21: présence de sondes OneSeq (encadré violet) à l'intérieur de la duplication segmentaire (encadré rouge) et absence de sondes d'ACPA (encadré bleu)

- **Cas n°5 et n°13 : chevauchement de 84 et 96%**

Le cas n°5 présente une délétion homozygote en 8p22 et le cas n°13 est un patient avec une délétion hémizygote en Xp22.31.

Bien que l'absence totale de sonde soit visible sur la plateforme Alissa Align & Call (Figure 22), ces deux délétions n'étaient pas détectées par le OneSeq. Pour le patient n°5, le OneSeq met en évidence une région d'homozygotie à la place d'une délétion.

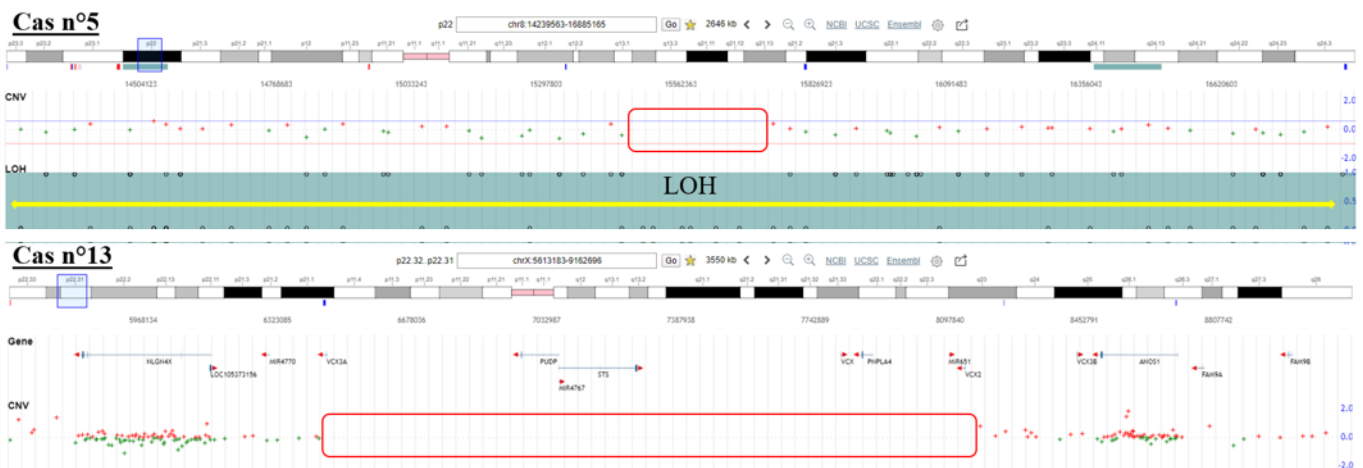


Figure 22: cas n°5 et 13, délétions non détectées par le OneSeq mais visibles sur la plateforme Alissa Align & Call. Les régions délétées sont encadrées en rouge, la région de LOH du cas n°5 est marquée par la flèche jaune.

Dans la seule étude publiée sur le OneSeq, Vetro *et al.* rapportent que l'unique cas étudié de délétion homozygote en ACPA n'a pas été détecté en OneSeq. Ils montrent également que cette délétion est située au sein d'une région d'homozygotie probablement en lien avec la consanguinité parentale [109]. Ce résultat est très similaire à celui que nous avons obtenu pour le patient n°5 et il existait donc un risque non négligeable de non détection d'une délétion homozygote potentiellement pathogène.

Dans ces régions, l'absence de reads liée à la délétion homozygote semblait être considérée par l'algorithme bio-informatique comme une région non couverte et n'était donc pas signalée comme un CNV potentiel. Cependant, la détection des CNV étant réalisée par comparaison

avec une référence, l'algorithme devrait détecter l'absence de reads de cette région chez le patient puisqu'ils sont présents sur la référence et donc détecter la présence d'une délétion homozygote.

Ainsi, nous nous sommes posé la question d'un problème lié au traitement bio-informatique des données et non accessible à notre niveau. Un travail sur cette question avec les ingénieurs d'Agilent Technologies a permis de confirmer qu'il s'agissait bien d'un problème lié à leur algorithme bio-informatique. Celui-ci a été modifié et le problème a été corrigé au cours du dernier changement de version de Alissa Align & Call. Les délétions homozygotes sont maintenant correctement détectées (Figure 23).

Les résultats des cas n°5 et n°13 sont présentés en Annexe I.

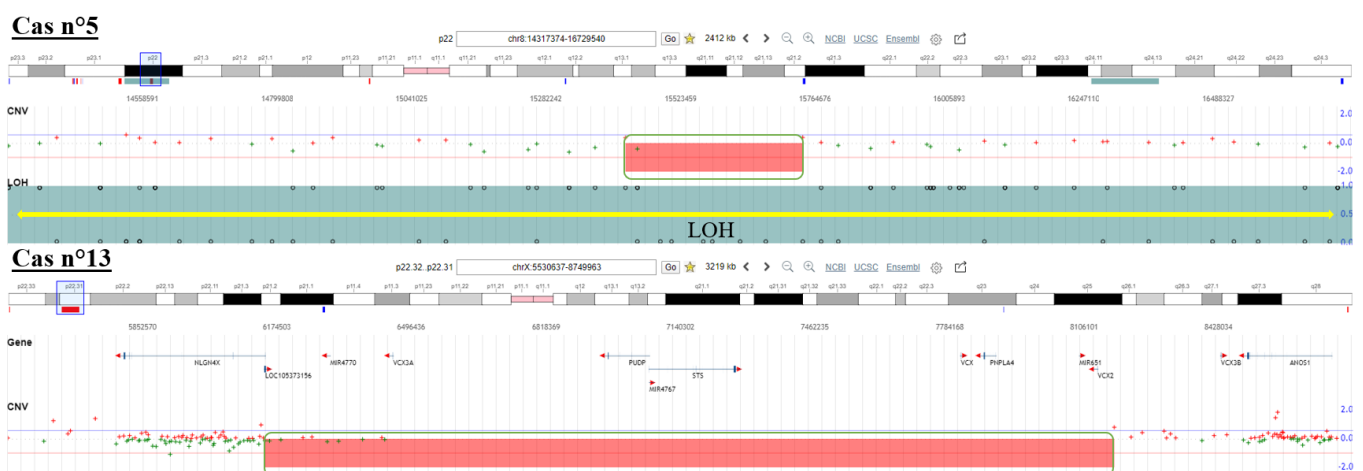


Figure 23: cas n°5 et 13, délétions détectées par le OneSeq après modification de l'algorithme sur Alissa Align & Call. Les régions délétées sont encadrées en vert, la région de LOH du cas n°5 est marquée par la flèche jaune.

- **Cas n°22**: région d'homozygotie en 15q11.2q21.2 et variation faux-sens c.858C>G dans le gène *CHST14*

Ce cas permet de mettre en évidence l'intérêt diagnostique de la détection simultanée des régions d'homozygotie et d'une mutation ponctuelle.

Un panel ciblant des gènes responsables d'arthrogrypose avait été demandé pour le patient n°22. Cette analyse a permis de mettre en évidence une variation faux-sens homozygote entraînant le remplacement d'une Cystéine par un Tryptophane dans le gène *CHST14* et prédite pathogène.

L'analyse par OneSeq retrouve le variant c.858C>G dans le gène *CHST14* mais également une région d'homozygotie d'environ 29 mégabases en 15q11.2q21.2. De façon intéressante, cette région comprend le gène *CHST14* (Figures 24 et 25).

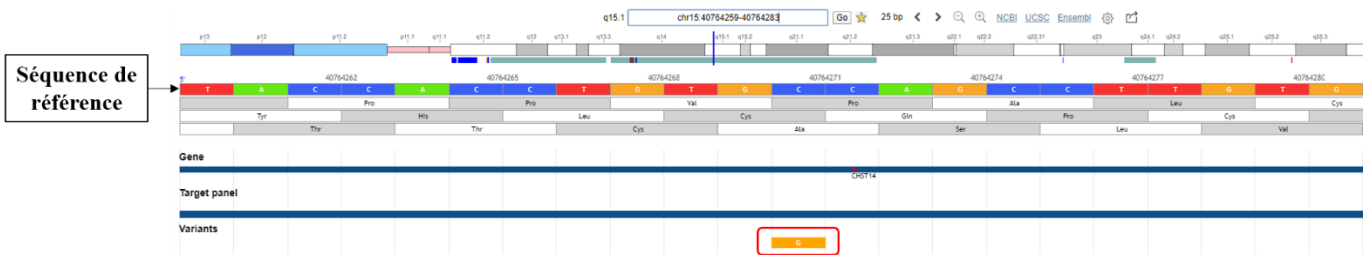


Figure 24: détection de la variation c.858C>G dans le gène CHST14. La Guanine remplaçant la Cytosine de référence est encadrée en rouge.



Figure 25: détection d'une région d'homozygotie en 15q11.2q21.2 comprenant le gène CHST14 (encadré rouge)

Une disomie uniparentale du chromosome 15 a été exclue par technique de MS-MLPA (Methylation-Specific Multiplex Ligation-Dependent Probe Amplification).

Cette grande région d'homozygotie nous a fait suspecter une consanguinité parentale ou une anomalie de structure parentale du chromosome 15 empêchant les mécanismes de recombinaison méiotique dans cette région. Dans ce contexte, un caryotype parental a été réalisé mais n'a pas retrouvé d'anomalie de structure pouvant expliquer cette région d'homozygotie.

L'hypothèse d'une consanguinité parentale à l'origine de cette région d'homozygotie était confortée par la détection de régions d'homozygotie disséminées sur d'autres chromosomes et a été confirmée secondairement par la famille.

Ainsi, la détection simultanée des variants nucléotidiques et des régions d'homozygotie donne des informations complémentaires qui n'auraient pas été détectées par les méthodes usuelles.



## 2) Analyse en aveugle

L'analyse en aveugle a permis de détecter les remaniements chromosomiques présents chez 7 des 8 patients analysés. Il est donc possible d'identifier les anomalies pathogènes sans orientation phénotypique de façon satisfaisante grâce à une arborescence filtrant les CNV, SNV et LOH.

Le patient n°29 présentait deux remaniements chromosomiques, une délétion en 1q21.1q21.2 et une duplication en 15q13.3. Ces deux CNV ont bien été détectés techniquement par le OneSeq mais lors de l'interprétation, la duplication 15q13.3 n'a pas été identifiée par les deux opérateurs.

L'opérateur 1 n'a pas identifié cette duplication en raison d'un biais lié à la construction de l'arborescence filtrant les variants. En effet, l'un des premiers filtres de l'arborescence correspond à la présence en population générale d'un CNV avec les mêmes coordonnées génomiques chez plus de 10 témoins. Les CNV pathogènes ne sont pas retenus et vont accéder aux filtres suivants tandis que les CNV a priori bénins sont arrêtés dès cette étape.

La duplication 15q13.3 est un remaniement récurrent de susceptibilité aux troubles neuro-développementaux à pénétrance incomplète et/ou expressivité variable (PIEV) et en conséquence, ce remaniement est également présent dans la population générale. Cette duplication a été retenue par le filtre de population car elle y est très représentée : plus de 100 patients dans la base de données Database of Genomic Variants (DGV) (Figure 26).

Le deuxième remaniement présent chez ce patient est également un PIEV, il s'agit d'une délétion 1q21.1q21.2. Ce remaniement ne correspondait pas aux critères du filtre de population générale et n'a donc pas été retenu. Il a ainsi été identifié par un filtre correspondant aux remaniements pathogènes et probablement pathogènes dans la base de données ClinGen. La présence de cette délétion a probablement contribué à l'absence d'identification de la duplication 15q13.3. En effet, il est rare d'identifier 2 CNV d'intérêt chez le même patient et l'analyse n'a donc pas été approfondie après détection de ce premier CNV.

L'opérateur 2 a choisi de filtrer les remaniements en fonction de leur présence dans la base de données interne et de leur classification dans la base de données ClinGen.

Ce remaniement n'étant pas considéré pathogène ou probablement pathogène dans cette base, la duplication a été orientée vers un filtre de taille où les duplications d'une taille importante sont systématiquement analysées. Cependant, la duplication 15q13.3 était d'une taille

inférieure à la taille minimale choisie pour ce filtre. Ainsi, elle a été orientée vers le filtre de présence en population générale et n'a donc pas été analysée en raison de sa présence chez des individus sains (Figure 27). La présence du deuxième remaniement a probablement contribué à l'absence d'identification de la duplication 15q13.3.

Dans ce contexte précis, l'absence d'identification de la duplication 15q13.3 n'aurait probablement pas eu d'impact particulier sur la prise en charge de ce patient.

En effet, l'interprétation de ce remaniement est controversée. La duplication récurrente 15q13.3 comprenant le gène *CHRNA7* est très présente en population générale (fréquence de 0,029%) et les bases de données françaises et internationales la classifient d'une manière variable en remaniement bénin, de signification incertaine ou pathogène [110]. Ces différences de classification s'expliquent notamment par un phénotype neuro-développemental peu spécifique et par sa fréquence en population générale.

De plus, les duplications 15q13.3 sont les CNV les plus fréquemment retrouvés par ACPA et sont le plus souvent héritées d'un parent sain. Il s'agit également de l'expérience de notre laboratoire où la fréquence de ces duplications est de 0,65% et où elles sont majoritairement héritées.

Enfin, les duplications 15q13.3 sont souvent accompagnées, comme chez notre patient, d'autres remaniements chromosomiques de signification incertaine.

Une hypothèse envisageable serait que le remaniement associé soit responsable du phénotype et que la duplication 15q13.3 soit bénigne et simplement le reflet de sa fréquence en population générale. Une autre hypothèse serait que l'association de ces deux remaniements soit responsable du phénotype dans un modèle de type « double hit » [111].

A l'heure actuelle, les recommandations du réseau français AChro-Puce sont de considérer cette duplication comme un remaniement récurrent de susceptibilité aux troubles neuro-développementaux à pénétrance incomplète et/ou expressivité variable. Son implication dans le phénotype du patient est donc incertaine.

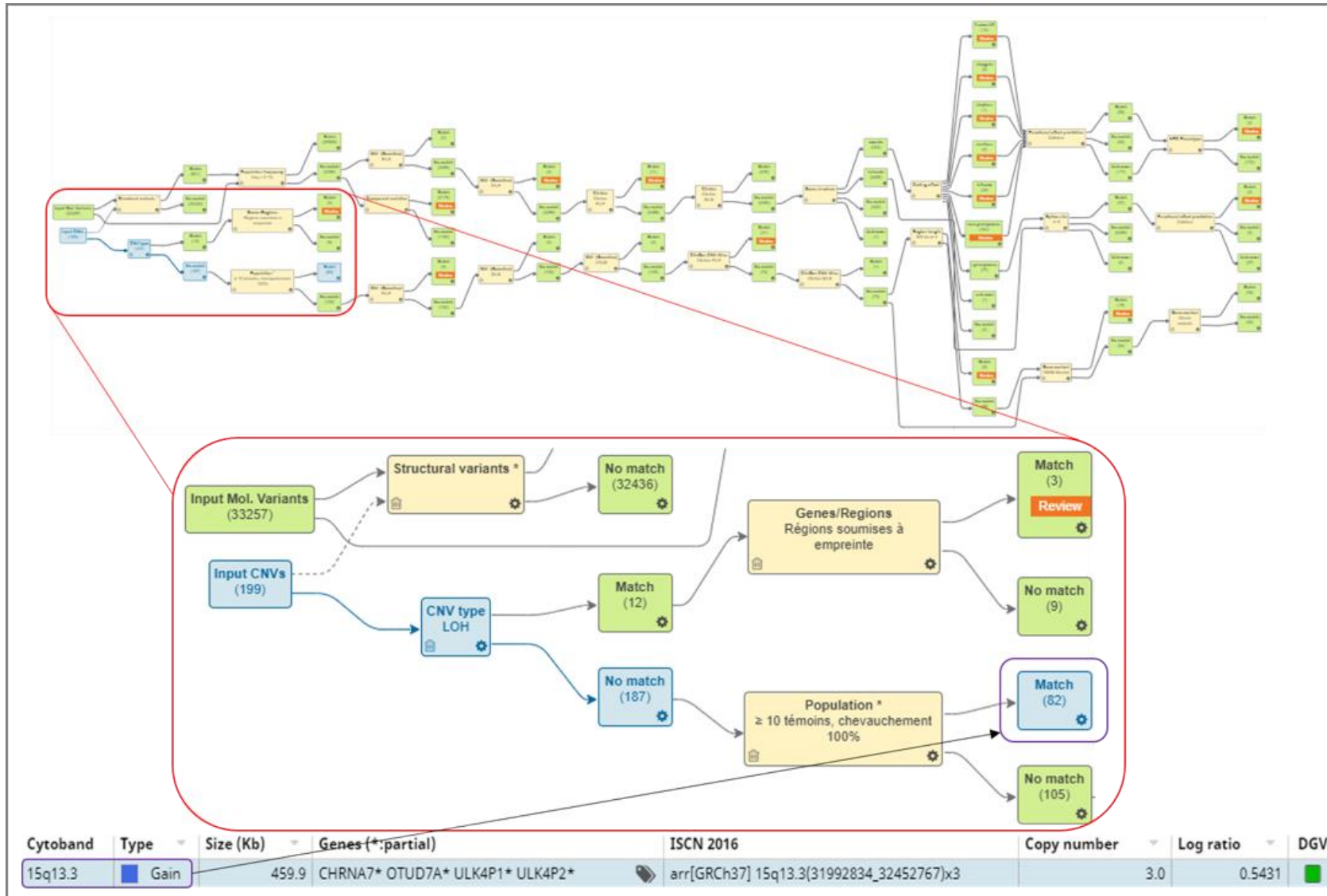


Figure 26: cheminement de la duplication 15q13.3 avec l'arborescence de l'opérateur 1, la duplication 15q13.3 est arrêtée par le filtre de population générale (encadrés violets liés par une flèche)

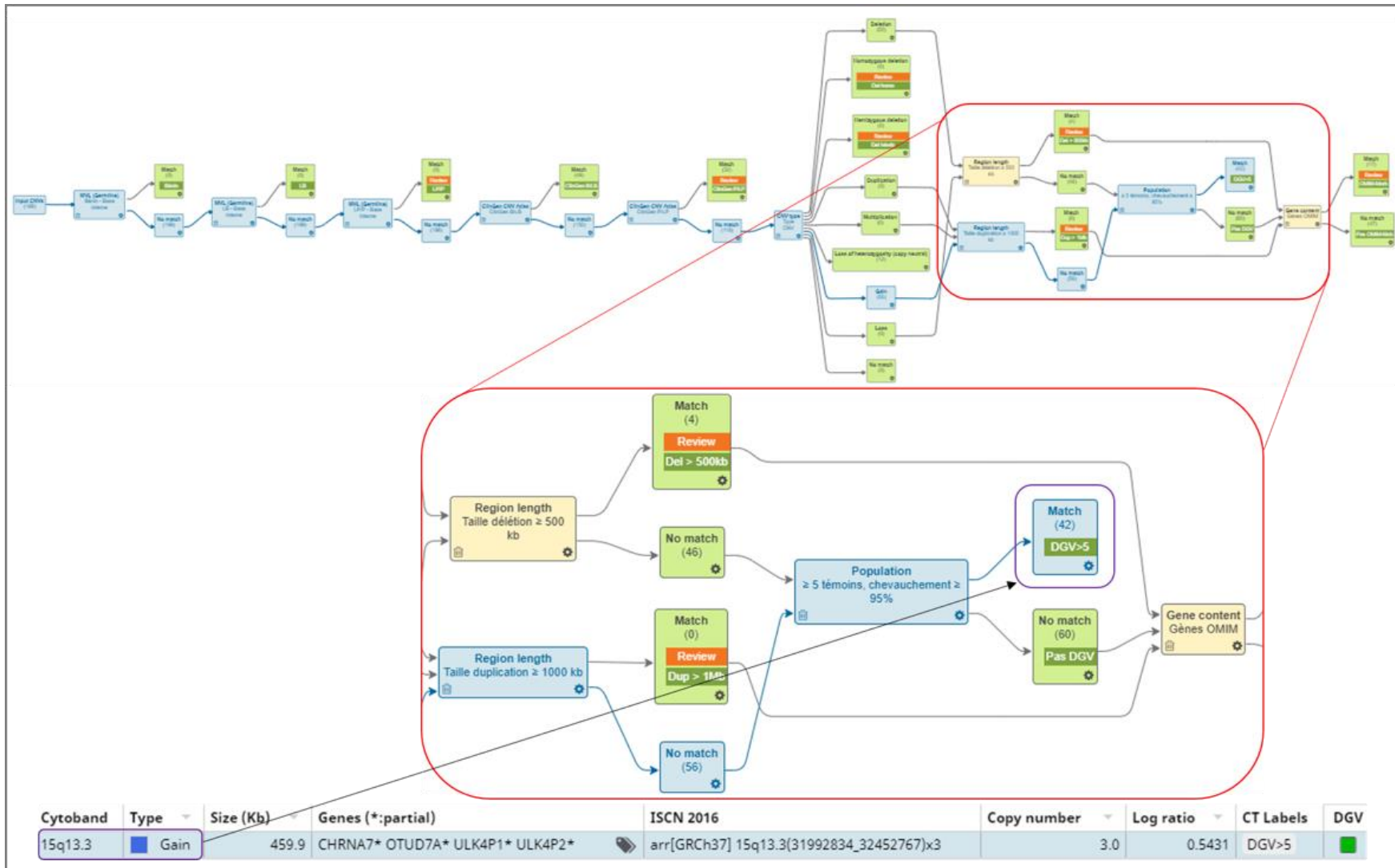


Figure 27: cheminement de la duplication 15q13.3 avec l'arborescence de l'opérateur 2, la duplication 15q13.3 est arrêtée par le filtre de population générale (encadrés violets liés par une flèche)

Cependant, l'absence de détection de ce remaniement lors de l'analyse en aveugle montre que les arborescences utilisées par les deux opérateurs présentent des défauts.

Ainsi, il était important d'améliorer l'arborescence permettant de filtrer les variants afin de détecter avec certitude les remaniements récurrents de susceptibilité aux troubles neuro-développementaux à pénétrance incomplète et/ou expressivité variable et les remaniements de signification incertaine.

Pour cela, nous avons mis en place une nouvelle arborescence pour les CNV qui comprend des filtres supplémentaires tels que les principaux gènes sensibles au dosage génique ou soumis à empreinte. Si le remaniement comprend l'un de ces gènes, il sera alors analysé de façon systématique (Figures 28 et 29).

Cette nouvelle arborescence a été testée sur le patient n°29 et a permis d'identifier la duplication 15q13.3 grâce à un filtre ciblant les remaniements récurrents et placé parallèlement au filtre de population générale.

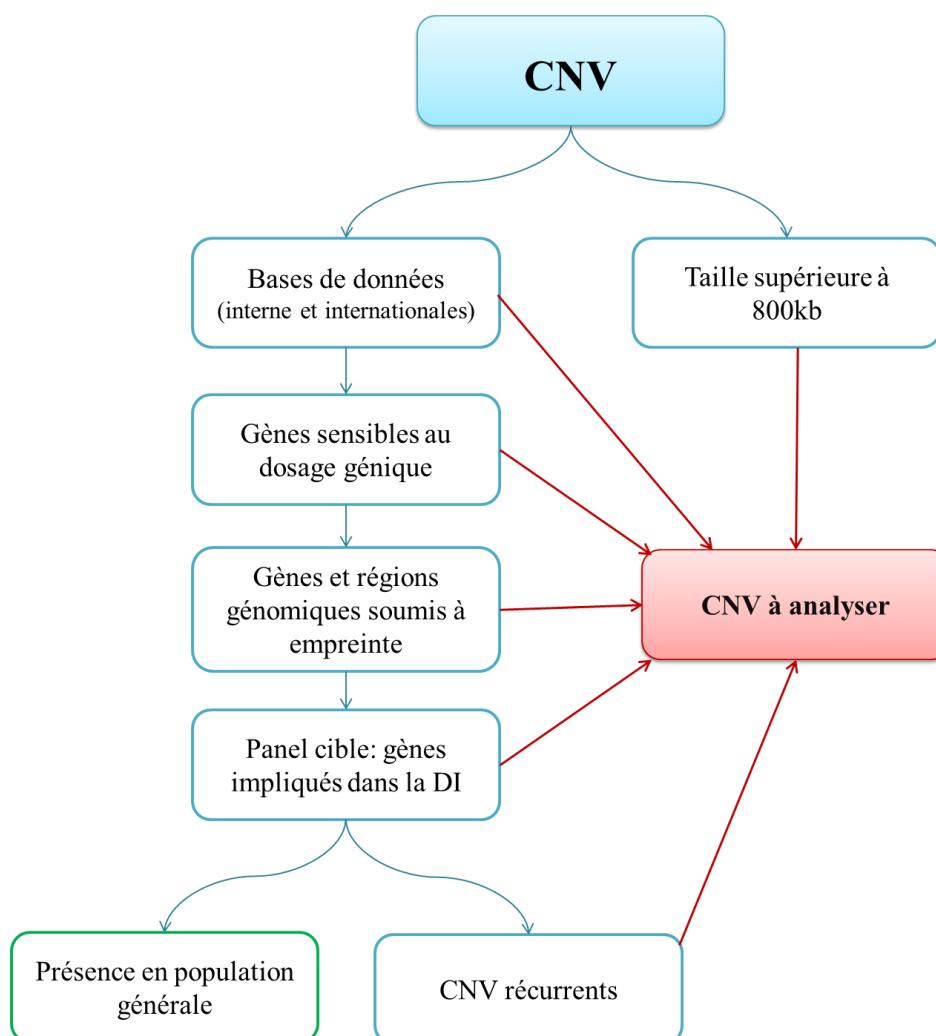


Figure 28: arborescence imaginée pour les CNV après l'analyse en aveugle

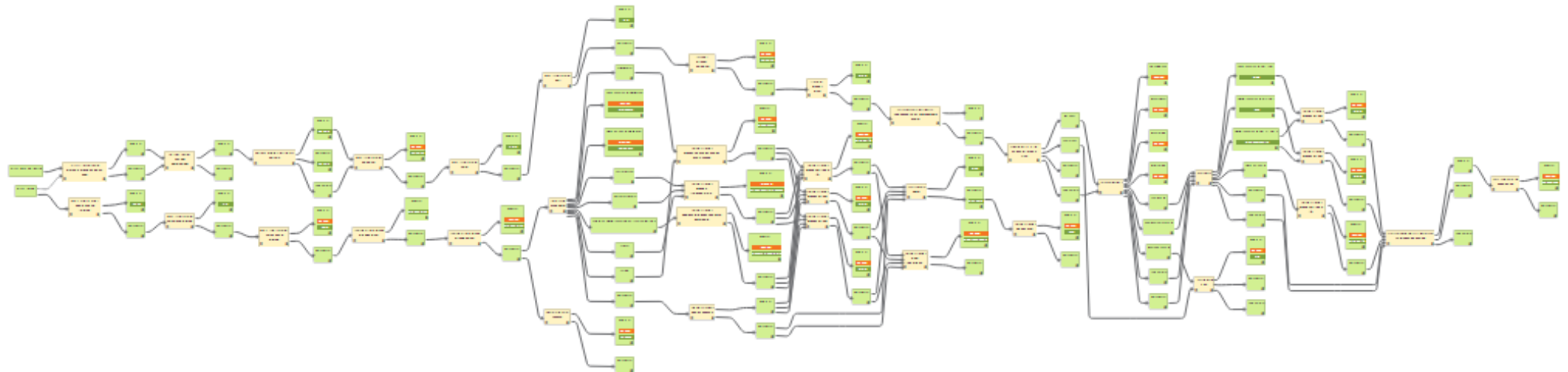


Figure 29: arborescence permettant de filtrer les CNV, SNV et LOH optimisée après l'analyse en aveugle (plateforme Alissa Interpret – Agilent Technologie)

À la suite de cette analyse en aveugle, il paraît envisageable d'utiliser cette technologie en routine diagnostique lorsque les optimisations de l'arborescence filtrant les variants auront été effectuées et testées avec plusieurs patients.

L'objectif sera de détecter de façon optimale et exhaustive les variants pathogènes, probablement pathogènes, les PIEV et les variants de signification inconnue pouvant être impliqués dans le phénotype.

### **3) Limite liée au nombre de CNV détectés**

Une des limites principales de cette technologie est le nombre de CNV détectés par patient.

En moyenne, 201 CNV par patient ont été identifiés en OneSeq contre seulement 10 à 15 en ACPA.

En OneSeq, le nombre de CNV identifiés dépend notamment du paramétrage bio-informatique et de la référence utilisée. Les CNV sont identifiés non pas par hybridation comparative comme l'ACPA mais par comparaison bio-informatique des données de profondeur de séquençage du patient par rapport à une référence.

Les paramètres bio-informatiques doivent permettre de détecter les anomalies d'intérêt mais sans être trop permissifs de façon à ne pas identifier un nombre trop important de CNV artéfactuels.

Dans cette optique, nous avons paramétré un filtre bio-informatique spécifique qui donne des résultats satisfaisants puisque les anomalies chromosomiques de 27 patients sur 30 ont été correctement identifiées.

Le nombre important de CNV est également lié à la référence utilisée. En effet, dans notre étude, les patients ont été analysés par rapport à une référence unique. L'utilisation d'un pool de références, en lissant les différences liées au séquençage et à la variabilité inter-individuelle, permettrait de diminuer le nombre de CNV identifiés et sera prochainement testé dans notre laboratoire. L'utilisation prochaine des données générées par les sondes exoniques pour identifier les CNV devrait également contribuer à la diminution du nombre de CNV artéfactuels.

Lors de l'interprétation d'une ACPA, les CNV sont analysés un par un en fonction de leur fréquence en population générale, de leur présence dans les bases de données et des gènes inclus. Une ACPA sans anomalie ou avec un CNV pathogène bien documenté nécessite environ 10 à 15 minutes d'interprétation biologique. Cependant, dans presque la moitié des cas, un CNV de signification inconnue sera identifié et nécessitera des recherches

bibliographiques complémentaires. Le temps d'interprétation biologique est alors beaucoup plus long.

Compte tenu du nombre de CNV détectés avec le OneSeq, l'analyse individuelle de chaque CNV n'est plus réalisable en raison du temps d'interprétation nécessaire.

L'analyse en aveugle de 8 patients par deux opérateurs indépendants montre que la création d'arborescences permettant de classer les CNV selon leur pathogénicité présumée permet de détecter les remaniements pathogènes de façon satisfaisante.

La mise en évidence fiable des variants de signification inconnue, absents en population générale et dans les bases de données, nécessite de filtrer les CNV en fonction des caractéristiques des gènes qu'ils contiennent. Il faut également tenir compte des PIEV pour optimiser l'arborescence de tri des variants puisqu'ils peuvent être présents en population générale mais participer au phénotype.

Le paramétrage des arborescences utilisées pour filtrer les variants est donc critique et doit encore être validé pour identifier avec certitudes tous les CNV d'intérêt lors de la réalisation de cette analyse en routine diagnostique.

## **II. Intérêt de la détection simultanée des CNV, SNV et LOH**

### **1) Rendement diagnostique**

A l'heure actuelle, devant une DI sans orientation clinique évidente, une ACPA et une recherche du syndrome de l'X fragile sont recommandés en première intention. Dans un second temps, en cas de négativité de ces deux examens, il est conseillé de s'orienter vers un panel de gènes impliqués dans la DI.

En raison de l'extrême hétérogénéité génétique de la DI, la réalisation séquentielle de plusieurs analyses permettant de rechercher différents types d'anomalies génétiques est nécessaire. En conséquence, le rendement diagnostique de l'ACPA et du NGS est variable dans la DI.

Plusieurs études ont montré que l'ACPA permettait l'identification d'une cause moléculaire dans 10 à 15% des déficiences intellectuelles [96,97].

Pour le NGS, le rendement diagnostique augmente de façon importante avec le nombre de gènes étudiés. L'analyse d'un panel comprenant 44 gènes fréquemment impliqués dans la DI permet un rendement diagnostique de 10 à 12% tandis que le séquençage de l'exome conduit



à un diagnostic dans 30 à 40% des cas [93,112–114]. Concernant l'exome clinique, l'étude de Chérot *et al.* montre un rendement diagnostique d'environ 32% dans la DI [115].

Selon une étude récente de C. Gao *et al.* réalisée dans une cohorte de 54 patients présentant des troubles neuro-développementaux, le rendement diagnostique de l'association de la recherche des CNV et du séquençage de l'exome par NGS serait de 57% [94].

Un intérêt supplémentaire de l'analyse couplée des CNV et des SNV est la capacité de détecter des pathologies récessives liées à un mécanisme hétérozygote composite expliqué par l'association d'une délétion ou d'une duplication sur un allèle du gène et d'une mutation ponctuelle sur l'autre allèle.

De plus, la recherche concomitante des régions d'homozygotie a un intérêt diagnostique dans la DI notamment lorsqu'il s'agit d'un chromosome soumis à empreinte ou dans les cas de consanguinité. Cette recherche est également importante dans un contexte de pathologie autosomique récessive où la découverte d'une disomie uniparentale, le plus souvent de survenue accidentelle, modifiera le conseil génétique puisque le risque de récurrence sera alors moins élevé [116]. Cependant, une hétérodisomie complète d'un chromosome ne pourra pas être détectée et concernerait environ 1/3 des cas de disomie uniparentale [117].

Dans notre étude, nous avons choisi une technologie permettant le séquençage de l'exome clinique, c'est-à-dire plus de 5000 gènes impliqués en pathologie, associé à la recherche des CNV et des LOH. Le rendement diagnostique attendu est compris entre 30 et 40% et devrait être équivalent à l'association d'une ACPA et d'un grand panel de gènes impliqués dans la DI. Il doit encore être évalué dans les conditions de notre laboratoire.

## **2) Intérêt économique et diminution du délai diagnostique**

De rares études ont étudié le coût des bilans étiologiques négatifs antérieurement réalisés et le délai pour parvenir au diagnostic étiologique de la DI.

En 2016, l'étude réalisée par Monroe *et al.* met en évidence une moyenne de 7 analyses génétiques réalisées pour chaque patient avant l'exome. Le temps écoulé pour parvenir au diagnostic était en moyenne de 6,6 ans (3,4 à 16,2 ans). Le coût des analyses génétiques séquentielles a été estimé à 5 972 euros contre 3 600 euros pour le séquençage de l'exome en trio. Le coût total des analyses et des consultations réalisées avant le diagnostic a été estimé à près de 15 000 euros par patient [118].

Selon une étude de Soden *et al.* portant sur 119 enfants ayant des troubles neuro-développementaux, le coût total des analyses réalisées est estimé à 17 000 euros par patient. Le temps écoulé pour obtenir le diagnostic était en moyenne de 7,9 ans (1,33 à 21,8 ans) [119].

Ces études mettent en évidence que la majorité des patients ayant une DI bénéficient de plusieurs explorations étiologiques séquentielles coûteuses et subissent une errance diagnostique supérieure à 5 ans en moyenne. Compte tenu de ces données, le séquençage de l'exome en première intention a déjà été proposé par Thevenon *et al.* et Srivastava *et al.* [114,120]. Ces données confortent l'intérêt du NGS dans les pathologies génétiquement hétérogènes comme la DI en permettant de limiter l'errance diagnostique et la multiplication d'examens étiologiques coûteux.

La mise en place du OneSeq en routine diagnostique va dans ce sens puisqu'un exome clinique et une ACPA seront réalisés puis interprétés dans un même temps analytique.

Nous avons réalisé une étude médico-économique comparant les coûts d'une analyse de OneSeq par rapport à une analyse séquentielle comprenant une ACPA puis un exome clinique. Cette étude montre que le coût d'une analyse de OneSeq est presque identique à celui de l'association ACPA-exome clinique tout en permettant également la détection des régions d'homozygotie (Tableau 14).

<b>Technique</b>	<b>Estimation du coût d'une analyse</b> (réactifs et temps technique)	<b>Anomalies génétiques détectées</b>	<b>Total</b>
<b>ACPA</b>	540 €	CNV	1210 €
<b>Exome clinique DI</b>	670 €	SNV	
<b>OneSeq</b>	1250 €	CNV, SNV, LOH	1250 €

Tableau 14: comparaison du coût estimé d'une analyse de OneSeq par rapport à l'association ACPA-exome clinique

En raison des recommandations nationales, l'ACPA est réalisée en première intention pour le diagnostic étiologique de la DI. Cependant dans ce contexte, le rendement diagnostique du séquençage d'un grand panel de gènes ou d'un exome est bien supérieur à celui de l'ACPA. Ainsi, chez une majorité de patients, un exome clinique sera réalisé consécutivement à une ACPA. Cette étude médico-économique permet de montrer que chez ces patients, le OneSeq contribuerait à limiter l'errance diagnostique en évitant la réalisation séquentielle de ces analyses et serait une alternative économiquement envisageable.

### **III. Avantages de la détection des CNV par utilisation de sondes intergéniques**

Bien qu'il soit possible de détecter les CNV par séquençage de l'exome en analysant la profondeur des reads, il ne s'agit pas de la méthode que nous avons choisie.

En effet, lors du séquençage de l'exome, seuls les CNV présents dans les régions codantes pourront être détectés, les régions non codantes telles que les régions régulatrices, promotrices ou introniques n'étant pas séquencées. Les avancées récentes réalisées en génétique ont montré qu'une altération de la régulation d'un gène peut être pathogène sans que le gène en lui-même ne soit altéré.

Le contrôle de la transcription peut être influencé par des modifications de la conformation chromatinienne ou par l'action des facteurs de transcription. Les éléments régulateurs peuvent être très éloignés du gène qu'ils régulent et sont parfois situés dans les introns des gènes avoisinants [121].

Récemment, il a été découvert que le génome était organisé en régions d'interactions chromatinienne préférentielles nommées Topologically Associating Domains (TAD). Les TAD sont hautement conservés chez les mammifères et représentent des unités régulatrices à l'intérieur desquelles les séquences amplificatrices (enhancer) et les promoteurs des gènes peuvent interagir. Les TAD sont délimités par des « frontières » composées notamment par des protéines isolantes comme CTCF. Ces frontières vont bloquer les interactions avec les TAD adjacents [122].

Lupiáñez *et al.* ont montré que l'interruption d'un TAD, en modifiant l'architecture du génome, peut altérer la régulation d'un gène éloigné et être responsable d'un phénotype anormal [123]. L'organisation chromatinienne peut être modifiée notamment par des CNV (délétions, duplications) mais également par des anomalies de structure équilibrées (inversions, translocations). Ces anomalies de structure, lorsqu'elles suppriment ou déplacent

les frontières entre les TAD, peuvent entraîner des interactions aberrantes entre deux TAD (Figure 30). Les unités régulatrices du premier TAD peuvent alors venir modifier l'expression des gènes cibles ectopiques appartenant au deuxième TAD [123]. Ces interactions peuvent entraîner l'expression d'un gène normalement non exprimé ou au contraire empêcher l'expression d'un gène qui devrait être exprimé. Jusqu'à 4,45% des délétions pourraient être pathogènes en raison d'une interaction ectopique entre un enhancer et un promoteur [124].

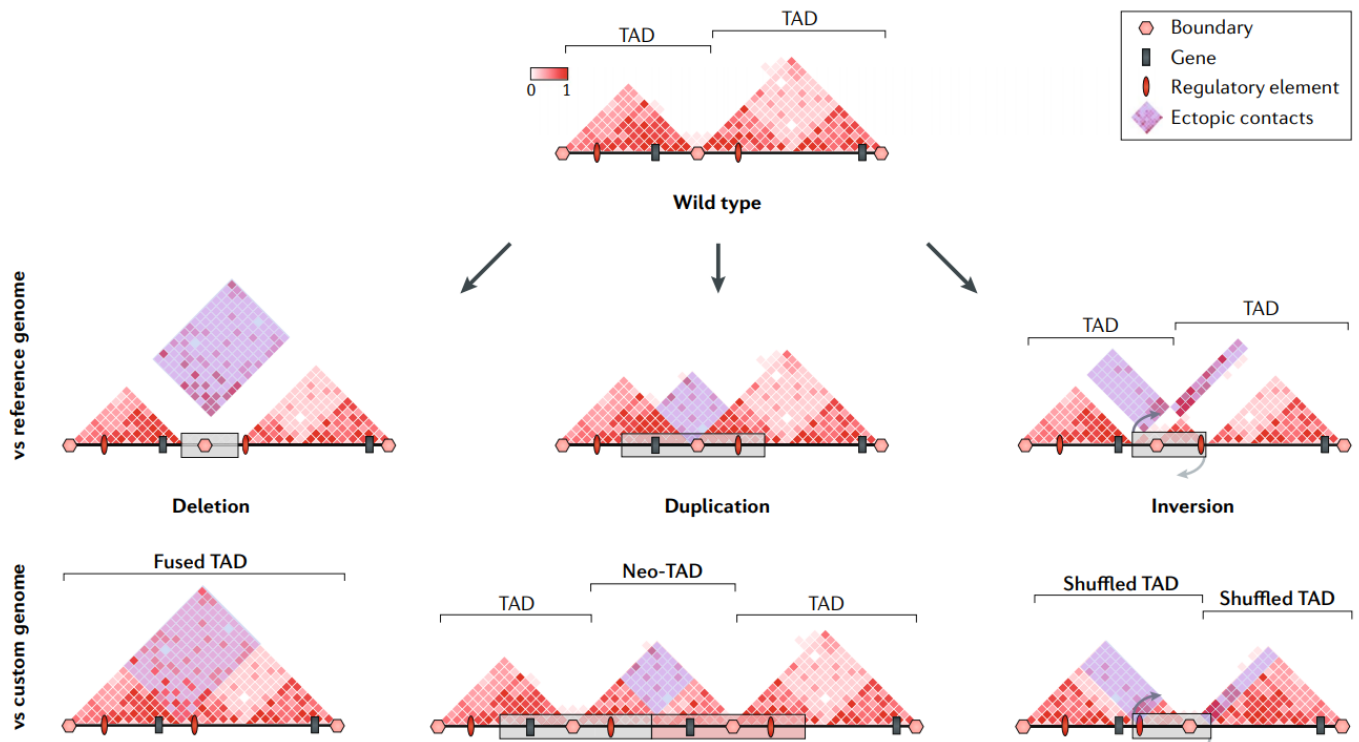


Figure 30: représentation de deux TAD et conséquences engendrées par une délétion, une duplication et une inversion sur l'architecture chromatinienne ; d'après Spielmann et al., 2018 [125]

Lors de l'interprétation des anomalies de structure, il est conseillé de tenir compte des TAD et en particulier de l'intégrité de leurs frontières pour prédire une interaction enhancer-promoteur ectopique et interpréter leur pathogénicité notamment dans les régions non codantes [123]. L'analyse de l'impact des anomalies de structure sur les TAD commence à être réalisée en routine diagnostique.

Ainsi, avec l'avancée des connaissances et le développement de bases de données référençant les TAD, l'analyse des régions intergéniques non codantes paraît aujourd'hui très importante. Il s'agit de l'un des avantages du OneSeq par rapport au séquençage de l'exome puisque la détection des CNV est faite *via* des sondes intergéniques pangénomiques qui permettront, à l'avenir, l'interprétation des TAD.

## CONCLUSION

Notre étude montre que la sensibilité du OneSeq pour la détection des CNV est similaire à celle de l'ACPA. 20 anomalies sur 23 préalablement identifiées en ACPA ont été correctement détectées avec un chevauchement médian de 94% (minimum 80%, maximum 99%). Toutes techniques diagnostiques confondues, les anomalies chromosomiques de 27 patients sur 30 ont été correctement identifiées. Ces résultats sont donc très satisfaisants.

L'utilisation prochaine des sondes de séquençage pour la détection des CNV et d'un pool de références diminuera le nombre de CNV détectés ce qui facilitera l'interprétation analytique.

La détection combinée des CNV, SNV et LOH par NGS devrait permettre d'augmenter le rendement diagnostique, de diminuer l'errance diagnostique et, à terme, de diminuer le coût analytique des examens génétiques nécessaires pour parvenir au diagnostic étiologique de la DI.

Cette technologie paraît être un bon compromis pour notre laboratoire en attendant la disponibilité du séquençage de génome entier en routine diagnostique.

Une étude médico-économique plus détaillée prenant en compte le temps d'interprétation biologique et le nombre de consultations pour les prescriptions multiples est nécessaire pour évaluer la possibilité de mettre en place cette technologie en routine notamment dans le cadre de la DI et des anomalies du développement.





## REFERENCES BIBLIOGRAPHIQUES

1. CIM-10. Classification internationale des maladies. World Health Organization. Dixième édition. 1993. 328p.
2. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5. Washington, DC: American Psychiatric Association. 2013. 1520 p.
3. Schalock RL, Borthwick-Duffy SA, Bradley VJ, Buntinx WHE, Coulter DL, Craig EM, et al. Intellectual Disability: Definition, Classification, and Systems of Supports. Eleventh Edition. American Association on Intellectual and Developmental Disabilities; 2010.
4. Inserm. Déficiences intellectuelles. Collection Expertise collective. Montrouge : EDP Sciences, 2016.
5. Leonard H, Wen X. The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev.* 2002;8(3):117–34.
6. Mercer JR. The myth of 3% prevalence. *Monogr Am Assoc Ment Defic.* 1973;(1):1–18.
7. Maulik PK, Mascarenhas MN, Mathers CD, Dua T, Saxena S. Prevalence of intellectual disability: A meta-analysis of population-based studies. *Research in Developmental Disabilities.* 2011 Mar 1;32(2):419–36.
8. David M, Dieterich K, Villemeur AB de, Jouk P-S, Counillon J, Larroque B, et al. Prevalence and characteristics of children with mild intellectual disability in a French county. *Journal of Intellectual Disability Research.* 2014;58(7):591–602.
9. Chiurazzi P, Schwartz CE, Gecz J, Neri G. XLMR genes: update 2007. *Eur J Hum Genet.* 2008 Apr;16(4):422–34.
10. Glasson EJ, Buckley N, Chen W, Leonard H, Epstein A, Skoss R, et al. Systematic Review and Meta-Analysis: Mental Health in Children With Neurogenetic Disorders Associated With Intellectual Disability. *Journal of the American Academy of Child & Adolescent Psychiatry.* 2020 Jan 13;0(0).



11. Petterson B, Bourke J, Leonard H, Jacoby P, Bower C. Co-occurrence of birth defects and intellectual disability. *Paediatr Perinat Epidemiol*. 2007 Jan;21(1):65–75.
12. Emerson E, Hatton C. Mental health of children and adolescents with intellectual disabilities in Britain. *Br J Psychiatry*. 2007 Dec;191(6):493–9.
13. Oeseburg B, Dijkstra GJ, Groothoff JW, Reijneveld SA, Jansen DEMC. Prevalence of Chronic Health Conditions in Children With Intellectual Disability: A Systematic Literature Review. *Intellectual and Developmental Disabilities*. 2011 Apr;49(2):59–85.
14. Matson JL, Shoemaker M. Intellectual disability and its relationship to autism spectrum disorders. *Research in Developmental Disabilities*. 2009 Nov;30(6):1107–14.
15. Korndewal MJ, Oudesluys-Murphy AM, Kroes ACM, van der Sande MAB, de Melker HE, Vossen ACTM. Long-term impairment attributable to congenital cytomegalovirus infection: a retrospective cohort study. *Dev Med Child Neurol*. 2017;59(12):1261–8.
16. Mwaniki MK, Atieno M, Lawn JE, Newton CR. Long-term neurodevelopmental outcomes after intrauterine and neonatal insults: a systematic review. *Lancet*. 2012 Feb 4;379(9814):445–52.
17. Inserm. Déficiences et handicaps d'origine périnatale. Expertise collective Inserm, Les Éditions Inserm, Paris 2004.
18. Haddow JE, Palomaki GE, Allan WC, Williams JR, Knight GJ, Gagnon J, et al. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N Engl J Med*. 1999 Aug 19;341(8):549–55.
19. Prick BW, Hop WCJ, Duvekot JJ. Maternal phenylketonuria and hyperphenylalaninemia in pregnancy: pregnancy complications and neonatal sequelae in untreated and treated pregnancies. *Am J Clin Nutr*. 2012 Feb;95(2):374–82.
20. Benaron HBW, Tucker BE, Andrews JP, Boshes B, Cohen J, Fromm E, et al. Effect of anoxia during labor and immediately after birth on the subsequent development of the child. *American Journal of Obstetrics & Gynecology*. 1960 Dec 1;80(6):1129–42.
21. Camp BW, Broman SH, Nichols PL, Leff M. Maternal and neonatal risk factors for mental retardation: defining the “at-risk” child. *Early Hum Dev*. 1998 Jan 9;50(2):159–73.

22. Hirvonen M, Ojala R, Korhonen P, Haataja P, Eriksson K, Rantanen K, et al. Intellectual disability in children aged less than seven years born moderately and late preterm compared with very preterm and term-born children – a nationwide birth cohort study. *Journal of Intellectual Disability Research*. 2017;61(11):1034–54.
23. Grimwood K, Anderson P, Anderson V, Tan L, Nolan T. Twelve year outcomes following bacterial meningitis: further evidence for persisting effects. *Arch Dis Child*. 2000 Aug;83(2):111–6.
24. Singhi P, Bansal A, Geeta P, Singhi S. Predictors of long term neurological outcome in bacterial meningitis. *Indian J Pediatr*. 2007 Apr;74(4):369–74.
25. Anderson V, Godfrey C, Rosenfeld JV, Catroppa C. Predictors of Cognitive Function and Recovery 10 Years After Traumatic Brain Injury in Young Children. *Pediatrics*. 2012 Feb 1;129(2):e254–61.
26. Anderson V, Catroppa C, Godfrey C, Rosenfeld JV. Intellectual Ability 10 Years after Traumatic Brain Injury in Infancy and Childhood: What Predicts Outcome? *Journal of Neurotrauma*. 2011 Oct 24;29(1):143–53.
27. Min MO, Singer LT, Kirchner HL, Minnes S, Short E, Hussain Z, et al. Cognitive Development and Low-Level Lead Exposure in Poly-Drug Exposed Children. *Neurotoxicol Teratol*. 2009;31(4):225–31.
28. Heikura U, Taanila A, Hartikainen A-L, Olsen P, Linna S-L, von Wendt L, et al. Variations in prenatal sociodemographic factors associated with intellectual disability: a study of the 20-year interval between two birth cohorts in northern Finland. *Am J Epidemiol*. 2008 Jan 15;167(2):169–77.
29. Bilder DA, Pinborough-Zimmerman J, Bakian AV, Miller JS, Dorius JT, Nangle B, et al. Prenatal and perinatal factors associated with intellectual disability. *Am J Intellect Dev Disabil*. 2013 Mar;118(2):156–76.
30. Chapman DA, Scott KG, Mason CA. Early risk factors for mental retardation: role of maternal age and maternal education. *Am J Ment Retard*. 2002 Jan;107(1):46–59.
31. Stevenson RE, Procopio-Allen AM, Schroer RJ, Collins JS. Genetic syndromes among individuals with mental retardation. *Am J Med Genet*. 2003 Nov 15;123A(1):29–32.

32. Hassold T, Abruazzo M, Adkins K, Griffin D, Merrill M, Millie E, et al. Human aneuploidy: Incidence, origin, and etiology. *Environmental and Molecular Mutagenesis*. 1996;28(3):167–75.
33. Cereda A, Carey JC. The trisomy 18 syndrome. *Orphanet Journal of Rare Diseases*. 2012 Oct 23;7(1):81.
34. Rasmussen SA, Wong L-YC, Yang Q, May KM, Friedman JM. Population-based analyses of mortality in trisomy 13 and trisomy 18. *Pediatrics*. 2003 Apr;111(4 Pt 1):777–84.
35. Webster A, Schuh M. Mechanisms of Aneuploidy in Human Eggs. *Trends in Cell Biology*. 2017 Jan;27(1):55–68.
36. Resta RG. Changing demographics of advanced maternal age (AMA) and the impact on the predicted incidence of Down syndrome in the United States: Implications for prenatal screening and genetic counseling. *American Journal of Medical Genetics Part A*. 2005;133A(1):31–6.
37. Loane M, Morris JK, Addor M-C, Arriola L, Budd J, Doray B, et al. Twenty-year trends in the prevalence of Down syndrome and other trisomies in Europe: impact of maternal age and prenatal screening. *Eur J Hum Genet*. 2013 Jan;21(1):27–33.
38. Linden MG, Bender BG, Robinson A. Sex chromosome tetrasomy and pentasomy. *Pediatrics*. 1995;96:672–682.
39. Deng HX, Abe K, Kondo I, Tsukahara M, Inagaki H, Hamada I, et al. Parental origin and mechanism of formation of polysomy X: an XXXXX case and four XXXXY cases determined with RFLPs. *Hum Genet*. 1991 Apr;86(6):541–4.
40. Karadima G, Bugge M, Nicolaidis P, Vassilopoulos D, Avramopoulos D, Grigoriadou M, et al. Origin of nondisjunction in trisomy 8 and trisomy 8 mosaicism. *Eur J Hum Genet*. 1998 Sep;6(5):432–8.
41. Bruns DA, Campbell E. Twenty-five additional cases of trisomy 9 mosaic: Birth information, medical conditions, and developmental status. *American Journal of Medical Genetics Part A*. 2015;167(5):997–1007.

42. Schluth-Bolard C, Labalme A, Cordier M-P, Till M, Nadeau G, Tevissen H, et al. Breakpoint mapping by next generation sequencing reveals causative gene disruption in patients carrying apparently balanced chromosome rearrangements with intellectual deficiency and/or congenital malformations. *J Med Genet.* 2013 Mar;50(3):144–50.
43. Kalscheuer VM, Tao J, Donnelly A, Hollway G, Schwinger E, Kübart S, et al. Disruption of the serine/threonine kinase 9 gene causes severe X-linked infantile spasms and mental retardation. *Am J Hum Genet.* 2003 Jun;72(6):1401–11.
44. Fantes J, Redeker B, Breen M, Boyle S, Brown J, Fletcher J, et al. Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. *Hum Mol Genet.* 1995 Mar;4(3):415–22.
45. Dupont J-M, Cuisset L, Cartigny M, Le Tessier D, Vasseur C, Rabineau D, et al. Familial reciprocal translocation t(7;16) associated with maternal uniparental disomy 7 in a Silver-Russell patient. *Am J Med Genet.* 2002 Sep 1;111(4):405–8.
46. Warburton D. De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. *Am J Hum Genet.* 1991 Nov;49(5):995–1013.
47. Madan K, Nieuwint AWM, van Bever Y. Recombination in a balanced complex translocation of a mother leading to a balanced reciprocal translocation in the child. Review of 60 cases of balanced complex translocations. *Human Genetics.* 1997 May 15;99(6):806–15.
48. Talkowski ME, Ernst C, Heilbut A, Chiang C, Hanscom C, Lindgren A, et al. Next-Generation Sequencing Strategies Enable Routine Detection of Balanced Chromosome Rearrangements for Clinical Diagnostics and Genetic Research. *Am J Hum Genet.* 2011 Apr 8;88(4):469–81.
49. Suzuki T, Tsurusaki Y, Nakashima M, Miyake N, Saitsu H, Takeda S, et al. Precise detection of chromosomal translocation or inversion breakpoints by whole-genome sequencing. *J Hum Genet.* 2014 Dec;59(12):649–54.
50. Liang D, Wang Y, Ji X, Hu H, Zhang J, Meng L, et al. Clinical application of whole-genome low-coverage next-generation sequencing to detect and characterize balanced chromosomal translocations. *Clin Genet.* 2017 Apr;91(4):605–10.

51. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet*. 2015 Mar;16(3):172–83.
52. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, et al. Population Analysis of Large Copy Number Variants and Hotspots of Human Genetic Disease. *Am J Hum Genet*. 2009 Feb 13;84(2):148–61.
53. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nature Reviews Genetics*. 2015 Mar;16(3):172–83.
54. Hastings P, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. *Nat Rev Genet*. 2009 Aug;10(8):551–64.
55. Henrichsen CN, Chaignat E, Reymond A. Copy number variants, diseases and gene expression. *Hum Mol Genet*. 2009 Apr 15;18(R1):R1–8.
56. Sanlaville D, Schluth-Bolard C, Turleau C. Distal Xq duplication and functional Xq disomy. *Orphanet J Rare Dis*. 2009 Feb 20;4:4.
57. Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH, et al. Relative Burden of Large CNVs on a Range of Neurodevelopmental Phenotypes. *PLoS Genet* [Internet]. 2011 Nov 10;7(11).
58. Pedoutour F, Turc-Carel C. Des chromosomes sans fin : les chromosomes en anneau. *Med Sci (Paris)*. 1997;13(11):1239.
59. Kubota T, Wakui K, Nakamura T, Ohashi H, Watanabe Y, Yoshino M, et al. The proportion of cells with functional X disomy is associated with the severity of mental retardation in mosaic ring X Turner syndrome females. *CGR*. 2002;99(1–4):276–84.
60. Migeon BR, Luo S, Stasiowski BA, Jani M, Axelman J, Van Dyke DL, et al. Deficient transcription of XIST from tiny ring X chromosomes in females with severe phenotypes. *Proc Natl Acad Sci U S A*. 1993 Dec 15;90(24):12025–9.
61. Daber RD, Conlin LK, Leonard LD, Canevini MP, Vignoli A, Hosain S, et al. Ring chromosome 20. *European Journal of Medical Genetics*. 2012 May;55(5):381–7.

62. Liehr T, Weise A. Frequency of small supernumerary marker chromosomes in prenatal, newborn, developmentally retarded and infertility diagnostics. *International Journal of Molecular Medicine*. 2007 May 1;19(5):719–31.
63. Liehr T, Claussen U, Starke H. Small supernumerary marker chromosomes (sSMC) in humans. *CGR*. 2004;107(1–2):55–67.
64. Finucane BM, Lusk L, Arkilo D, Chamberlain S, Devinsky O, Dindot S, et al. 15q Duplication Syndrome and Related Disorders. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJ, Stephens K, et al., editors. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 1993
65. Wiszniewska J, Bi W, Shaw C, Stankiewicz P, Kang S-HL, Pursley AN, et al. Combined array CGH plus SNP genome analyses in a single assay for optimized clinical testing. *Eur J Hum Genet*. 2014 Jan;22(1):79–87.
66. Wang J-C, Ross L, Mahon LW, Owen R, Hemmat M, Wang BT, et al. Regions of homozygosity identified by oligonucleotide SNP arrays: evaluating the incidence and clinical utility. *Eur J Hum Genet*. 2015 May;23(5):663–71.
67. Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet*. 2006 Mar 1;15(5):789–95.
68. Wang J-C, Ross L, Mahon LW, Owen R, Hemmat M, Wang BT, et al. Regions of homozygosity identified by oligonucleotide SNP arrays: evaluating the incidence and clinical utility. *Eur J Hum Genet*. 2015 May;23(5):663–71.
69. Buiting K. Prader–Willi syndrome and Angelman syndrome. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*. 2010;154C(3):365–76.
70. Gurrieri F, Accadia M. Genetic Imprinting: The Paradigm of Prader-Willi and Angelman Syndromes. *Endocrine Involvement in Developmental Syndromes*. 2009;14:20–8.
71. Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nature Reviews Genetics*. 2016 Jan;17(1):9–18.

72. Tejada MI, Ibarluzea N. Non-syndromic X linked intellectual disability: Current knowledge in light of the recent advances in molecular and functional studies. *Clin Genet*. 2020 Jan 9;cge.13698.
73. Kleefstra T, Schenck A, Kramer JM, van Bokhoven H. The genetics of cognitive epigenetics. *Neuropharmacology*. 2014 May 1;80:83–94.
74. Wieczorek D. Autosomal dominant intellectual disability. *Med Genet*. 2018;30(3):318–22.
75. Prevalence and architecture of de novo mutations in developmental disorders. *Nature*. 2017 Feb 23;542(7642):433–8.
76. McRae JF, Clayton S, Fitzgerald TW, Kaplanis J, Prigmore E, Rajan D, et al. Prevalence, phenotype and architecture of developmental disorders caused by de novo mutation: The Deciphering Developmental Disorders Study. *bioRxiv*. 2016 Jun 16;049056.
77. Fitzgerald T, Gerety S, Jones W, van Kogelenberg M, King D, McRae J, et al. Large-scale discovery of novel genetic causes of developmental disorders. *Nature*. 2015 Mar 12;519(7542):223–8.
78. Musante L, Ropers HH. Genetics of recessive cognitive disorders. *Trends in Genetics*. 2014 Jan 1;30(1):32–9.
79. Rauch A, Wieczorek D, Graf E, Wieland T, Ende S, Schwarzmayr T, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *The Lancet*. 2012 Nov 10;380(9854):1674–82.
80. Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, et al. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature*. 2011 Oct;478(7367):57–63.
81. Neri G, Schwartz CE, Lubs HA, Stevenson RE. XLID Update 2017. *Am J Med Genet A*. 2018 Jun;176(6):1375–88.
82. Hunter J, Rivero-Arias O, Angelov A, Kim E, Fotheringham I, Leal J. Epidemiology of fragile X syndrome: A systematic review and meta-analysis. *American Journal of Medical Genetics Part A*. 2014;164(7):1648–58.

83. Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*. 1991 May 31;65(5):905–14.
84. De Boulle K, Verkerk AJ, Reyniers E, Vits L, Hendrickx J, Van Roy B, et al. A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat Genet*. 1993 Jan;3(1):31–5.
85. Lugenbeel KA, Peier AM, Carson NL, Chudley AE, Nelson DL. Intragenic loss of function mutations demonstrate the primary role of FMR1 in fragile X syndrome. *Nat Genet*. 1995 Aug;10(4):483–5.
86. Hunter JE, Berry-Kravis E, Hipp H, Todd PK. FMR1 Disorders. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJ, Stephens K, et al., editors. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 1993
87. Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, et al. Phenotypic Heterogeneity of Genomic Disorders and Rare Copy-Number Variants. *N Engl J Med*. 2012 Oct 4;367(14):1321–31.
88. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A, et al. A recurrent 16p12.1 microdeletion suggests a two-hit model for severe developmental delay. *Nat Genet*. 2010 Mar;42(3):203–9.
89. Girirajan S, Eichler EE. Phenotypic variability and genetic susceptibility to genomic disorders. *Hum Mol Genet*. 2010 Oct 15;19(R2):R176–87.
90. Paciorkowski AR, Keppler-Noreuil K, Robinson L, Sullivan C, Sajan S, Christian SL, et al. Deletion 16p13.11 uncovers NDE1 mutations on the non-deleted homolog and extends the spectrum of severe microcephaly to include fetal brain disruption. *Am J Med Genet A*. 2013 Jul;161(7):1523–30.
91. McDonald-McGinn DM, Fahiminiya S, Revil T, Nowakowska BA, Suhl J, Bailey A, et al. Hemizygous mutations in SNAP29 unmask autosomal recessive conditions and contribute to atypical findings in patients with 22q11.2DS. *J Med Genet*. 2013 Feb;50(2):80–90.
92. ANPGM\_122-Deficiencie\_intellectuelle.pdf [Internet]. [Available from: [https://anpgm.fr/media/documents/ANPGM\\_122-Deficiencie\\_intellectuelle.pdf](https://anpgm.fr/media/documents/ANPGM_122-Deficiencie_intellectuelle.pdf)]



93. Dillon OJ, Lunke S, Stark Z, Yeung A, Thorne N, Gaff C, et al. Exome sequencing has higher diagnostic yield compared to simulated disease-specific panels in children with suspected monogenic disorders. *Eur J Hum Genet*. 2018 May;26(5):644–51.
94. Gao C, Wang X, Mei S, Li D, Duan J, Zhang P, et al. Diagnostic Yields of Trio-WES Accompanied by CNVseq for Rare Neurodevelopmental Disorders. *Front Genet* [Internet]. 2019 May 24;10.
95. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*. 2010 Jan;42(1):30–5.
96. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. *Am J Hum Genet*. 2010 May 14;86(5):749–64.
97. Sagoo GS, Butterworth AS, Sanderson S, Shaw-Smith C, Higgins JPT, Burton H. Array CGH in patients with learning disability (mental retardation) and congenital anomalies: updated systematic review and meta-analysis of 19 studies and 13,926 subjects. *Genet Med*. 2009 Mar;11(3):139–46.
98. Waggoner D, Wain KE, Dubuc AM, Conlin L, Hickey SE, Lamb AN, et al. Yield of Additional Genetic Testing after Chromosomal Microarray for Diagnosis of Neurodevelopmental Disability and Congenital Anomalies: a clinical practice resource of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2018 Oct;20(10):1105–13.
99. Bi W, Borgan C, Pursley AN, Hixson P, Shaw CA, Bacino CA, et al. Comparison of chromosome analysis and chromosomal microarray analysis: what is the value of chromosome analysis in today’s genomic array era? *Genet Med*. 2013 Jun;15(6):450–7.
100. Hochstenbach R, van Binsbergen E, Engelen J, Nieuwint A, Polstra A, Poddighe P, et al. Array analysis and karyotyping: Workflow consequences based on a retrospective study of 36,325 patients with idiopathic developmental delay in the Netherlands. *European Journal of Medical Genetics*. 2009 Jul 1;52(4):161–9.

101. Pham J, Shaw C, Pursley A, Hixson P, Sampath S, Roney E, et al. Somatic mosaicism detected by exon-targeted, high-resolution aCGH in 10 362 consecutive cases. *Eur J Hum Genet.* 2014 Aug;22(8):969–78.
102. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature.* 1953 Apr 25;171(4356):737–8.
103. Tjio JH, Levan A. The Chromosome Number of Man. *Hereditas.* 1956;42(1–2):1–6.
104. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, et al. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature.* 1977 Feb 24;265(5596):687–95.
105. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol.* 1986;51 Pt 1:263–73.
106. Lander E, Chen C, Linton L, Birren B, Nusbaum C, Zody M, et al. Initial Sequencing and Analysis of the Human Genome. *Nature.* 2001 Feb 15;409.
107. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet.* 2010 Oct;11(10):685–96.
108. Nakka P, Pattillo Smith S, O'Donnell-Luria AH, McManus KF, Mountain JL, Ramachandran S, et al. Characterization of Prevalence and Health Consequences of Uniparental Disomy in Four Million Individuals from the General Population. *The American Journal of Human Genetics.* 2019 Nov;105(5):921–32.
109. Vetro A, Goidin D, Lesende I, Limongelli I, Ranzani G n., Novara F, et al. Diagnostic application of a capture based NGS test for the concurrent detection of variants in sequence and copy number as well as LOH. *Clinical Genetics.* 2018 Mar 1;93(3):545–56.
110. van Bon BW, Mefford HC, de Vries BB. 15q13.3 Microdeletion. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJ, Stephens K, et al., editors. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 1993
111. Szafranski P, Schaaf CP, Person RE, Gibson IB, Xia Z, Mahadevan S, et al. Structures and Molecular Mechanisms for Common 15q13.3 Microduplications Involving CHRNA7: Benign or Pathological? *Hum Mutat.* 2010 Jul;31(7):840–50.

112. Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, Weiss ME, Köster J, Marais A, et al. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *Eur J Hum Genet.* 2017 Feb;25(2):176–82.
113. Srivastava S, Cohen JS, Vernon H, Barañano K, McClellan R, Jamal L, et al. Clinical whole exome sequencing in child neurology practice. *Annals of Neurology.* 2014;76(4):473–83.
114. Srivastava S, Love-Nichols JA, Dies KA, Ledbetter DH, Martin CL, Chung WK, et al. Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet Med.* 2019;21(11):2413–21.
115. Chérot E, Keren B, Dubourg C, Carré W, Fradin M, Lavillaureix A, et al. Using medical exome sequencing to identify the causes of neurodevelopmental disorders: Experience of 2 clinical units and 216 patients. *Clin Genet.* 2018;93(3):567–76.
116. Wiszniewska J, Bi W, Shaw C, Stankiewicz P, Kang S-HL, Pursley AN, et al. Combined array CGH plus SNP genome analyses in a single assay for optimized clinical testing. *Eur J Hum Genet.* 2014 Jan;22(1):79–87.
117. Hoppman N, Rumilla K, Lauer E, Kearney H, Thorland E. Patterns of homozygosity in patients with uniparental disomy: detection rate and suggested reporting thresholds for SNP microarrays. *Genetics in Medicine.* 2018 Dec;20(12):1522–7.
118. Monroe GR, Frederix GW, Savelberg SMC, de Vries TI, Duran KJ, van der Smagt JJ, et al. Effectiveness of whole-exome sequencing and costs of the traditional diagnostic trajectory in children with intellectual disability. *Genetics in Medicine.* 2016 Sep;18(9):949–56.
119. Soden SE, Saunders CJ, Willig LK, Farrow EG, Smith LD, Petrikin JE, et al. Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Sci Transl Med.* 2014 Dec 3;6(265):265ra168.
120. Thevenon J, Duffourd Y, Masurel-Paulet A, Lefebvre M, Feillet F, El Chehadeh-Djebbar S, et al. Diagnostic odyssey in severe neurodevelopmental disorders: toward

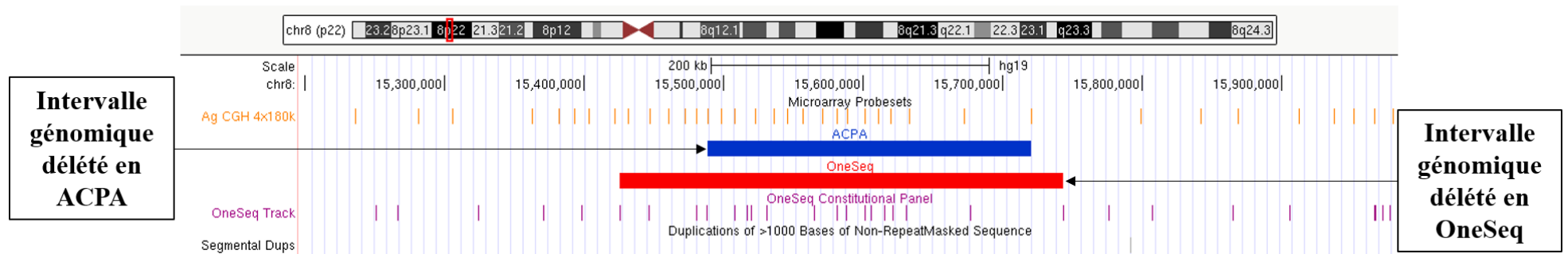
- clinical whole-exome sequencing as a first-line diagnostic test. *Clin Genet*. 2016;89(6):700–7.
121. Kleinjan DA, Lettice LA. Long-range gene control and genetic disease. *Adv Genet*. 2008;61:339–88.
  122. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature*. 2012 Apr 11;485(7398):376–80.
  123. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell*. 2015 May 21;161(5):1012–25.
  124. Ibn-Salem J, Köhler S, Love MI, Chung H-R, Huang N, Hurles ME, et al. Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol [Internet]*. 2014;15(9).
  125. Spielmann M, Lupiáñez DG, Mundlos S. Structural variation in the 3D genome. *Nature Reviews Genetics*. 2018 Jul;19(7):453–67.



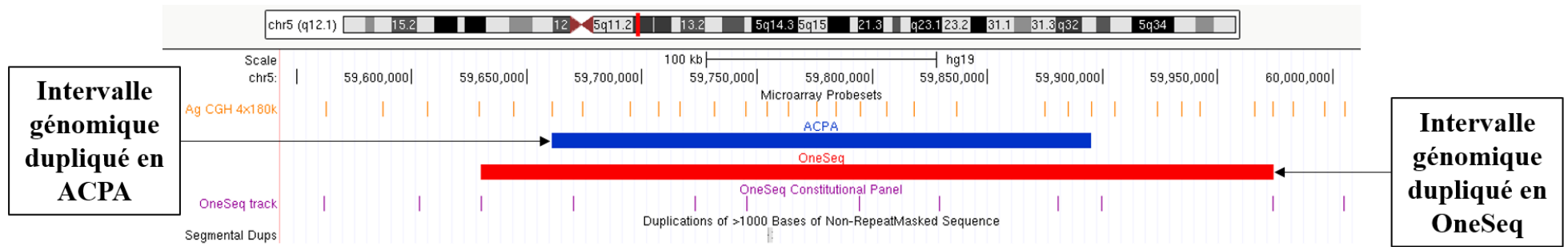
## **ANNEXES**

### **I. ANNEXE 1 : chevauchement des coordonnées génomiques entre le OneSeq et l'ACPA pour les cas n°5, 8, 9, 13, 14, 16, 18 et 19**

# Cas n°5 : délétion homozygote 8p22, chevauchement des coordonnées génomiques entre le OneSeq et l'ACPA de 84%

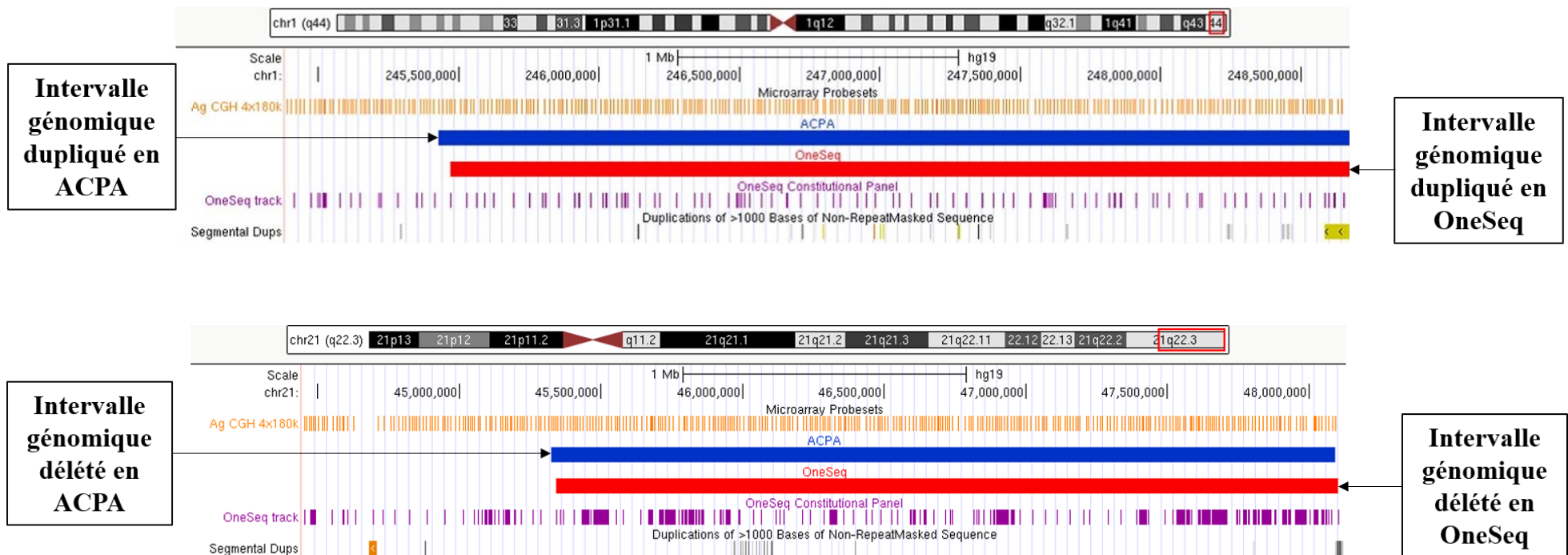


# Cas n°8 : duplication 5q12.1, chevauchement des coordonnées génomiques entre le OneSeq et l'ACPA de 80%

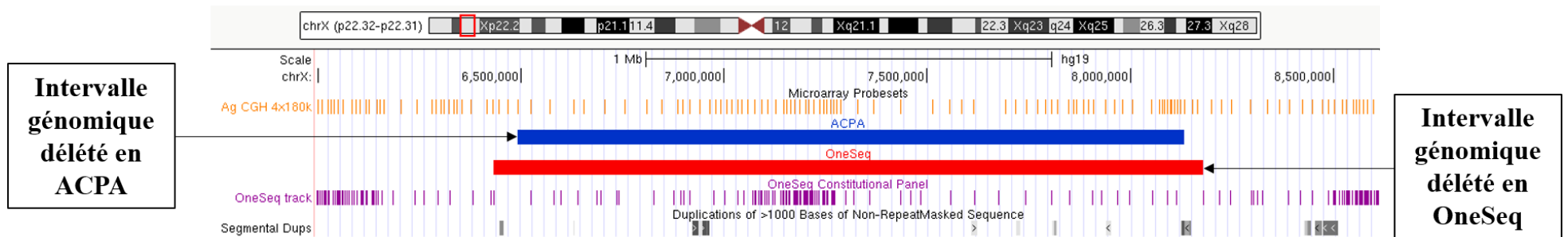




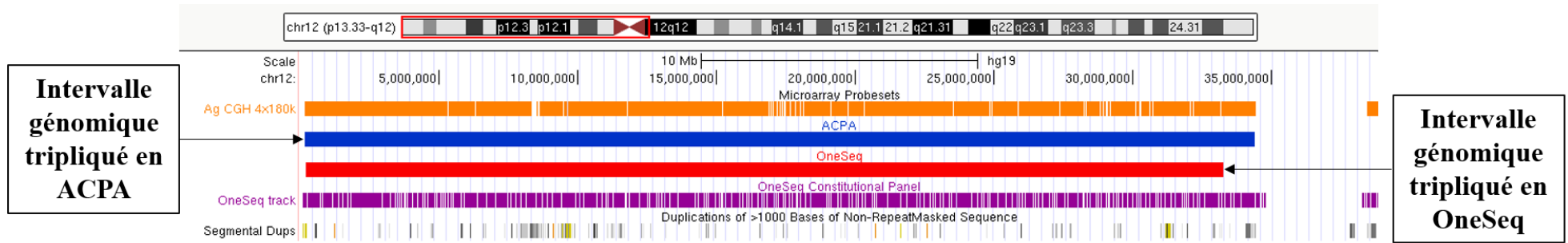
# Cas n°9 : dérivé de translocation (1;21), chevauchement des coordonnées génomiques entre le OneSeq et l'ACPA de 99%



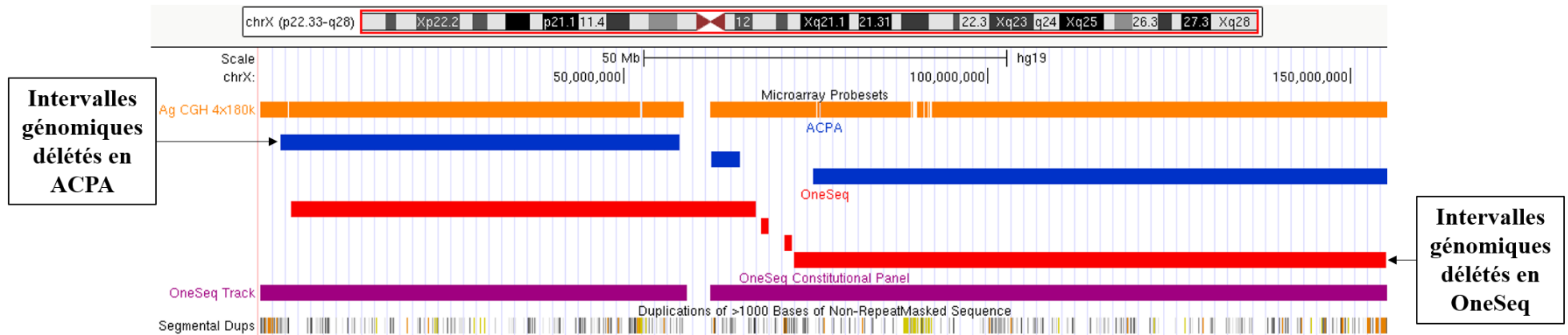
# Cas n°13 : délétion Xp22.31, chevauchement des coordonnées génomiques entre le OneSeq et l'ACPA de 96%



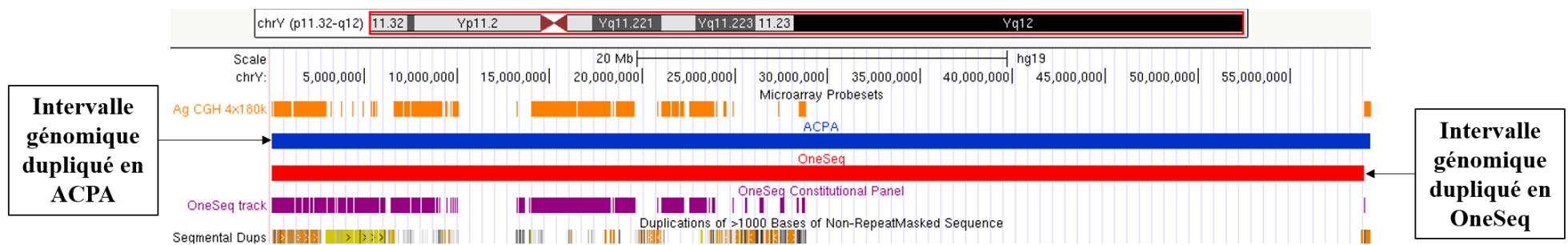
# Cas n°14 : tétrasomie 12p, chevauchement des coordonnées génomiques entre le OneSeq et l'ACPA de 98%



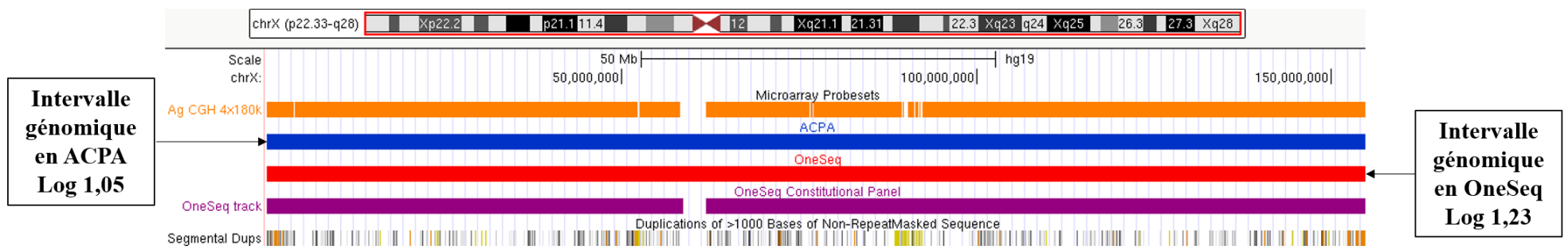
# Cas n°16 : délétion complexe en mosaïque du chromosome X, chevauchement des coordonnées génomiques entre le OneSeq et l'ACPA de 94%



# Cas n°18 : 47,XYY, chevauchement des coordonnées génomiques entre le OneSeq et l'ACPA de 99%



# Cas n°19 : tétrasomie ou pentasomie X, chevauchement des coordonnées génomiques entre le OneSeq et l'ACPA de 99%





## II. ANNEXE 2 : liste des gènes impliqués dans la DI en mars 2020 selon SysID et inclus dans l'exome clinique du OneSeq

A2ML1	AHDC1	ARID1B	B9D1	CAD	CHD4	COX5A	DHX30	EFNB1	FASN
AAAS	AHI1	ARID2	B9D2	CAMK2A	CHD7	COX6B1	DHX34	EFTUD2	FASTKD2
AARS	AIFM1	ARIH1	BBIP1	CAMK2G	CHD8	COX7B	DHX37	EGLN2	FAT1
AARS2	AIMP1	ARL13B	BBS1	CAMTA1	CHKB	COX8A	DIAPH1	EHMT1	FAT4
AASS	AK1	ARL14EP	BBS10	CAPN10	CHL1	CPLX1	DIAPH2	EIF2AK3	FBXL4
ABAT	AKT3	ARL6	BBS12	CAPN15	CHMP1A	CPS1	DIP2B	EIF2S3	FBXO11
ABCA5	ALDH18A1	ARL6IP1	BBS2	CAPN3	CHRM3	CPT1A	DISP1	EIF4A3	FBXO31
ABCC8	ALDH1A3	ARSA	BBS4	CAPRIN1	CHRNA4	CRADD	DKC1	EIF4G1	FCRL6
ABCC9	ALDH3A2	ARSE	BBS5	CARS	CHRNA7	CRB2	DLAT	ELAC2	FDPS
ABCD1	ALDH4A1	ARV1	BBS7	CARS2	CHRN2	CRBN	DLC1	ELOVL4	FGD1
ABCD4	ALDH5A1	ARX	BBS9	CASC5	CHST5	CREBBP	DLD	ELP2	FGF12
ABHD5	ALDH6A1	ASAH1	BCAP31	CASK	CIC	CRIP1	DLG2	EMC1	FGF14
ABI2	ALDH7A1	ASCC1	BCAT2	CASP2	CKAP2L	CRLF1	DLG3	EMG1	FGFR1
ABL1	ALDOA	ASCC3	BCHE	CBL	CLCN2	CSF1R	DLG4	EML1	FGFR2
ACACA	ALG1	ASCL1	BCKDHA	CBS	CLCN4	CSNK2A1	DLL1	EMX2	FGFR3
ACAD9	ALG11	ASHIL	BCKDHB	CC2D1A	CLCN7	CSPP1	DLL4	ENO3	FH
ACADS	ALG12	ASL	BCKDK	CC2D2A	CLIC2	CTBP1	DLX3	ENTPD1	FIG4
ACADSB	ALG13	ASNS	BCL11A	CCBE1	CLIP1	CTC1	DLX6	EOGT	FKRP
ACAT1	ALG14	ASPA	BCOR	CCDC115	CLN3	CTCF	DMD	EP300	FKTN
ACBD5	ALG2	ASPM	BCORL1	CCDC174	CLN5	CTDP1	DMPK	EPB41L1	FLG
ACBD6	ALG3	ASS1	BCS1L	CCDC22	CLN6	CTNNA1	DMXL2	EPG5	FLNA
ACHE	ALG6	ASTN2	BDNF	CCDC78	CLN8	CTNND2	DNA2	EPHB2	FLNB
ACO2	ALG8	ASXL1	BICD2	CCDC88C	CLP1	CTSA	DNAH3	EPRS	FLVCR1
ACOX1	ALG9	ASXL2	BLM	CCNA2	CLPB	CTSD	DNAJC19	ERCC1	FLVCR2
ACOX2	ALMS1	ASXL3	BMP4	CCND2	CLPP	CTTNBP2	DNAJC6	ERCC2	FMN2
ACP5	ALS2	ATAD3A	BMPER	CD96	CLTC	CUL3	DNM1	ERCC3	FMO4
ACSF3	ALX4	ATCAY	BOLA3	CDC42	CLTCL1	CUL4B	DNM1L	ERCC4	FMR1
ACSL4	AMHR2	ATG5	BPIFB6	CDC45	CNKSRI	CUX2	DNMT3A	ERCC5	FOLR1
ACTB	AMMECR1	ATIC	BPTF	CDC6	CNKSRI2	CWF19L1	DNMT3B	ERCC6	FOXG1
ACTC1	AMPD2	ATM	BRAF	CDH15	CNNM2	CX3CR1	DOCK6	ERCC6L2	FOXP1
ACTG1	AMT	ATN1	BRAT1	CDK13	CNOT3	CYB5R3	DOCK7	ERCC8	FOXP2
ACTL6A	ANK2	ATOH1	BRD4	CDK16	CNTN4	CYP27A1	DOCK8	ERF	FOXRED1
ACTL6B	ANK3	ATP10D	BRF1	CDK5	CNTNAP1	CYP2U1	DOLK	ERLIN2	FRAS1
ACVR1	ANKH	ATP1A2	BRPF1	CDK5RAP2	CNTNAP2	D2HGDH	DPAGT1	ERMARD	FREM2
ACY1	ANKLE2	ATP1A3	BRWD3	CDK6	COA3	DAG1	DPF2	ESCO2	FRMD4A
ADAR	ANKRD11	ATP1B1	BSCL2	CDK8	COASY	DARS	DPH1	ETHE1	FRMPD4
ADAT3	ANO10	ATP2A1	BSN	CDKL5	COG1	DARS2	DPM1	EXOSC3	FRRS1L
ADCK3	ANO5	ATP2A2	BTD	CDON	COG4	DBT	DPM2	EXOSC8	FRY
ADCY5	AP1S1	ATP2B3	BUB1B	CELSR2	COG5	DCAF17	DPM3	EXT2	FTCD
ADD3	AP1S2	ATP5A1	C10orf2	CENPE	COG6	DCC	DPP6	EZH2	FTO
ADGRG1	AP3B1	ATP5D	C12orf57	CENPF	COG7	DCHS1	DPYD	EZR	FTSJ1
ADGRG6	AP4B1	ATP6AP2	C12orf65	CENPJ	COG8	DCPS	DRP2	FA2H	FUCA1
ADIPOR1	AP4E1	ATP6V0A2	C2CD3	CEP104	COL18A1	DCX	DST	FAAH2	FZD2
ADK	AP4M1	ATP6V1B2	C5orf42	CEP120	COL27A1	DDC	DUOX1	FAM120C	FZD3
ADNP	AP4S1	ATP7A	C8orf37	CEP135	COL3A1	DDHD2	DUOX2	FAM126A	G6PD
ADRA2B	APBA2	ATP8A2	C9orf72	CEP152	COL4A1	DDOST	DUOX2	FAM20C	GABBR2
ADSL	APC2	ATPAF2	CA2	CEP19	COL4A2	DDX11	DVL2	FANCA	GABRA1
AFF2	APOPT1	ATR	CA5A	CEP290	COL4A3BP	DDX3X	DVL3	FANCB	GABRA2
AFF3	APP	ATRX	CA8	CEP41	COLEC11	DDX59	DYM	FANCC	GABRA3
AFF4	APTIX	AUH	CACNA1A	CEP57	COQ2	DEAF1	DYNC1H1	FANCD2	GABRA5
AFG3L2	AQP4	AUTS2	CACNA1B	CEP63	COQ4	DENND5A	DYRK1A	FANCE	GABRB1
AGA	ARCNI	AVPR2	CACNA1C	CEP83	COQ5	DEPDC5	EARS2	FANCF	GABRB2
AGK	ARFGEF2	B3GALNT2	CACNA1D	CEP89	COQ9	DHCR24	EBF3	FANCG	GABRB3
AGO1	ARG1	B3GALT6	CACNA1G	CFAP44	COX10	DHCR7	EBP	FANCI	GABRG2
AGPAT2	ARHGAP31	B4GALNT1	CACNA1H	CHAMP1	COX14	DHDDS	ECHS1	FANCL	GAD1
AGPS	ARHGEF6	B4GALT1	CACNA2D1	CHD1	COX15	DHFR	EDC3	FANCM	GALC
AGTR2	ARHGEF9	B4GALT7	CACNB2	CHDIL	COX20	DHRS3	EEF1A2	FAR1	GALE
AHCY	ARID1A	B4GAT1	CACNG2	CHD2	COX4I1	DHTKD1	EEF1B2	FARS2	GALT



GAMT	GRIN2B	HYLS1	KDM5A	LRRK1	MFSD8	NALCN	NPRL2	PDE10A	PITX3
GAN	GRIN2D	IBA57	KDM5B	LSS	MGAT2	NARS2	NPRL3	PDE4D	PLA2G6
GATA6	GRM1	IDH2	KDM5C	LYRM4	MGME1	NAXE	NR1I3	PDE6D	PLCB1
GATAD2B	GRM7	IDS	KDM6A	LYRM7	MIB1	NBAS	NR2F1	PDGFRB	PLCD1
GATM	GSPT2	IDUA	KDM6B	LYST	MICU1	NBN	NR3C2	PDHA1	PLEKHG2
GBA2	GSS	IER3IP1	KIAA0196	LZTFL1	MICU2	NDE1	NR4A2	PDHX	PLIN1
GCDH	GTF2E2	IFIH1	KIAA0586	LZTR1	MID1	NDNL2	NRAS	PKD2	PLK4
GCH1	GTF2H5	IFT172	KIAA1033	MAB21L2	MID2	NDP	NRXN1	PDP1	PLOD3
GCSH	GTPBP3	IFT27	KIAA2022	MAF	MIPEP	NDST1	NRXN2	PDPR	PLP1
GDI1	GUCY2D	IGBP1	KIDINS220	MAG	MIR17HG	NDUFA1	NRXN3	PDSS1	PLXNA1
GFAP	GUSB	IGF1	KIF11	MAGEA11	MKS1	NDUFA10	NSD1	PDSS2	PMM2
GFER	HACE1	IGF1R	KIF14	MAGED2	MKS1	NDUFA11	NSDHL	PECR	PMPCA
GFM1	HAX1	IGF2	KIF1A	MAGEL2	MLC1	NDUFA12	NSUN2	PEPD	PNKP
GFM2	HCCS	IKBKAP	KIF1BP	MAGI1	MLYCD	NDUFA2	NT5C2	PET100	PNP
GIGYF2	HCFC1	IKBKKG	KIF23	MAGI2	MMAA	NDUFA6	NTN1	PEX1	PNPO
GJA1	HCN1	IL1RAPL1	KIF26A	MAGT1	MMACHC	NDUFA9	NTNG1	PEX10	PNPT1
GJC2	HCN4	INPP4A	KIF2A	MAN1B1	MMADHC	NDUFAF1	NTRK1	PEX11B	POC1A
GK	HCRT	INPP5E	KIF4A	MAN2B1	MN1	NDUFAF2	NTRK2	PEX12	POFUT1
GLB1	HDAC4	INPP5K	KIF5C	MANBA	MOCS1	NDUFAF3	NUAK1	PEX13	POGZ
GLDC	HDAC6	IQSEC2	KIF7	MAOA	MOCS2	NDUFAF4	NUBPL	PEX14	POLA1
GLI2	HDAC8	IRX5	KIRREL3	MAP1B	MOGS	NDUFAF5	NUP107	PEX16	POLG
GLI3	HECW2	ISCA1	KLHL15	MAP2	MPDU1	NDUFB11	NUP214	PEX19	POLR1C
GLIS3	HEPACAM	ISCA2	KLHL24	MAP2K1	MPDZ	NDUFB3	NUP62	PEX2	POLR3A
GLRA1	HERC1	ISPD	KLHL40	MAP2K2	MPI	NDUFB8	NUS1	PEX26	POLR3B
GLRA2	HERC2	ITCH	KLHL7	MAP3K7	MPLKIP	NDUFB9	NXF5	PEX3	POMGNT1
GLS	HESX1	ITGA7	KMT2A	MAPK1	MPV17	NDUFS1	OAT	PEX5	POMGNT2
GLUL	HEXA	ITIH6	KMT2B	MAPK10	MRE11A	NDUFS2	OBSCN	PEX6	POMK
GLYCTK	HEXB	ITPA	KMT2C	MAPRE2	MRPL10	NDUFS3	OCLN	PEX7	POMT1
GM2A	HGSNAT	ITPR1	KMT2D	MASP1	MRPL3	NDUFS4	OCRL	PGAP1	POMT2
GMNN	HHAT	ITSN1	KPNA7	MAT1A	MRPS14	NDUFS6	OFD1	PGAP2	POR
GMPPA	HIBCH	IVD	KPTN	MBD5	MRPS16	NDUFS7	OGT	PGAP3	PORCN
GMPPB	HIRA	IYD	KRAS	MBNL3	MRPS2	NDUFS8	OPHN1	PGK1	POUIF1
GNAI1	HIST1H4B	JAG1	L1CAM	MBOAT7	MRPS22	NDUFV1	ORC1	PGM1	PPARGC1A
GNAI2	HIST3H3	JAM3	L2HGDH	MBTPS2	MRPS34	NDUFV2	OTC	PGM3	PPM1D
GNAO1	HIVEP2	KANK1	LAMA1	MCCC1	MRPS35	NEB	OTX2	PHACTR1	PPOX
GNAS	HIVEP3	KANSL1	LAMA2	MCCC2	MRPS7	NECAP1	PACS1	PHF21A	PPP1CB
GNB1	HK1	KARS	LAMA5	MCEE	MSL3	NEDD4L	PAFAH1B1	PHF6	PPP1R15B
GNE	HLCS	KAT6A	LAMB1	MCOLN1	MSMO1	NEU1	PAH	PHF8	PPP2R1A
GNPAT	HMGB3	KAT6B	LAMC1	MCPH1	MTF1	NF1	PAK3	PHGDH	PPP2R5D
GNPTAB	HMGCL	KATNB1	LAMC3	MCTP2	MTFMT	NFIA	PANK2	PHIP	PPT1
GNPTG	HNF1B	KCNA1	LAMP2	MDGA2	MTHFR	NFIX	PARK2	PI4K2A	PQBP1
GNS	HNMT	KCNA2	LARP7	MDH2	MTO1	NFU1	PARN	PIBF1	PREPL
GON4L	HNRNPH1	KCNA4	LARS2	MDM2	MTOR	NGLY1	PARP1	PIEZO2	PRICKLE1
GPC3	HNRNPK	KCNB1	LAS1L	MECP2	MTR	NHS	PARS2	PIGA	PRICKLE2
GPC4	HNRNPU	KCNC1	LDB3	MECR	MTRR	NIN	PAX1	PIGG	PRKCG
GPHN	HOXA1	KCNC3	LHX3	MED12	MUT	NIPBL	PAX5	PIGL	PRKD1
GPI	HPD	KCND3	LIAS	MED13	MVK	NKAP	PAX6	PIGM	PRKDC
GPSM2	HPRT1	KCNH1	LIG4	MED13L	MYCN	NKX2-1	PAX7	PIGN	PRKRA
GPT2	HRAS	KCNJ10	LINS	MED17	MYO18B	NLGN2	PAX8	PIGO	PRMT7
GPX4	HS6ST2	KCNJ11	LIPT2	MED23	MYO1A	NLGN3	PBX1	PIGT	PRMT9
GRAMD1B	HSD17B10	KCNJ6	LMAN2L	MED25	MYO1E	NLGN4X	PC	PIGV	PRODH
GRIA1	HSD17B4	KCNK9	LMBRD1	MEF2C	MYO5A	NLGN4Y	PCCA	PIGW	PROX2
GRIA2	HSPD1	KCNMA1	LMNB2	MEGF8	MYT1L	NLRP3	PCCB	PIGY	PRPS1
GRIA3	HSPG2	KCNQ2	LONP1	MEIS2	NAA10	NONO	PCDH19	PIK3C3	PRRT2
GRID2	HTR2A	KCNQ3	LPIN3	METTL23	NAA15	NOTCH1	PCDHB4	PIK3CA	PRRX1
GRIK2	HTRA2	KCNT1	LRP1	MFF	NAGA	NPC1	PCGF2	PIK3R1	PRSS12
GRIN1	HTT	KCTD7	LRP2	MFRP	NAGLU	NPC2	PCLO	PIK3R2	PSAP
GRIN2A	HUWE1	KDM1A	LRPPRC	MFSD2A	NAGS	NPHP1	PCNT	PITRM1	PSMA7

PSPH	RMND1	SEPSECS	SLC4A4	SQSTM1	TBX6	TRAPPC9	UPB1	ZFX
PTCH1	RMRP	SERAC1	SLC5A5	SRCAP	TCF12	TREX1	UPF3B	ZFYVE26
PTCHD1	RNASEH2A	SET	SLC5A7	SRD5A3	TCF20	TRIM32	UQCRQ	ZIC1
PTDSS1	RNASEH2B	SETBP1	SLC6A1	SRGAP3	TCF4	TRIM37	UROC1	ZIC2
PTEN	RNASEH2C	SETD1A	SLC6A17	SRPX2	TCF7L2	TRIO	USP27X	ZMYM3
PTF1A	RNASET2	SETD2	SLC6A3	SSR4	TCN2	TRIP4	USP9X	ZMYM6
PTK7	RNF113A	SETD5	SLC6A8	ST3GAL3	TCTN1	TRIT1	VAC14	ZMYND11
PTPN11	RNF125	SF3B1	SLC6A9	ST3GAL5	TCTN2	TRMT1	VARS	ZNF148
PTPN23	RNF135	SFXN4	SLC7A7	STAG1	TCTN3	TRMT10A	VARS2	ZNF335
PTPRD	RNF168	SGCE	SLC9A1	STAMPB	TDP1	TRMT10C	VAX1	ZNF41
PTRH2	RNU4ATAC	SGSH	SLC9A6	STARD9	TDP2	TRMT5	VCP	ZNF526
PTS	ROGD1	SHANK1	SLC9A9	STIL	TECPR2	TRPM3	VLDLR	ZNF592
PUF60	RORA	SHANK2	SLCO1C1	STIM1	TECR	TRPM6	VPS11	ZNF599
PURA	ROS1	SHANK3	SLITRK5	STRA6	TELO2	TRPS1	VPS13B	ZNF711
PUS1	RPGRIP1	SHH	SMAD4	STRADA	TFG	TRRAP	VPS33B	ZNF713
PUS3	RPGRIP1L	SHOC2	SMAD6	STT3A	TG	TSC1	VPS35	ZNF81
PYCR1	RPIA	SHOX	SMARCA2	STT3B	TGFBR1	TSC2	VPS37A	ZSWIM6
PYCR2	RPL10	SHROOM4	SMARCA4	STUB1	TGFBR2	TSEN15	VPS53	
QARS	RPL26	SIK1	SMARCB1	STX1A	TGIF1	TSEN2	VRK1	
QDPR	RPS6KA3	SIL1	SMARCC2	STX1B	TGM3	TSEN34	VWA3B	
QRICH1	RRAS2	SIX3	SMARCD1	STX3	TH	TSEN54	WAC	
QRSL1	RRM2B	SKI	SMARCD2	STXBP1	THG1L	TFSM	WARS2	
RAB18	RSPRY1	SKIDA1	SMARCE1	SUCLA2	THOC2	TSHB	WDPCP	
RAB23	RTEL1	SLC12A2	SMC1A	SUCLG1	THOC6	TSHZ3	WDR13	
RAB27A	RTN4IP1	SLC12A5	SMC3	SUFU	THRA	TSPAN7	WDR35	
RAB39B	RTTN	SLC12A6	SMG9	SUMF1	THRB	TTC19	WDR45	
RAB3GAP1	RYR2	SLC13A5	SMOC1	SUOX	THUMPD1	TTC8	WDR45B	
RAB3GAP2	RYR3	SLC16A2	SMPD1	SURF1	TIMM8A	TTI2	WDR62	
RAB40AL	SACS	SLC17A5	SMS	SUV420H1	TKT	TTN	WDR73	
RABGGTA	SAMD9	SLC18A2	SNAP25	SYN1	TLK2	TUB	WDR81	
RABL6	SAMHD1	SLC19A3	SNAP29	SYNCRIP	TM4SF20	TUBA1A	WFS1	
RAC1	SAR1B	SLC1A1	SNIP1	SYNE1	TMCO1	TUBA8	WHSC1	
RAD21	SARS	SLC1A4	SNRPB	SYNGAP1	TMEM107	TUBAL3	WNT1	
RAD50	SARS2	SLC22A5	SNRPE	SYNJ1	TMEM135	TUBB	WRAP53	
RAF1	SASS6	SLC25A1	SNTG1	SYP	TMEM138	TUBB2A	WT1	
RAI1	SATB2	SLC25A12	SNX14	SYT1	TMEM165	TUBB2B	WWOX	
RALGDS	SBDS	SLC25A15	SNX3	SYT14	TMEM199	TUBB3	XIRP1	
RANBP2	SBF1	SLC25A16	SOBP	SZT2	TMEM216	TUBB4A	XPA	
RAPGEF1	SC5D	SLC25A19	SON	TAB2	TMEM231	TUBG1	XPNPEP2	
RARB	SCAPER	SLC25A20	SOS1	TACO1	TMEM237	TUBGCP4	XRCC4	
RARS	SCN10A	SLC25A22	SOS2	TAF1	TMEM240	TUBGCP6	XYLT1	
RARS2	SCN1A	SLC25A23	SOX10	TAF2	TMEM5	TUFM	XYLT2	
RBBP8	SCN1B	SLC25A24	SOX11	TANC2	TMEM67	TUSC3	YAP1	
RBM10	SCN2A	SLC25A26	SOX2	TANGO2	TMEM70	TWIST1	YARS	
RBM28	SCN3A	SLC25A39	SOX3	TARS2	TMTC3	TXN2	YME1L1	
RBPJ	SCN8A	SLC25A4	SOX5	TASP1	TNIK	TXNRD2	YY1	
RCBTB1	SCO1	SLC25A42	SPAG17	TAT	TNPO2	UBA5	ZBTB16	
RDH11	SCO2	SLC25A46	SPAST	TBC1D20	TNPO3	UBE2A	ZBTB18	
RELN	SCP2	SLC2A1	SPATA13	TBC1D23	TNR	UBE3A	ZBTB20	
RERE	SCYL1	SLC31A1	SPATA5	TBC1D24	TP53	UBE3B	ZBTB40	
RFT1	SDCCAG8	SLC33A1	SPECC1L	TBC1D7	TP73	UBR1	ZC3H14	
RGR	SDHA	SLC35A1	SPG11	TBCD	TPI1	UBR4	ZC4H2	
RGS6	SDHAF1	SLC35A2	SPG20	TBCE	TPK1	UBR7	ZCCHC8	
RGS7	SDHB	SLC35A3	SPR	TBCK	TPO	UBTF	ZDHHC15	
RIMS1	SEC24D	SLC35C1	SPRED1	TBL1XR1	TPP1	UMPS	ZDHHC9	
RIN2	SEC61A1	SLC39A14	SPRY2	TBP	TRAIP	UNC13A	ZEB2	
RIT1	SEMA3E	SLC39A8	SPTAN1	TBR1	TRAP1	UNC45B	ZFAT	
RLIM	SEMA5A	SLC46A1	SPTBN2	TBX2	TRAPPC11	UNC80	ZFHX4	



### **III. ANNEXE 3 : protocole SureSelect XT HS – Agilent Technologies**

#### **I. Préparation de la librairie**

##### **1) Fragmentation enzymatique**

L'ADN de chaque patient était dilué dans de l'eau exempte de nucléase (ENF) pour obtenir 200ng d'ADN dans un volume final de 7 µl dans une barrette de PCR comportant 8 puits.

Pour l'étape de fragmentation enzymatique, le kit SureSelect XT HS and XT Low Input Enzymatic Fragmentation a été utilisé.

La fragmentation débutait par la préparation du mélange réactionnel de fragmentation contenant 18 µl de tampon de fragmentation et 9 µl d'enzyme de fragmentation. 3 µl de ce mélange était ajouté dans chaque puits puis placé dans le thermocycleur pour permettre l'activation enzymatique. Le programme utilisé pour le thermocycleur était le suivant : 15 minutes à 37°C, 5 minutes à 65°C puis 4°C infini.

A la fin du programme, 40 µl d'ENF était ajouté dans chaque puits.

##### **2) Réparation des extrémités des fragments d'ADN et ajout de la déoxyadénosine 5'-monophosphate non matricée**

Le mélange réactionnel de ligation était réalisé comme suit : 207 µl de tampon de ligation et 18 µl de T4 DNA Ligase. Ce mélange était conservé à température ambiante pendant au moins 45 minutes afin d'activer l'enzyme de ligation.

Le mélange réactionnel permettant la réparation des extrémités des fragments d'ADN et l'ajout de la déoxyadénosine 5'-monophosphate (dAMP) était réalisé en suivant ces indications : 144 µl de tampon End Repair-A Tailing et 36 µl d'enzyme End Repair-A Tailing.

20 µl de ce mélange était ajouté dans chaque puits contenant l'ADN fragmenté puis la barrette était placée dans le thermocycleur pour permettre d'activer l'enzyme en utilisant le programme suivant : 15 minutes à 20°C, 15 minutes à 72°C puis 4°C infini. A la fin du cycle du thermocycleur, 25 µl de mélange réactionnel de ligation était ajouté dans chaque puits. Suivait l'ajout de 5 µl d'Adaptor Oligo-mix correspondant à l'ajout de l'adaptateur et du code-barre. La barrette était ensuite placée dans le thermocycleur en suivant ce programme : 30 minutes à 20°C puis 4°C infini.

### **3) Purification des échantillons**

La purification des échantillons était réalisée avec les billes AMPure XP. Celles-ci doivent avoir été conservées à température ambiante pendant une durée supérieure à 30 minutes avant usage. A la fin du programme du thermocycleur, 80 µl de billes AMPure XP étaient ajoutées dans chaque puits et une incubation de 5 minutes à température ambiante était réalisée.

La barrette était ensuite placée sur un séparateur magnétique. Lorsque la solution était claire, le surnageant était aspiré avec une pipette multicanaux et jeté. Ensuite, il était ajouté 200 µl d'éthanol à 70% dans chaque tube et après une minute, celui-ci était enlevé. Cette étape était répétée une nouvelle fois pour réaliser deux lavages.

Lorsque tout l'éthanol avait été soigneusement enlevé, il était ajouté 35 µl d'ENF par tube. Après remise en suspension des billes et une incubation de deux minutes, la barrette était remise sur le séparateur magnétique et le surnageant était aspiré avec une pipette multicanaux et déposée dans une nouvelle barrette. Les billes pouvaient alors être jetées.

### **4) Amplification de la librairie liée à l'adaptateur**

Le mélange réactionnel de PCR pré-capture était réalisé comme suit : 90 µl de 5X Herculase Reaction Buffer, 4,5 µl de 100 nM dNTP Mix, 18 µl de Forward Primer et 9 µl d'Herculase II Fusion DNA Polymerase. 13,5 µl de ce mélange était ajouté dans chaque puits d'échantillon purifié. Il était ensuite ajouté 2 µl de SureSelect XT HS Index Primer choisi par tube.

La barrette était ensuite placée dans le thermocycleur et le programme suivant était lancé : 2 minutes à 98°C puis 8 cycles avec 30 secondes à 98°C, 30 secondes à 60°C et 1 minute à 72°C puis 5 minutes à 72°C et 4°C infini.

### **5) Purification des échantillons**

A la fin du programme du thermocycleur, 50 µl de billes AMPure XP étaient ajoutées puis cette purification était réalisée dans les mêmes conditions que précédemment sauf qu'il était rajouté seulement 15 µl d'ENF par tube après retrait de l'éthanol.

A l'issue de cette purification, un nouveau dosage de l'ADN par technique fluorométrique avec le Qubit<sup>TM</sup> (ThermoFisher Scientific) était réalisée.

## **6) Hybridation des échantillons d'ADN avec la librairie de capture**

L'ADN de chaque patient était dilué dans de l'eau exempte de nucléase (ENF) pour obtenir 1000ng d'ADN dans un volume final de 12 µl dans une barrette de PCR comportant 8 puits.

5 µl de SureSelect XT HS and XT Low Input Blocker Mix était ajouté dans chaque tube. Cet ajout permet de bloquer l'adaptateur pour éviter des hybridations non spécifiques.

La barrette était ensuite placée dans le thermocycleur en suivant ce programme : 5 minutes à 95°C, 10 minutes à 65°C puis 1 minute à 65°C avec une pause du thermocycleur.

Pendant le temps des deux premiers cycles du thermocycleur, le mélange réactionnel de bloquant la RNase était réalisé comme suit : 4,5 µl de SureSelect RNase Block et 13,5 µl d'ENF. Juste avant la pause du thermocycleur à 65°C, le mélange réactionnel d'hybridation de la librairie de capture était réalisé comme suit : 18 µl de mélange réactionnel de RNase, 45 µl de librairie de capture OneSeq et 54 µl de SureSelect Fast Hybridization Buffer.

Quand le thermocycleur débute sa pause à 65°C, ajouter 13 µl du mélange réactionnel d'hybridation dans chaque puits directement dans le thermocycleur. Le thermocycleur poursuit alors avec le programme suivant : 60 cycles avec 1 minute à 65°C et 3 secondes à 37°C puis 65°C infini.

## **7) Capture des ADN hybridés avec les billes de streptavidine**

50 µl de billes de streptavidine par échantillon étaient d'abord lavées avec du SureSelect Binding Buffer. Elles étaient ensuite placées sur un séparateur magnétique et le surnageant était jeté. Cette étape était répétée trois fois puis les billes étaient re-suspendues dans 200 µl de SureSelect Binding Buffer. A la fin du programme du thermocycleur, l'intégralité de chaque échantillon était transférée dans les puits contenant les billes lavées. Les tubes étaient ensuite incubés 30 minutes en mixant à 1800 tours par minute à température ambiante.

Pendant ce temps, 6 aliquots de 200 µl de Wash Buffer 2 étaient réalisés pour chaque échantillon. Ils étaient ensuite incubés à 70°C.

A la fin de l'incubation, la barrette était placée sur le séparateur magnétique puis le surnageant était aspiré et jeté. Les billes étaient ensuite re-suspendue dans 200 µl de SureSelect Wash Buffer 1. La barrette était à nouveau placée sur le séparateur magnétique et quand le surnageant était clair, il était enlevé et jeté.

Les billes étaient ensuite lavées en gardant la température à 70°C pour permettre une meilleure stringence :

- La barrette était mise dans le thermocycleur à 70°C
- Les billes étaient mises en suspension dans 200 µl de SureSelect Wash Buffer 2 à 70°C puis incubées 5 minutes à 70°C
- La barrette était ensuite placée sur le séparateur magnétique, quand le surnageant était clair il était enlevé et jeté

Il fallait ainsi réaliser 6 lavages. Après le dernier lavage, 25 µl d'ENF était ajoutée dans chaque échantillon.

## **8) Amplification des bibliothèques capturées**

Le mélange réactionnel de PCR post-capture était réalisé comme suit : 112,5 µl d'ENF, 90 µl de 5x Herculase II Reaction Buffer, 9 µl d'Herculase II Fusion DNA Polymerase, 4,5 µl de 100 nM dNTP Mix et 9 µl de SureSelect Post Capture Primer Mix. 25 µl de ce mélange réactionnel était ajouté dans chaque puits.

La barrette était ensuite placée dans le thermocycleur et le programme suivant était lancé : 2 minutes à 98°C puis 9 cycles avec 30 secondes à 98°C, 30 secondes à 60°C et 1 minute à 72°C puis 5 minutes à 72°C et 4°C infini. Il s'agit ici d'une PCR sur billes.

A la fin du programme, la barrette était placée sur le séparateur magnétique. Le surnageant était alors déposé dans une nouvelle barrette puis les billes étaient jetées.

## **9) Purification des bibliothèques amplifiées**

A la fin du programme du thermocycleur, 50 µl de billes AMPure XP étaient ajoutées puis cette purification était réalisée dans les mêmes conditions que précédemment sauf qu'il était rajouté seulement 25 µl d'ENF par tube après retrait de l'éthanol.

A l'issue de cette purification, un nouveau dosage de l'ADN par technique fluorométrique avec le Qubit™ (ThermoFisher Scientific) était réalisée.

La concentration de chaque bibliothèque était déterminée en fonction de la taille moyenne des fragments vérifiée sur le 2100 Bioanalyzer instrument™ (Agilent Technologies).

Les bibliothèques étaient ensuite combinées de façon à ce que chaque échantillon indexé soit présent de façon équimolaire dans le mélange.

## **II. Préparation des échantillons pour le séquençage**

### **1) Préparation de la cassette de réactifs et de la flow cell**

La cassette de réactifs Illumina High Output était mise à décongeler dans un bain d'eau à température ambiante pendant 1 heure. Après vérification de la décongélation, la cassette était séchée avec du papier absorbant puis retournée 5 fois afin de mélanger les réactifs.

La flow cell High Output était sortie du frigo et gardée à température ambiante pendant 30 minutes. Elle était ensuite sortie de son emballage et la surface en verre était nettoyée avec un tissu anti-poussière.

### **2) Préparation de la librairie**

La librairie était d'abord dénaturée avec de la soude (NaOH) pour obtenir de l'ADN simple-brin.

Pour une librairie concentrée à 4nM, il était ajouté 5 µl de librairie et 5 µl de NaOH à 0,2 N.

Après 5 minutes d'incubation, 5 µl de tampon HT1 était ajouté afin d'arrêter la dénaturation.

La librairie dénaturée était ensuite diluée à 20 pM en ajoutant 985 µl de HT1 froid.

Enfin, la librairie était diluée à une concentration de 1,8 pM pour permettre le chargement de la cassette de réactifs. Pour cela, 1183 µl de HT1 froid était ajouté à 117 µl de librairie dénaturée. Cette dernière dilution permet d'obtenir un volume final de librairie de 1,3 ml.

La librairie était ensuite entièrement chargée au niveau de l'emplacement « load library here » de la cassette de réactifs après avoir percé l'opercule en aluminium.

### **3) Lancement du séquenceur NextSeq**

Le séquenceur NextSeq550<sup>TM</sup> d'Illumina était ensuite lancé avec les paramètres suivants en Paired End : Read 1: 150 cycles, Index 1: 8 cycles, Index 2: 10 cycles, Read 2: 150 cycles

La durée du séquençage était d'environ 24 heures.

A la fin du séquençage, la sample sheet contenant les paramètres du run et notamment les séquences associées aux index spécifiques des patients étaient téléversés avec les fichiers .BCL sur le logiciel « bcl2fastq » pour permettre le démultiplexage des échantillons.

Les fichiers .FASTQ étaient ensuite téléversés sur la plateforme Alissa Align & Call (Agilent Technologies) pour la suite du traitement bio-informatique.





#### **IV. ANNEXE 4 : Serment d'Hippocrate**

Au moment d'être admis à exercer la médecine, je promets et je jure d'être fidèle aux lois de l'honneur et de la probité.

Mon premier souci sera de rétablir, de préserver ou de promouvoir la santé dans tous ses éléments, physiques et mentaux, individuels et sociaux.

Je respecterai toutes les personnes, leur autonomie et leur volonté, sans aucune discrimination selon leur état ou leurs convictions. J'interviendrai pour les protéger si elles sont affaiblies, vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l'humanité.

J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences. Je ne tromperai jamais leur confiance et n'exploiterai pas le pouvoir hérité des circonstances pour forcer les consciences.

Je donnerai mes soins à l'indigent et à quiconque me les demandera. Je ne me laisserai pas influencer par la soif du gain ou la recherche de la gloire.

Admis dans l'intimité des personnes, je tairai les secrets qui me seront confiés. Reçu à l'intérieur des maisons, je respecterai les secrets des foyers et ma conduite ne servira pas à corrompre les mœurs.

Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement les agonies. Je ne provoquerai jamais la mort délibérément.

Je préserverai l'indépendance nécessaire à l'accomplissement de ma mission. Je n'entreprendrai rien qui dépasse mes compétences. Je les entretiendrai et les perfectionnerai pour assurer au mieux les services qui me seront demandés.

J'apporterai mon aide à mes confrères ainsi qu'à leurs familles dans l'adversité. Que les hommes et mes confrères m'accordent leur estime si je suis fidèle à mes promesses ; que je sois déshonoré et méprisé si j'y manque.



## ETUDE DES PERFORMANCES D'UNE TECHNOLOGIE DE SEQUENCAGE HAUT DEBIT PANGENOMIQUE POUR LA DETECTION DES COPY NUMBER VARIATIONS : COMPARAISON AVEC L'ACPA

**Introduction :** La Déficience Intellectuelle (DI) est fréquente et hétérogène. Des causes génétiques d'origine génique (20-30%) ou chromosomique (10-15%) sont impliquées dans environ 50% des cas. Le diagnostic étiologique génétique comprend notamment une Analyse Chromosomique sur Puce à ADN (ACPA) puis le séquençage d'un ou plusieurs panels de gènes. L'étude portait sur une technologie innovante permettant la détection simultanée des mutations ponctuelles, d'anomalies chromosomiques déséquilibrées et de régions d'homozygotie. L'objectif était d'évaluer ses performances dans l'identification des délétions et duplications chromosomiques (Copy Number Variations) par comparaison avec l'ACPA.

**Matériel et méthodes :** 30 patients à génotype connu ont été testés avec une capture OneSeq Constitutional Research Panel.

**Résultats :** l'analyse a permis d'obtenir le bon résultat chez 27 des 30 patients testés (33 variants sur 36).

**Discussion :** les CNV non détectés sont liés soit au paramétrage bio-informatique (1 CNV), soit à l'arborescence filtrant les variants (1 CNV d'interprétation débattue), soit au design de l'analyse (1 CNV trop petit). Ils pourraient être identifiés par optimisation de certains paramètres. Une étude médico-économique a montré un coût équivalent à celui de l'association ACPA et séquençage d'un panel de gènes.

**Conclusion :** la sensibilité de cette technologie est très similaire à celle de l'ACPA et pourrait être une alternative à l'association ACPA-séquençage d'un panel de gènes en routine. Le rendement diagnostique serait identique ce qui devrait permettre de diminuer l'errance diagnostique et le coût analytique liés à la multiplication d'exams génétiques.

**Mots-clefs :** déficience intellectuelle, séquençage haut débit, copy number variation, mutation ponctuelle, single nucleotide polymorphism, région d'homozygotie

---

## PERFORMANCE STUDY OF A CAPTURE BASED NGS TEST FOR DETECTION OF COPY NUMBER VARIATIONS: COMPARISON TO ARRAY-CGH

**Introduction:** Intellectual Disability (ID) is a common condition and is extremely heterogeneous. Genetic causes are involved in approximately 50% of cases and may be the result from gene mutations (20-30%) or chromosomal aberrations (10-15%). The genetic etiological diagnosis includes a Comparative Genomic Hybridization Array (array-CGH) then, next-generation sequencing (NGS) genes panels. We tested a novel NGS technology enabling combined detection of gene mutations, chromosomal unbalanced aberrations and regions of homozygosity. Our objective was to assess this method performance to detect chromosomal copy number variations (CNV) compared to array-CGH.

**Material and methods:** 30 patients with known genotype were analyzed with the OneSeq Constitutional Research Panel capture design.

**Results:** correct identification was established in 27 of the 30 patients tested (33 variants/36)

**Discussion:** undetected CNV are related either to chosen bioinformatics parameters (1 CNV), to the variant classification tree (1 CNV of questionable clinical significance) or the analysis design (1 very small CNV). They could be identified by optimizing some parameters. A medico-economic study has shown a comparable cost than array-CGH and genes sequencing panel association.

**Conclusion:** this NGS technology sensitivity is very similar to array-CGH for CNV detection and could be an alternative to array-CGH and NGS genes panel association in routine care. The expected diagnosis yield is identical and should reduce time to diagnosis and analytical cost.

**Keywords:** intellectual disability, next generation sequencing, copy number variation, point mutation, single nucleotide polymorphism, region of homozygosity