



HAL
open science

**Caractérisation moléculaire des remaniements
chromosomiques chez des patients atteints d'anomalies
congénitales ou de troubles de la reproduction par
technique de séquençage Haut-débit**

Tony Yammine

► **To cite this version:**

Tony Yammine. Caractérisation moléculaire des remaniements chromosomiques chez des patients atteints d'anomalies congénitales ou de troubles de la reproduction par technique de séquençage Haut-débit. Génétique humaine. Université de Lyon; Université Saint-Joseph (Beyrouth), 2020. Français. NNT : 2020LYSE1339 . tel-03618968

HAL Id: tel-03618968

<https://theses.hal.science/tel-03618968>

Submitted on 24 Mar 2022

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.



N° d'ordre NNT : 2020LYSE1339

**THESE de DOCTORAT en cotutelle DE L'UNIVERSITE
DE LYON et l'UNIVERSITE SAINT-JOSEPH DE
BEYROUTH**

opérée au sein de
l'Université Claude Bernard Lyon 1

**Ecole Doctorale N° 476
Neurosciences et cognition**

Spécialité de doctorat : Science de la vie

Discipline : Génétique Humaine

Et

l'Université Saint-Joseph de Beyrouth

Ecole Doctorale Sciences et santé

Soutenue publiquement le 17/12/2020, par :

Tony YAMMINE

**Caractérisation moléculaire des remaniements
chromosomiques chez des patients atteints d'anomalies
congénitales ou de troubles de la reproduction par
technique de séquençage Haut-débit**

Devant le jury composé de :

SCHAEFFER, Laurent	PU-PH	Université Claude Bernard Lyon 1	Président
VIALARD, François	PU-PH	Université Versailles St Quentin	Rapporteur
KHOUEIRY, Pierre	MCU	American University of Beirut	Rapporteur
MALAN, Valérie	MCU-PH	Université Paris Descartes	Examinatrice
ABI FADEL, Marianne	PU	Université Saint-Joseph de Beyrouth	Examinatrice
SANLAVILLE, Damien	PU-PH	Université Claude Bernard Lyon 1	Directeur
SCHLUTH-BOLARD, Caroline	MCU-PH	Université Claude Bernard Lyon 1	Co-directrice
FARRA, Chantal	PU-PH	Université Saint-Joseph de Beyrouth	Directrice

Remerciements

La réalisation de cette thèse et la contribution au succès de mon parcours ont été possible grâce à plusieurs personnes à qui je voudrais témoigner toute ma gratitude et mes remerciements les plus chaleureux,

Professeur Damien Sanlaville, je voudrais vous adresser toute ma reconnaissance et mes remerciements. Merci pour vos judicieux conseils qui ont assuré un grand professionnalisme et pour toutes les précieuses remarques qui ont guidé et alimenté mes réflexions. Merci pour toutes les brillantes idées qui m'ont toujours inspiré pour faire de mon mieux. Je vous remercie d'avoir cru en mon projet de thèse et lui avoir permis de voir le jour. Je vous remercie de m'avoir fait l'honneur d'être le directeur de ma thèse.

Docteure Caroline Schluth-Bolard, je vous remercie très sincèrement d'avoir accepté avec beaucoup de gentillesse d'être la co-directrice. Vous avez dédié votre temps à la lecture intelligente et respectueuse de ma thèse, veuillez recevoir mes remerciements les plus chaleureux. Je vous remercie pour le sérieux et l'efficacité avec laquelle vous avez pris en charge ma clé de réussite. Je vous remercie pour votre soutien sincère et généreux, pour vos conseils bienveillants et consécutifs qui m'ont énormément aidé tout au long de mon parcours.

Professeur Chantal Farra, je vous remercie pour votre professionnalisme, votre disponibilité, votre gentillesse, votre patience et vos conseils. Je vous remercie d'avoir cru en moi et d'avoir parcouru avec moi le chemin menant à la réalisation de cette thèse. Je tiens à vous témoigner ma plus vive reconnaissance.

Aux membres du jury, Pr. François Vialard, Pr. Ass. Pierre Khoueiry, Pr. Laurent Schaeffer, Dr. Valérie Malan et Pr. Marianne Abi Fadel. Je vous remercie d'avoir accepté de participer à ce jury. Merci pour votre écoute, votre attention, et votre implication. Votre présence au sein du jury est un grand honneur pour moi.

Professeur Roland Tomb, doyen de la faculté de médecine de l'USJ, je vous remercie chaleureusement de m'avoir permis de réaliser cette thèse. Votre soutien compte beaucoup pour moi. Votre confiance en moi, votre support et votre aide m'ont poussé à franchir les

obstacles et à continuer droit vers le but de l'excellence. Veuillez recevoir mes plus sincères remerciements.

Révérénd Père professeur Michel Sheuer, je vous remercie profondément d'avoir fourni tous les outils nécessaires à la réussite de mes études universitaires. Votre soutien inestimable a donné naissance à tous mes projets. Veuillez recevoir mes remerciements les plus chaleureux,

Pr. Ass. Eliane Kfoury, malgré les conditions qui ne nous ont pas permis de continuer cette dernière étape ensemble, je vous remercie d'avoir accepté de m'encadrer pendant les premières années de mon parcours. Merci pour tous vos conseils qui resteront gravés pour toujours.

Dr. Marianne Till, Dr. Gaëtan Lesca, Dr. Nicolas Chatron, Dr. Pierre-Antoine, Flavie Diguët, à toute l'équipe du laboratoire de cytogénétique constitutionnelle et le service de génétique du CHU de Lyon. Les internes Maude, Kevin, Idriss, Dorian, Thibaut, Afef, Rahma, ... et ma collègue Julie, vous m'avez accueilli à Lyon les bras grands ouverts avec le sentiment d'être chez moi, dans mon propre pays. Depuis ma première année, vous m'avez formé et aidé dans chaque étape. Vos encouragements et votre soutien moral, intellectuel et inconditionnel tout au long de ma démarche a été d'une grande aide.

À tous mes collègues de l'Unité de Génétique Médicale et toute l'équipe pédagogique de l'Université Saint-Joseph de Beyrouth qui m'ont toujours motivé et encouragé. Ce travail n'aurait pu voir le jour sans votre présence et votre soutien. Merci pour votre implication et vos conseils. Je vous remercie d'être toujours à mes côtés, dans les bons comme dans les mauvais moments. Votre amour et votre sagesse m'ont toujours poussé à atteindre mon but. Je suis très heureux et très fier de vous connaître.

Je tiens également à remercier les patients et leur famille d'avoir accepté de participer à cette étude de recherche. Votre participation a été d'une grande aide pour la réalisation de mon succès.

À ma famille, votre soutien sans faille, votre support et vos encouragements m'ont

permis d'arriver là où j'en suis aujourd'hui et être qui je suis. Grâce à vous et à votre confiance en moi, je me réalise et j'avance. Grâce à votre amour et votre éducation, vous avez fait de mon parcours une réussite. Grâce à vous, je me construis et je m'épanouis, tant d'un point de vue personnel que professionnel. Vous m'avez toujours encouragé à aller plus loin, à en faire plus, à me dépasser et toujours aller de l'avant. Cet ouvrage est l'aboutissement de vos efforts et de vos sacrifices. Je vous en suis infiniment reconnaissant.

Résumé

Titre : Caractérisation moléculaire de remaniements chromosomiques chez des patients atteints d'anomalies congénitales ou de troubles de la reproduction par technique de séquençage Haut-débit.

L'avènement des nouvelles technologies, notamment, le séquençage génomique à haut débit a montré son efficacité dans la caractérisation au niveau moléculaire des points de cassures de remaniements chromosomiques équilibrés et déséquilibrés. Cette technique a de plus, l'avantage d'identifier non seulement des variations génomiques de structure équilibrée ou déséquilibrées (CNV) mais également de variations nucléotidiques (SNV). Dans le cadre de ce travail, nous avons tout d'abord appliqué cette approche pour étudier trois patients présentant des altérations spermatiques et un remaniement chromosomique apparemment équilibrés (RCAE). Environ 7% de la population masculine générale est touchée par des problèmes d'infertilité. Les RCAEs sont identifiés chez 0,4 à 1,4% des hommes infertiles et sont considérés comme une cause bien établie de l'infertilité masculine. En plus de l'altération mécanique de la spermatogenèse, ces RCAEs pourraient hypothétiquement interrompre ou modifier l'expression par un effet de position d'un ou plusieurs gènes. De plus une association fortuite de ces RCAEs avec des variations nucléotidiques géniques pourrait être une cause de l'infertilité masculine. Jusqu'à présent peu d'études sur la caractérisation moléculaire de ces RCAE dans les infertilités ont été publiées. Nous avons pu confirmer que l'étude complet du génome, couplé à d'autres techniques de cytogénétiques et cytogénétiques moléculaires, permet une caractérisation précise des RCAEs. Par ailleurs, notre étude a permis d'identifier parmi nos trois patients deux gènes candidats, *LACTB2* et *DNAH5*, impliqués dans l'infertilité masculine. Dans un deuxième temps, nous avons utilisé le séquençage haut débit pour caractériser les remaniements chromosomiques équilibrés et déséquilibrés de neuf patients (cinq déséquilibrés et quatre équilibrés) atteints d'anomalies congénitales multiples. La totalité des points de cassures identifiés au niveau du caryotype ont été caractérisés, montrant que la majorité interviennent au niveau des éléments répétés. La caractérisation précise des points de cassures étant plus facilement réalisable, le défi actuel réside dans l'implication de gènes ou de variants nucléotidiques impliqués dans la pathologie présentée par les patients. Dans notre étude, onze gènes candidats ont été identifiés et devraient être l'objet d'analyses fonctionnelles afin de confirmer ou d'infirmer leur implication dans le phénotype des patients. Enfin, la technique de séquençage à haut débit a permis d'identifier la présence de remaniement complexes de type *chromoanagenesis* chez quatre patients soulignant que les RCAEs sont probablement plus fréquents, que nous pouvons l'imaginer en pathologie constitutionnelle. En conclusion, le séquençage à haut débit est une technique efficace qui permet rapidement de caractériser les remaniements chromosomiques et d'identifier les CNV et les SNV grâce à des outils bio-informatiques adéquats. Avec les progrès technologiques et l'arrivée des séquenceurs de troisième génération permettant de séquencer de longs fragments d'ADN, la question du transfert en diagnostic est posée.

Mots clés : Séquençage à haut débit, remaniements chromosomiques, infertilités masculines, anomalies congénitales.

Abstract

Title: Molecular characterization of chromosomal rearrangements in patients with congenital anomalies or reproductive disorders by high-throughput sequencing

The rise of new technologies, in particular, whole genome sequencing has shown its efficiency in characterizing at the molecular level the breakpoints of balanced and unbalanced chromosomal rearrangements. In addition, this technique has the advantage of detecting the presence of genomic copy number variations (CNVs) and nucleotide variations (SNVs). In this study, we first applied this approach to characterize apparently balanced chromosomal rearrangements (ABCR) in three patients with sperm alterations. Approximately 7% of the general male population are affected by infertility problems. ABCR are identified in 0.4 to 1.4% of infertile men and are considered to be a well-established cause of male infertility. In addition to the mechanical alteration of spermatogenesis, these ABCRs could hypothetically disrupt one or more genes or modify their expression through position effect. Furthermore, a fortuitous association of these RCAEs with gene variations could be a cause of male infertility. So far, few studies on the molecular characterization of these ABCRs in infertility have been published. This study confirms the relevance of this technique with other cytogenetic and molecular cytogenetic techniques in the characterization of these rearrangements. Our analysis identified two candidate genes, *LACTB2* and *DNAH5*, involved in male infertility. In a second approach, this technique was used to characterize balanced and unbalanced chromosomal rearrangements in nine patients (five unbalanced and four balanced) with multiple congenital anomalies. All the breakpoints identified by karyotype have been characterized, showing that the majority occur at the level of repeated elements. Nowadays the characterization at the molecular level of chromosomal rearrangements breakpoints is easier, the current challenge lies in the authentication of genes or nucleotide variants involved in the pathology. In this study, eleven candidate genes were identified and should be subjected to functional analysis to confirm or rule out their involvement in the phenotype of patients. The high-throughput sequencing technique used in our study also revealed in four patients, the presence of complex rearrangement, *chromoanagenesis*, highlighting that ABCR are probably more frequent than we can imagine in constitutional anomalies. In conclusion, high-throughput sequencing is an efficient technique that allows rapid characterization of chromosomal rearrangements and identification of CNVs and SNVs using appropriate bioinformatics tools. With technological progress and the arrival of third-generation sequencers allowing the sequencing of long DNA fragments, the question of the transfer of this technique in diagnosis is raised.

Key words: High-throughput sequencing, chromosomal rearrangement, male infertility, congenital anomalies.

Table des matières

1. Introduction.....	4
1.1. Anomalies chromosomiques.....	6
1.1.1. Les différents types d'anomalies chromosomiques.....	6
1.1.1.1. Les anomalies de nombres.....	6
1.1.1.2. Les anomalies de structure.....	7
1.1.1.2.1. Les anomalies chromosomiques de structure impliquant un seul chromosome.....	8
1.1.1.2.2. Les anomalies chromosomiques de structure impliquant plusieurs chromosomes ..	12
1.1.1.3. Remaniement chromosomique complexe.....	17
1.1.1.4. Autres types d'anomalies chromosomiques.....	22
1.2. Mécanismes de formation des anomalies chromosomiques de structure.....	24
1.2.1. Les réarrangements récurrents.....	24
1.2.2. Les réarrangements non récurrents.....	26
1.2.2.1. Fork Stalling and Template Switching (FoSTeS).....	27
1.2.2.2. Microhomology-mediated break-induced replication (MMBIR).....	28
1.3. Méthodes d'étude des remaniements chromosomiques.....	29
1.3.1. Caryotype.....	29
1.3.2. FISH.....	31
1.3.3. a-CGH ou ACPA.....	34
1.4. Les conséquences des anomalies chromosomiques.....	38
1.5. Troubles de la reproduction : infertilité masculine.....	42
1.5.1. Généralité.....	42
1.5.2. La spermatogenèse.....	43
1.5.3. La spermocytogenèse.....	44
1.5.3.1. Phase de multiplication.....	44
1.5.3.2. Phase d'accroissement.....	44
1.5.3.3. Phase de maturation.....	45
1.5.4. La spermiogenèse.....	47
1.5.4.1. La formation de l'acrosome.....	49
1.5.4.2. La formation du flagelle.....	49
1.5.5. Contrôle neuro-endocrinien de la fonction testiculaire.....	50
1.5.6. Régulation thermique.....	51
1.6. L'infertilité masculine.....	51
1.6.1. Les causes de l'infertilité masculine.....	51
1.6.2. Le diagnostic de l'infertilité masculine.....	52
1.6.3. Spermogramme et spermocytogramme.....	53
1.6.4. Génétique de l'infertilité masculine.....	57

1.6.4.1.	Les anomalies chromosomiques.....	57
1.6.4.2.	Anomalies chromosomiques de nombre.....	57
1.6.4.3.	Anomalies chromosomiques de structures.....	58
1.6.4.4.	Microremaniements cryptiques et infertilité	60
1.6.4.5.	Gènes et infertilité masculine.....	61
1.7.	Les anomalies du développement humain.....	65
1.7.1.	Introduction.....	65
1.7.2.	Les étiologies des anomalies congénitales.....	67
1.7.3.	Les causes génétiques.....	67
1.7.3.1.	Les syndromes microdélétionnels et microduplicationnels	68
1.7.3.2.	Les causes géniques.....	69
1.7.3.3.	Les facteurs extrinsèques	69
1.8.	Apport du NGS dans l'étude des remaniements chromosomiques	70
1.8.1.	Bref historique du séquençage	70
1.8.2.	Principe et technique du séquençage du génome entier de la technologie Illumina	73
1.8.2.1.	Le Whole Genome Sequencing.....	74
2.	Matériels et Méthodes.....	83
2.1.	Hybridation in situ en fluorescence (FISH).....	83
2.2.	Analyse chromosomique sur puce à ADN de la technologie Affymetrix	83
2.3.	Caractérisation des points de cassure par WGS.....	83
2.3.1.	Analyse bio-informatique	84
2.3.2.	Amplification par PCR de fragments de jonction contenant les points de cassure et séquençage Sanger.....	85
2.4.	Séquençage d'exome.....	86
2.5.	Etude d'expression par Reverse transcription (RT) qPCR	88
2.5.1.1.	PCR quantitative (qPCR).....	88
3.	Caractérisation moléculaire des RCAE chez 3 patients présentant une infertilité masculine	90
3.1.	Patient 1 :.....	91
3.2.	Patient 2 :.....	92
3.3.	Patient 3 :.....	92
3.3.1.	Résultat.....	93
3.3.1.1.	Patient 1	93
3.3.1.2.	Patient 2	100
3.3.1.3.	Patient 3	103
3.4.	Discussion.....	106
4.	Caractérisation moléculaire de remaniements chromosomiques chez des patients atteints de déficience intellectuelle et/ou de malformations congénitales.....	108

4.1.	Caractérisation moléculaire de remaniements chromosomiques déséquilibrés chez des patients atteints de déficience intellectuelle et/ou de malformations congénitales.....	108
4.1.1.	Patient 1	109
4.1.2.	Patient 2.....	114
4.1.3.	Patient 3.....	117
4.1.4.	Patient 4 (Fœtus 1)	121
4.1.5.	Patiente 5	129
4.2.	Caractérisation moléculaire de remaniements chromosomiques apparemment équilibrés chez des patients atteints de déficience intellectuelle et/ou de malformations congénitales.....	138
4.2.1.	Patient 6.....	138
4.2.2.	Patient 7.....	144
4.2.3.	Patient 8.....	149
4.2.4.	Patient 9 (Fœtus 2)	154
4.3.	Discussion.....	163
5.	Discussion	165
6.	Conclusion	170
7.	Annexe 1 :.....	171
8.	Annexe 2 :.....	172
9.	Bibliographie.....	174

Liste des figures :

- Figure 1:** La non-disjonction méiotique des chromosomes au cours de la première et deuxième division méiotique
- Figure 2:** Mécanisme de formation d'une délétion terminale et d'une délétion interstitielle
- Figure 3:** Mécanisme de formation d'une duplication en tandem ou inversée
- Figure 4:** Mécanisme de formation d'un chromosome en anneau
- Figure 5:** Mécanisme de formation d'une inversion péricentrique et paracentrique.
- Figure 6:** Etude de la ségrégation de l'inversion péricentrique.
- Figure 7:** Etude de la ségrégation de l'inversion paracentrique d'un chromosome acrocentrique.
- Figure 8:** Mécanisme de formation d'un isochromosome monocentrique ou dicentrique.
- Figure 9:** Mécanisme de formation d'une insertion.
- Figure 10:** Mécanisme de formation d'une translocation robertsonienne.
- Figure 11:** Etude de la ségrégation d'une translocation robertsonienne.
- Figure 12:** Mécanisme de formation d'une translocation réciproque
- Figure 13:** Etude de la ségrégation d'une translocation réciproque.
- Figure 14:** Les trois types de réarrangements chromosomiques complexes.
- Figure 15:** Mécanisme du chromothripsis
- Figure 16:** Principe du processus de chromoanasythesis.
- Figure 17:** Mécanisme de chromoplexie
- Figure 18:** Apoptose abortive.
- Figure 19:** Schéma illustrant l'hypothèse de la formation des chromoanagenesis et chromothripsis dans le micronoyau
- Figure 20:** Vue à l'échelle du génome des duplications segmentaires
- Figure 21:** Réarrangements génomiques résultant de la recombinaison homologue entre les duplications segmentaires
- Figure 22:** Le mécanisme du NHEJ
- Figure 23:** Le mécanisme de FoSTeS
- Figure 24:** Le mécanisme de MMBIR
- Figure 25:** Caryotype humain masculin pour chaque paire le chromosome de gauche est en bande G et celui de droite en bande R (ICSN).
- Figure 26:** Principe de la technique FISH
- Figure 27:** Marche sur chromosome par FISH
- Figure 28:** Principe de la CGH sur chromosome

- Figure 29:** Principe a-CGH.
- Figure 30:** Principe de SNP-array.
- Figure 31:** Les différents mécanismes expliquant un effet de position.
- Figure 32:** Illustration schématique de l'appareil génital masculin.
- Figure 33:** Les différentes étapes de la spermatogenèse.
- Figure 34:** Les étapes et les principaux événements de la méiose.
- Figure 35:** Spermiogenèse.
- Figure 36:** Condensation de la chromatine.
- Figure 37:** Régulation hormonale de la spermatogenèse.
- Figure 38:** Types d'anomalies selon la classification de David modifiée
- Figure 39:** Les anomalies morphologiques des spermatozoïdes selon la classification de Kruger.
- Figure 40:** Schéma du chromosome Y indiquant la position des régions précédemment définies AZFa, AZFb et AZFc et la position des délétions récurrentes actuellement définies basés sur les répétitions palindromiques.
- Figure 41:** Aperçu biologique de la génétique de l'infertilité masculine.
- Figure 42:** Les différents mécanismes provoquant une anomalie de morphogenèse
- Figure 43:** Les trois classes de polymalformations : Séquence, Association et Syndrome.
- Figure 44:** Vue d'ensemble de la majorité des microdélétions et microduplications génomiques cités au moins deux fois dans la littérature (année 2012).
- Figure 45:** Principe du séquençage Sanger.
- Figure 46:** Les principales technologies de séquençage à haut débit.
- Figure 47:** Les étapes du WES et WGS.
- Figure 48:** Préparation de la librairie.
- Figure 49:** Les différentes étapes pour la préparation de la librairie selon le kit TruSeq PCR Free
- Figure 50:** Amplification clonale par Bridge PCR.
- Figure 51:** Séquençage par synthèse.
- Figure 52:** Séquençage Paire-end.
- Figure 53:** Schématisation de la méthode de Read-depth.
- Figure 54:** Méthode Read-pair.
- Figure 55:** Méthode Split-read.
- Figure 56:** Pipeline bio-informatique utilisé pour la détection des variants structuraux.
- Figure 57:** Pipeline bio-informatique utilisé pour la détection des SNV.
- Figure 58:** La courbe obtenue d'une amplification par qPCR montrant les trois phases : initiation, exponentielle et le plateau.
- Figure 59:** Caryotype en bande R du patient 1 : 46,XY,t(8;22)(q13;p11).
- Figure 60:** Idéogramme illustrant la translocation t(1;18)(p21;q12) chez le patient 2.
- Figure 61:** Idéogramme illustrant la translocation t(9;21)(q12;q22.3) chez le patient 3.
- Figure 62:** Images d'hybridation *in situ* en fluorescence pour les sondes : a- RP11-152C15 (BAC) (8q13.3) (rouge) et 8pter (vert) (Cytocell, Cambridge, Royaume-Uni). La sonde RP11-152C15 est située sur le dérivé (8). b- RP11-11K9 BAC (8q13.3) (rouge) et 8pter (vert) (Cytocell, Cambridge, Royaume-Uni).

- Figure 63:** Visualisation par IGV du point de cassure sur le chromosome 8 chez le patient 1.
- Figure 64:** a- Amplification des fragments de jonction des points de cassures chez le patient 1 par PCR (visualisation Labchip) pour le der (8), le der (22) et une séquence de contrôle (ATP1A3 exons 7-8). b- Séquences des fragments de jonction et coordonnées des points de cassures définies par le WGS et le séquençage de Sanger.
- Figure 65:** Images d'hybridation in situ en fluorescence pour les sondes NOR et 8 pter (vert) (Cytocell, Cambridge, Royaume-Uni) montrant la présence de régions NOR sur le dérivé (8).
- Figure 66:** Exploration de la région de part et d'autre du point de cassure sur le chromosome 8 : 70520702_70521076 sur la carte tridimensionnelle dérivés par la base de données Hi-C (résolution 50 Kb).
- Figure 67:** a- Etude de l'expression des gènes *NCOA2*, *TRAM1* et *LACTB2* dans le tissu sanguin chez le patient 1 par rapport aux témoins fertiles. b- Etude de l'expression du gène *LACTB2* dans le tissu sanguin chez le patient 1 par rapport aux témoins fertiles et infertiles.
- Figure 68:** Visualisation par IGV des points de cassures sur les chromosomes 1 et 18 chez le patient 2.
- Figure 69:** a- Amplification des fragments de jonction des points de cassures chez le patient 2 par PCR (visualisation Labchip) pour le der (1), le der (18) et une séquence de contrôle (ATP1A3 exons 7-8). b- Séquences des fragments de jonction et coordonnées de points de cassures telles que définies par le WGS et le séquençage de Sanger.
- Figure 70:** Etude de l'expression du gène *CTDPI* dans le tissu sanguin chez le patient 3 par rapport aux témoins fertiles.
- Figure 71:** Images d'hybridation in situ en fluorescence chez le patient 3 pour les sondes : a- RP1-265B9 b- RP11-113F1 (rouge) et RP11-317L14 (vert).
- Figure 72:** Visualisation par IGV du point de cassure sur les chromosomes 21 chez le patient 3
- Figure 73:** a- Amplification des fragments de jonction des points de cassures chez le patient 3 par PCR (visualisation Labchip) pour le der (9), le der (21) et une séquence de contrôle (ATP1A3 exons 7-8). b- Séquences des fragments de jonction et coordonnées de points de cassures telles que définies par le WGS et le séquençage de Sanger.
- Figure 74:** Image du patient 1
- Figure 75:** Caryogramme (en bande R) du patient 1 présentant un fragment additionnel sur le bras long du chromosome 12.
- Figure 76:** Schéma de la duplication détectée par la puce Affymetrix. Le nombre de copie de la région 12q23.1q24.13 est décalée par rapport à la ligne de base normale indiquant une duplication.
- Figure 77:** Visualisation par IGV montrant une augmentation de la profondeur de lecture sur le bras long du chromosome 12 en chr12:101499266_113290369 Mb (hg19).

- Figure 78:** Séquence de jonction de la duplication chez le patient 1 d'après l'analyse des split-read.
- Figure 79:** A- Caryogramme (en bande R) de la mère du patiente 2 présentant une translocation réciproque entre les chromosomes 16 et 18. B- Caryogramme (en bande R) de la patiente 2 présentant un dérivé 18 de la translocation maternelle.
- Figure 80:** Profil de l'ACPA illustrant la duplication sur le chromosome 16 et la délétion sur le chromosome 18 chez la patiente 2.
- Figure 81:** Image illustrant les points de cassures sur le chromosome 16 et le chromosome 18 détectés par WGS chez la patiente 4.
- Figure 82:** Séquence de jonction du dérivé 18 d'après l'analyse des split-read chez la patiente 2.
- Figure 83:** Caryogramme (en bande R) du patient 3 présentant un chromosome Y surnuméraire et une délétion au niveau du bras long du chromosome 4.
- Figure 84:** Profil de l'ACPA illustrant la délétion au niveau du chromosome 4 chez le patient 3.
- Figure 85:** Profil de l'ACPA illustrant la présence d'une copie surnuméraire du chromosome Y mais les régions PAR sont visualisée comme gain sur le chromosome X.
- Figure 86:** Visualisation par IGV montrant une diminution de la profondeur de lecture sur le bras long du chromosome 12 en chr4:77310419_97795300Mb (hg19) Mb (hg19).
- Figure 87:** Séquence de jonction de la délétion chez le patient 3 d'après l'analyse des split-read.
- Figure 88:** A- Caryogramme (bande G) partielle du fœtus 1 de l'examen direct des villosités choriales : rob(13;15)(q10;q10),+rob(13;15)(q10;q10). B- Caryogramme (bande G) partielle de la culture longue des villosités choriales : rob(13;15)(q10;q10),+13. C- Caryogramme (bande G) partielle de la culture du liquide amniotique : rob(13;15) avec anomalie de structure au niveau du dérivé du chromosome 13. D- FISH métaphasiques sur la culture des liquides amniotiques en utilisant 2 sondes 13q14.3(Rouge) et 13qter (Vert) indiquant la présence des anomalies structurales au niveau du der(13).
- Figure 89:** Schéma illustrant les gains et les pertes au niveau du chromosome 13 détectés par la puce Agilent.
- Figure 90:** Un « circos plot » illustrant les points de cassures au niveau du chromosome 13 chez le fœtus 1
- Figure 91:** Caryogramme (en bande R) de la patiente 5 présentant un chromosome 19 dicentrique avec des anomalies structurales.
- Figure 92:** Images d'hybridation *in situ* en fluorescence chez la patiente 7 A : pour les sondes : RP11-259D4 en 19q12 (rouge) et RP11-1101H19 en 19q13.11 (vert) révélant un gain de copie et RP11-317L14 (vert). B : pour les sondes : RP101-D9 en 19q11q12 (vert) et WI2-1841M21 en 19q12 (rouge) révélant une perte de copie.

- Figure 93:** Schéma illustrant les gains et la perte au niveau du chromosome 19 détecté par la puce Agilent.
- Figure 94:** Image du patient 6
- Figure 95:** Caryogramme (en bande R) du patient 6 présentant une inversion péricentrique au niveau du chromosome 8.
- Figure 96:** Schéma de la délétion détectée par la puce Affymetrix. Le nombre de copie de la région 8p11.22p12 est décalée par rapport à la ligne de base normale indiquant une délétion.
- Figure 97:** Image d'hybridation *in situ* en fluorescence chez le patient 6 pour les sondes: 8pter (Cytocell) en rouge et RP11-100B16 en vert.
- Figure 98:** Image illustrant les points de cassures sur le chromosome 8 détectés par WGS chez le patient 6.
- Figure 99:** Amplification des fragments de jonction des points de cassures chez le patient 6 par PCR (visualisation Labchip) pour les points de cassures de l'inversion der(8) et une séquence de contrôle (ATP1A3 exons 7-8).
- Figure 100:** Séquences des fragments de jonction et coordonnées de points de cassures telles que définies par le WGS et le séquençage de Sanger chez le patient 6.
- Figure 101:** Séquence du fragment de jonction de la délétion d'après split-read au niveau du bras court du chromosome 8 chez le patient 6.
- Figure 102:** Photo du patient 7.
- Figure 103:** Caryogramme (en bande R) du patient 7 présentant une inversion péricentrique au niveau du chromosome 11.
- Figure 104:** Image illustrant les points de cassures de l'inversion sur le chromosome 11 détectés par WGS chez le patient 7.
- Figure 105:** a- Amplification des fragments de jonction des points de cassures chez le patient 3 par PCR (visualisation Labchip) pour le der (11p), le der (11q) et une séquence de contrôle (ATP1A3 exons 7-8). b- Séquences des fragments de jonction et coordonnées de points de cassures telles que définies par le WGS et le séquençage de Sanger. en bleu séquence de référence sur le chromosome 11p, en rouge: séquence de référence sur le chromosome 11q, en vert les insertions.
- Figure 106:** Visualisation par IGV de la délétion c.2954_2955delCT au niveau du gène *NSDI*.
- Figure 107:** Chromatogramme du séquençage de l'exon 5 du gène *NSDI*, confirmant la délétion de deux bases CT par la méthode Sanger.
- Figure 108:** Caryogramme (en bande R) du patient 8 présentant une translocation réciproque entre les bras longs du chromosome 1 et les bras courts du chromosome 2.
- Figure 109:** Schéma des deux délétions détecté par la puce Affymetrix au niveau du bras court du chromosome 2 chez le patient 8. Le nombre de copie des régions 2p16.3 et 2p22.2p23.2 sont décalées par rapport à la ligne de base normale indiquant une délétion.

- Figure 110:** Amplification des 5 fragments de jonction des points de cassures chez le patient 8 par PCR (visualisation Labchip) pour le der(1) et le der (2) et une séquence de contrôle (ATP1A3 exons 7-8).
- Figure 111:** Séquences des 5 fragments de jonction et coordonnées de points de cassures telles que définies par le WGS et le séquençage de Sanger chez le patient 8.
- Figure 112:** Construction des chromosomes dérivés 1 et 2 d'après les données générées par le WGS chez le patient 8.
- Figure 113:** Caryogramme (en bande R) du patient 9 présentant un réarrangement complexe entre les chromosomes 1, 3, 6 et 13.
- Figure 114:** Images d'hybridation *in situ* en fluorescence chez le patient 9.
- Figure 115:** Schéma illustrant les points de cassure présent chez le patient 9 sur les idéogrammes des chromosomes 1, 3, 6 et 13 impliqués dans le réarrangement chromosomique.
- Figure 116:** Construction des chromosomes dérivés 1, 3, 6 et 13 après la combinaison des données générées par le WGS et par le caryotype et la technique FISH chez le patient 9.

Liste des tableaux :

- Tableau I :** Paramètres spermatiques définis selon les critères de l'OMS
- Tableau II :** Liste des gènes identifiés responsable d'une infertilité masculine non syndromique.
- Tableau III :** Résultats du NGS et de la caractérisation des points de cassure chez les cinq patients présentant un déséquilibre au niveau du caryotype.
- Tableau IV :** Liste des gènes impliquées dans la duplication chez le patient 1.
- Tableau V :** Liste des gènes dominants impliqués dans la délétion au niveau du bras long du chromosome 4 chez le patient 3.
- Tableau VI :** Liste des points de cassure au niveau du chromosome 13 chez le patient 4. La séquence en Noire désigne la séquence du partenaire A, en rouge celle de B
- Tableau VII :** Liste des points de cassure au niveau du chromosome 19 chez la patiente 5. La séquence en Noire désigne la séquence du partenaire A, en rouge celle de B.
- Tableau VIII :** Résultats du NGS et de la caractérisation des points de cassure chez les 4 patients présentant un RCAE au niveau du caryotype.
- Tableau IV :** Liste des gènes impliqués dans la délétion 8p11.22p12 chez le patient 6.
- Tableau X :** Liste des points de cassure chez le patient 9. La séquence en Noire désigne la séquence du partenaire A et en rouge celle de B.

Liste des abréviations :

ACPA	Analyse chromosomique sur puce à ADN
ADN	Acide désoxyribonucléique
ARN	Acide ribonucléique
CGH	Comparative Genomic Hybridization
CNV	Copy Number Variant
FISH	Fluorescent <i>in situ</i> Hybridization
FoSTeS	Fork Stalling and Template Switching
Kb	kilobase
LINE	Long Interspersed Nuclear Elements
Mb	Mégabase
MMBIR	Microhomology-Mediated Breakage Induced Repair
NAHR	Non Allelic Homologous Recombination
NGS	Next generation sequencing (séquençage à Haut-débit)
NHEJ	Non Homologous End Joining
PCR	Polymerase Chain Reaction
RCAE	Remaniement chromosomique apparemment équilibré
SINE	Short Interspersed Nuclear Elements
SNV	Single Nucleotide Variant
SV	Structural Variant
TAD	Topologically Associated Domain
WES	Whole exome sequencing
WGS	Whole genome sequencing

1. Introduction

La génétique est la science qui étudie les caractères héréditaires des individus, leur transmission au fil des générations et leurs variations. Les maladies génétiques résultent de toute modification du génome aboutissant à un processus pathologique pour l'individu.

La cytogénétique est une discipline qui a pour objet l'étude de la physiopathologie des chromosomes (condensation, recombinaison, réparation, ségrégation, transmission) et de la chromatine (organisation et rôle dans la régulation de l'expression des gènes) (1).

D'un point de vue historique, le terme chromosome a été introduit par le Biologiste allemand Wilhelm von Waldeyer-Hartz, en 1888 (2). Plus en détail, la cytogénétique a pu se développer grâce aux découvertes suivantes :

- Wilhelm Roux, en 1885, a réussi à maintenir en vie des cellules en culture (3).
- Theodor Boveri et Walter Sutton ont proposé en 1902 que les chromosomes étaient les porteurs de l'information génétique (4).
- Ross Harrison, en 1907, a obtenu les premières multiplications cellulaires (5).
- Albert Francis Blakeslee et Amos G. Avery, deux botanistes américains, ont découvert en 1937, que la colchicine permettait l'arrêt de la réplication des chromosomes au stade de métaphase (6).
- C'est en 1952 que le Dr T.C Hsu a découvert le rôle du choc hypotonique, en plaçant les cellules avant leur fixation quelques minutes dans une solution hypotonique à la place d'une solution isotonique permettant un meilleur étalement des métaphases (7).
- En 1956, Tjio et Levan ont établi à 46 le nombre de chromosomes dans l'espèce humaine (8).
- Trois ans après, l'équipe Lejeune, Gautier et Turpin a établi pour la première fois au monde, le lien entre un syndrome clinique avec déficience intellectuelle et une anomalie chromosomique : la trisomie 21. C'était la première fois qu'une corrélation entre une anomalie chromosomique et un phénotype était établie. Cette découverte a donné naissance à une nouvelle discipline : la cytogénétique médicale.

La cytogénétique médicale est une discipline qui a pour objectifs la détection d'anomalies chromosomiques grâce à des techniques microscopiques ou moléculaires afin d'établir un diagnostic biologique et d'assurer un conseil génétique.

Par la suite de nombreuses découvertes concernant la structure et la transmission des chromosomes ont été mises en évidence grâce au développement des techniques de banding : bandes Q (9), bandes C (10), bandes G puis bandes R (11) et des bandes T (12) et de techniques de plus en plus résolutive (13). Le caryotype permet une analyse pangénomique et la détection des anomalies de nombre ou de structure équilibrées ou déséquilibrées, homogènes ou en mosaïque, mais avec une résolution limitée comprises entre 5 à 10 Mb.

En complément du caryotype, la technique d'hybridation *in situ* en fluorescence (FISH) s'est développée pour donner naissance à un nouveau domaine : la cytogénétique moléculaire, basée sur les travaux de Gall et Pardue qui utilisaient des radioisotopes pour la localisation de séquences uniques (14). Ensuite ces sondes radioactives ont été remplacées par des vecteurs, Yeast Artificial Chromosome (YAC), cosmides, Bacterial Artificial Chromosome (BAC) et des oligonucléotides marqués par un fluorochrome. Cette technique ciblée, plus résolutive que le caryotype, permet une visualisation métaphasique ou interphasique des sondes marquées. Elle peut détecter des microdélétions ou microduplications non décelables par le caryotype, et permet aussi de localiser les points de cassures des remaniements équilibrés. Plus tard, la technologie des puces à ADN (ACPA, Analyse chromosomique sur Puce à ADN) ou CGH-array a été mise en place. Cet outil de cytogénétique moléculaire permet une analyse pangénomique (15). Elle repose sur l'hybridation d'oligonucléotides fixés sur un support solide avec l'ADN génomique du patient et d'un témoin marqué par deux fluorochromes différents. Cette technique met en évidence le nombre de copies (CNV, Copy Number Variation) de chaque locus étudié dans un échantillon par rapport à un témoin. Malgré sa résolution de l'ordre de 10 kb, cette technique n'informe pas sur la structure des chromosomes et la mécanique chromosomique de même qu'elle ne détecte pas les anomalies de structures équilibrées. Plus récemment, la technologie de séquençage massif en parallèle, notamment le séquençage génome entier (WGS), s'est révélée une technique efficace pour caractériser les remaniements chromosomiques au niveau moléculaire, à la base près (16). Outre leur apport diagnostique, l'application de ces technologies haut-débit à l'étude des remaniements chromosomiques a permis de découvrir de nouveaux gènes, de mieux définir les corrélations génotype phénotype, mais aussi de découvrir de nouveaux types de remaniement comme les chromoanagenesis (17).

L'objectif principal de ma thèse est de caractériser des remaniements chromosomiques sur le plan moléculaire afin de mieux comprendre les mécanismes ainsi que les conséquences de ces remaniements et d'identifier de nouveaux gènes. Dans une première partie, nous nous intéresserons à l'apport de l'étude génome entier chez trois patients présentant des RCAE et

atteints de trouble de la spermatogenèse. Dans une deuxième partie, nous réaliserons une étude génome entier chez neuf patients atteints d'anomalies congénitales multiples. Cinq sont porteurs d'un remaniement chromosomique équilibré et quatre d'un remaniement chromosomique déséquilibré.

1.1. Anomalies chromosomiques

Les anomalies chromosomiques constitutionnelles surviennent soit dès la conception aboutissant alors à une aberration homogène touchant toutes les cellules de l'organisme, soit au cours des premières divisions du zygote se traduisant par une anomalie en mosaïque. Les anomalies chromosomiques acquises correspondent aux anomalies chromosomiques apparues au cours de la vie et ne touchant qu'un seul organe. Ces anomalies sont le plus souvent en rapport avec un processus de cancérisation, cette thématique ne sera pas abordée dans ce manuscrit.

On distingue deux types d'anomalies chromosomiques : les anomalies de nombre et les anomalies de structure.

1.1.1. Les différents types d'anomalies chromosomiques

1.1.1.1. Les anomalies de nombres

Les anomalies de nombres sont divisées en deux catégories : aneuploïdie et polyploïdie.

Les aneuploïdies :

La déviation du nombre normal des chromosomes dans une (ou plusieurs) paire(s) de chromosome(s) est appelée aneuploïdie. Elle est caractérisée par la perte d'un chromosome (monosomie) ou la présence d'un chromosome en plus (trisomie) du jeu chromosomique diploïde normal (46 chromosomes : 22 paires autosomiques et une paire gonosomique). Ces anomalies sont dues le plus souvent à un défaut de ségrégation (non disjonction) (Figure 1). Une non-disjonction est définie par le fait que deux chromosomes homologues ou deux chromatides sœurs migrent vers le même pôle lors de l'anaphase et donc dans la même cellule fille, au lieu de migrer chacun dans une cellule fille. Ce défaut de ségrégation peut se produire lors d'une division méiotique maternelle ou paternelle. Dans ce cas l'aneuploïdie sera présente dans tout l'organisme. On parle d'anomalie homogène. Plus rarement, ces anomalies peuvent être dues à une erreur mitotique postzygotique conduisant à une aberration uniquement dans

recolllements anormaux. Elles peuvent affecter un ou deux chromosomes, homologues ou non homologues; elles peuvent être équilibrées (pas de matériel chromosomique perdu ou gagné), ou déséquilibrées (avec matériel chromosomique ajouté ou perdu). Le mécanisme de survenue de ces anomalies sera détaillé dans le paragraphe suivant.

1.1.1.2.1. *Les anomalies chromosomiques de structure impliquant un seul chromosome*

Délétion :

Elle résulte d'une cassure chromosomique avec une perte de l'extrémité distale d'un bras chromosomique : délétion terminale. Une délétion interstitielle survient à la suite de deux cassures avec perte du matériel chromosomique situé entre ces deux cassures (Figure 2).

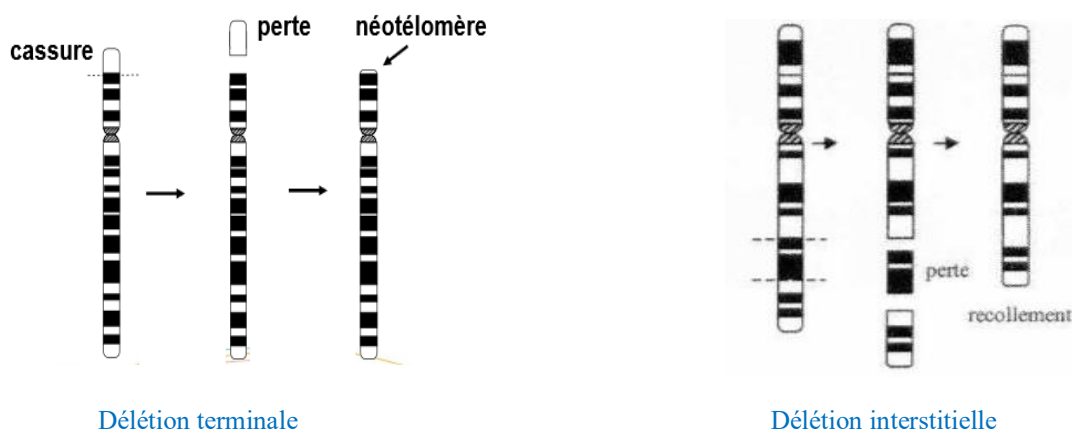
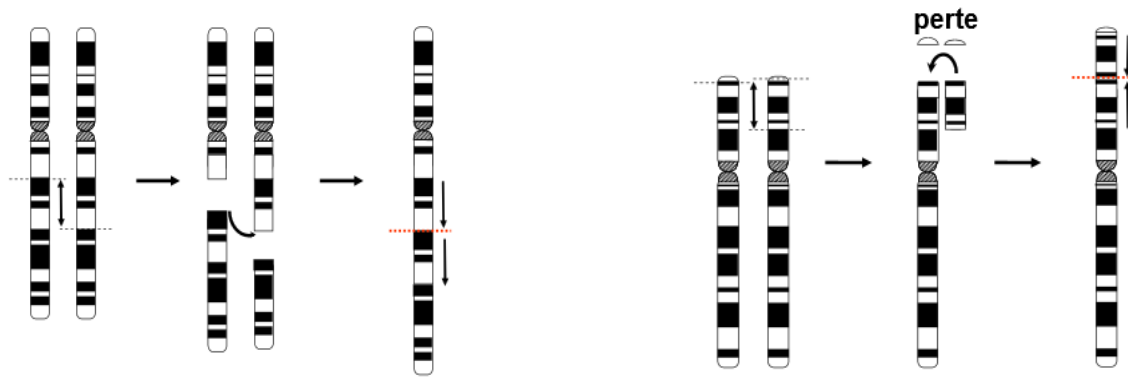


Figure 2 : Mécanisme de formation d'une délétion terminale et d'une délétion interstitielle (19).

Duplication :

Il s'agit de la répétition d'un fragment chromosomique de taille variable à sa position d'origine, entraînant un gain de matériel génétique. En règle générale, cette aberration est observée suite à la formation d'une boucle avant la phase S du cycle cellulaire. Elle peut se produire soit en tandem (le fragment dupliqué est dans le même sens : duplication directe), soit en miroir (le fragment dupliqué est dans le sens inverse : duplication inversée) (Figure 3).



Duplication en tandem

Duplication inversée

Figure 3 : Mécanisme de formation d'une duplication en tandem ou inversée (19).

Chromosome en anneau :

Les anneaux sont des chromosomes anormalement circulaires. Ils surviennent après deux cassures dans un même chromosome, une sur chaque bras chromosomique, avec perte des segments terminaux et fusion des extrémités cassées portant le centromère. Les segments distaux sont perdus, par conséquent un chromosome en anneau est déséquilibré (Figure 4).

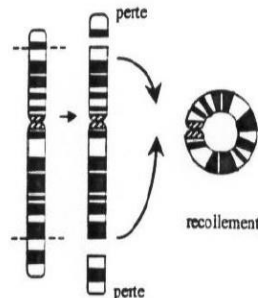


Figure 4 : Mécanisme de formation d'un chromosome en anneau (19).

Inversion :

Elles sont liées à la survenue de deux cassures chromosomiques sur le même chromosome avec retournement de 180° du fragment entre les deux cassures. Elles résultent de la formation d'une boucle dans l'un des chromosomes homologues durant leur appariement en prophase I. Elles peuvent être paracentriques si les cassures sont produites dans le même bras, ou péricentriques si le centromère est inclus dans le segment intermédiaire des points de cassures (Figure 25).

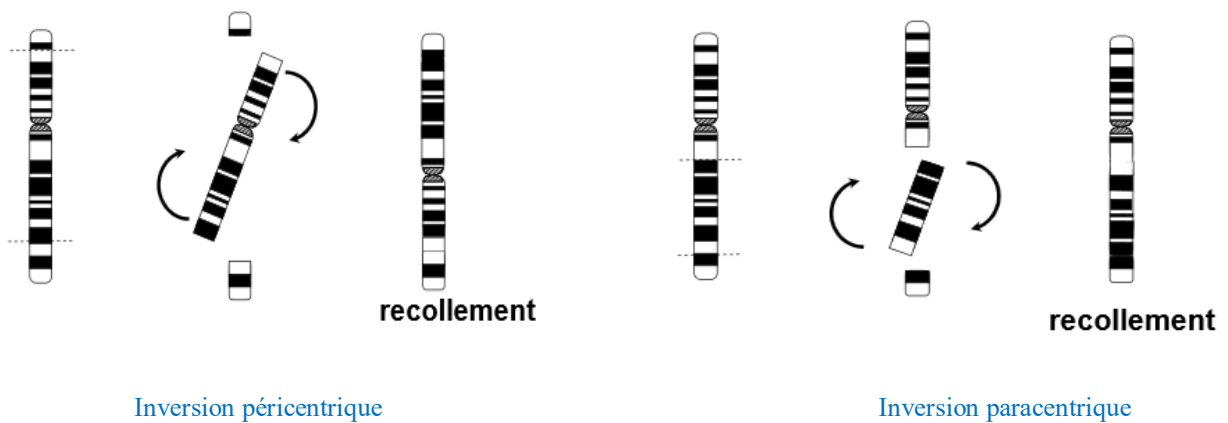


Figure 5: Mécanisme de formation d’une inversion péricentrique et paracentrique (19).

Lors de la méiose, durant la prophase I, l’appariement des chromosomes homologues dans la région de l’inversion nécessite la formation d’une boucle. En l’absence de crossing-over dans cette boucle, deux chromosomes sont produits, un normal et l’autre inversé. Si une recombinaison a lieu dans la boucle, il y a formation de gamètes déséquilibrés. Dans le cas d’une inversion péricentrique des duplications et des délétions des fragments terminaux sont générés (duplication-déficience) (Figure 6).

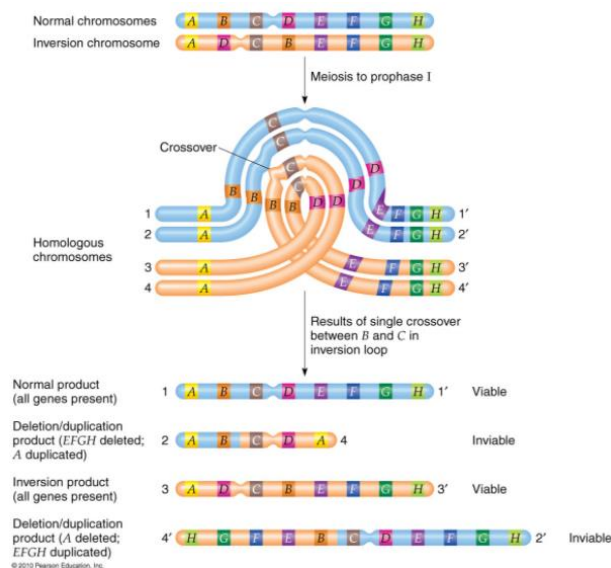


Figure 6 : Etude de la ségrégation de l’inversion péricentrique
 (https://www.mun.ca/biology/scarr/iGen3_16-09.html)

Dans le cas d'une inversion paracentrique, des chromosomes acentriques (qui seront perdus) et dicentriques sont produits (Figure 7).

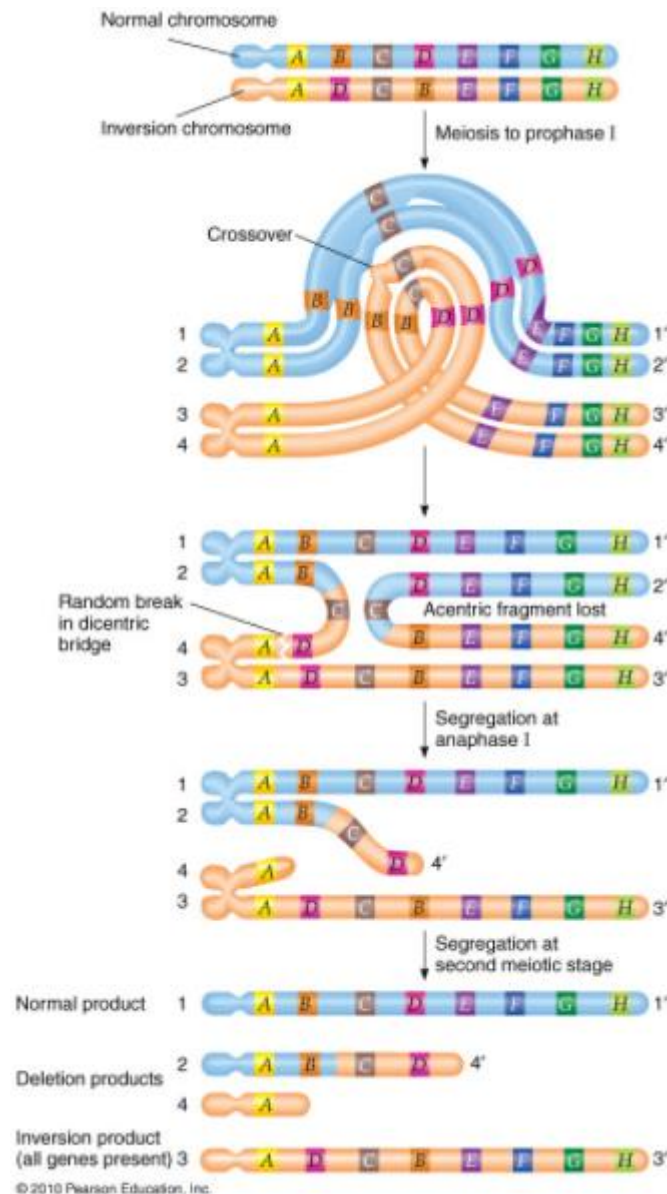


Figure 7 : Etude de la ségrégation de l'inversion paracentrique d'un chromosome acrocentrique (https://www.mun.ca/biology/scarr/iGen3_16-08.html).

Isochromosome :

Un isochromosome se forme lorsqu'un chromosome normal se divise transversalement et non longitudinalement de sorte qu'il sera composé de deux bras longs ou de deux bras courts d'un avec perte de l'autre bras. Chaque bras est composé de gènes identiques dans le même ordre en partant du centromère. Il peut être monocentrique ou dicentrique selon le mécanisme de formation (Figure 8).

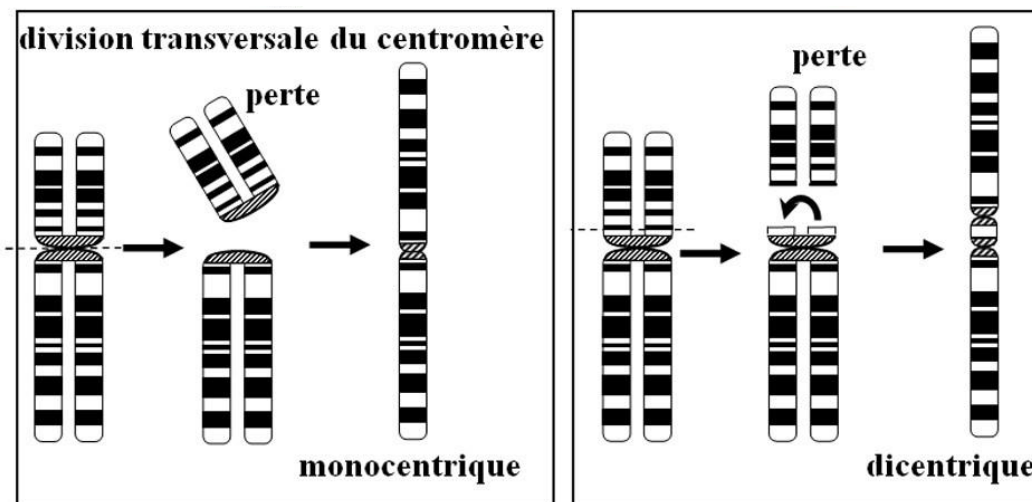


Figure 8 : Mécanisme de formation d'un isochromosome monocentrique ou dicentrique (19).

Marqueur :

Il s'agit de petit fragment chromosomique surnuméraire dont la taille est inférieure à celle d'un chromosome 20 de la même mitose et dont il n'est pas possible de déterminer l'origine avec les techniques de cytogénétique conventionnelles seules. Il existe donc un gain de matériel chromosomique générant une trisomie partielle pour une région du génome. Leurs conséquences phénotypiques dépendent de leur contenu génique. Ces marqueurs perturbent aussi la ségrégation méiotique (20).

1.1.1.2.2. Les anomalies chromosomiques de structure impliquant plusieurs chromosomes

Insertion :

Les insertions se caractérisent par le transfert d'un segment d'un chromosome donneur, généré par deux points de cassure, dans un chromosome accepteur s'intercalant au niveau d'un nouveau point de cassure (Figure 9). Le segment inséré peut avoir gardé son orientation ou prendre une orientation inversée. Ainsi, trois cassures sont nécessaires pour qu'une insertion se produise.

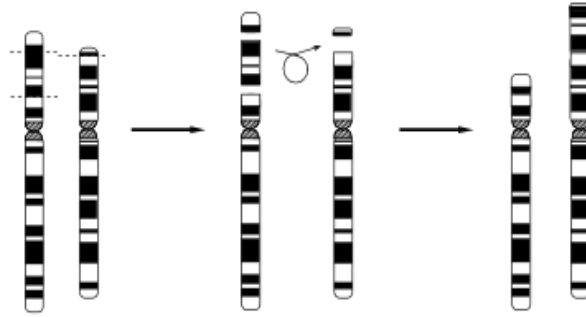


Figure 9 : Mécanisme de formation d'une insertion (19).

Translocation :

Elle se caractérise par deux cassures sur deux chromosomes différents, le plus souvent non homologues, et recollement après échange des segments distaux. On distingue les translocations robertsoniennes et les translocations réciproques.

Translocations robertsoniennes :

Elles se produisent uniquement au niveau des chromosomes acrocentriques (13,14,15,21 et 22) et entraînent la perte des deux bras courts et une réunion des deux bras longs pour former un chromosome dérivé (Figure 10). Les bras courts des chromosomes acrocentriques sont très petits et constitués d'hétérochromatine et des régions organisateurs des nucléoles, qui sont présentes en plusieurs exemplaires dans le génome et dont la perte n'a pas de conséquence clinique (21). Malgré la perte des bras courts, la quasi-totalité de l'information génétique est présente et, cette anomalie est considérée comme équilibrée avec un nombre de chromosome à 45.

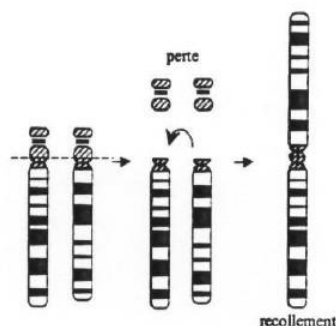


Figure 10 : Mécanisme de formation d'une translocation robertsonienne (19).

La translocation robertsonienne a comme conséquence sur la ségrégation des chromosomes, la formation d'un trivalent entre le chromosome dérivé et les deux chromosomes acrocentriques normaux de même paire pendant la méiose I, au stade pachytène.

Normalement la formation de ce trivalent se produit en configuration *cis* ou *trans*, mais en pratique seule la configuration *cis* est observée. Il y aura génération de 3 modes de ségrégation méiotique : le mode alterne, le mode adjacent et le mode 3:0. Dans le mode alterne, le chromosome dérivé migre dans un gamète et les deux chromosomes acrocentriques normaux migrent dans un autre, produisant respectivement des gamètes équilibrés et normaux. Dans le mode adjacent, quatre combinaisons sont possibles. Le chromosome dérivé migre dans un gamète avec un chromosome acrocentrique normal, et le deuxième chromosome acrocentrique migre dans un autre. Deux types de gamètes sont obtenus avec 23 ou 22 chromosomes (Figure 11).

Dans le mode de ségrégation 3:0, il y a production d'un gamète avec 21 chromosomes et d'un gamète avec 24 chromosomes. Ce mode de ségrégation est très rare. Ces deux derniers modes (adjacent et 3:0) ne produisent que des gamètes déséquilibrés.

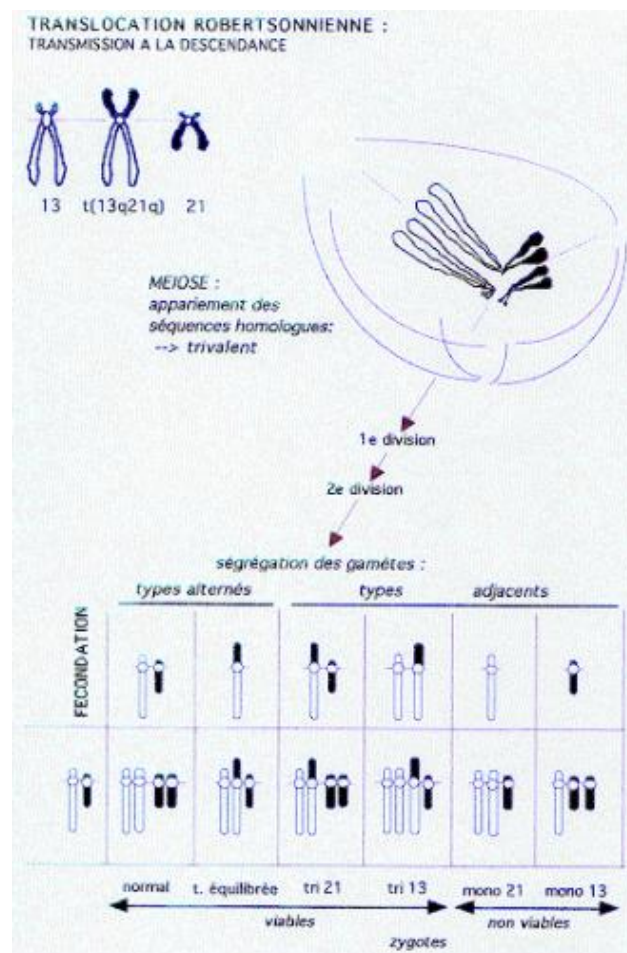


Figure 11 : Etude de la ségrégation d'une translocation robertsonienne (18).

Translocations réciproques :

Cette anomalie consiste en l'échange de matériel chromosomique entre deux chromosomes non homologues sans perte ni gain de matériel, elle est alors dite équilibrée (Figure 12). Elles peuvent affecter tous les chromosomes.

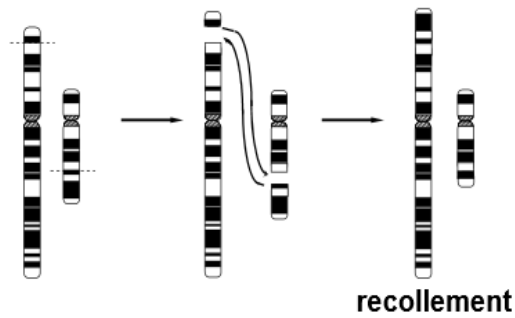


Figure 12: Mécanisme de formation d'une translocation réciproque (19).

Toutefois, les porteurs d'une translocation réciproque peuvent former des gamètes non équilibrés. Durant la méiose I, les chromosomes impliqués dans la translocation s'apparient, et forment une configuration caractéristique, un quadrivalent, appariement en croix des quatre chromosomes impliqués.

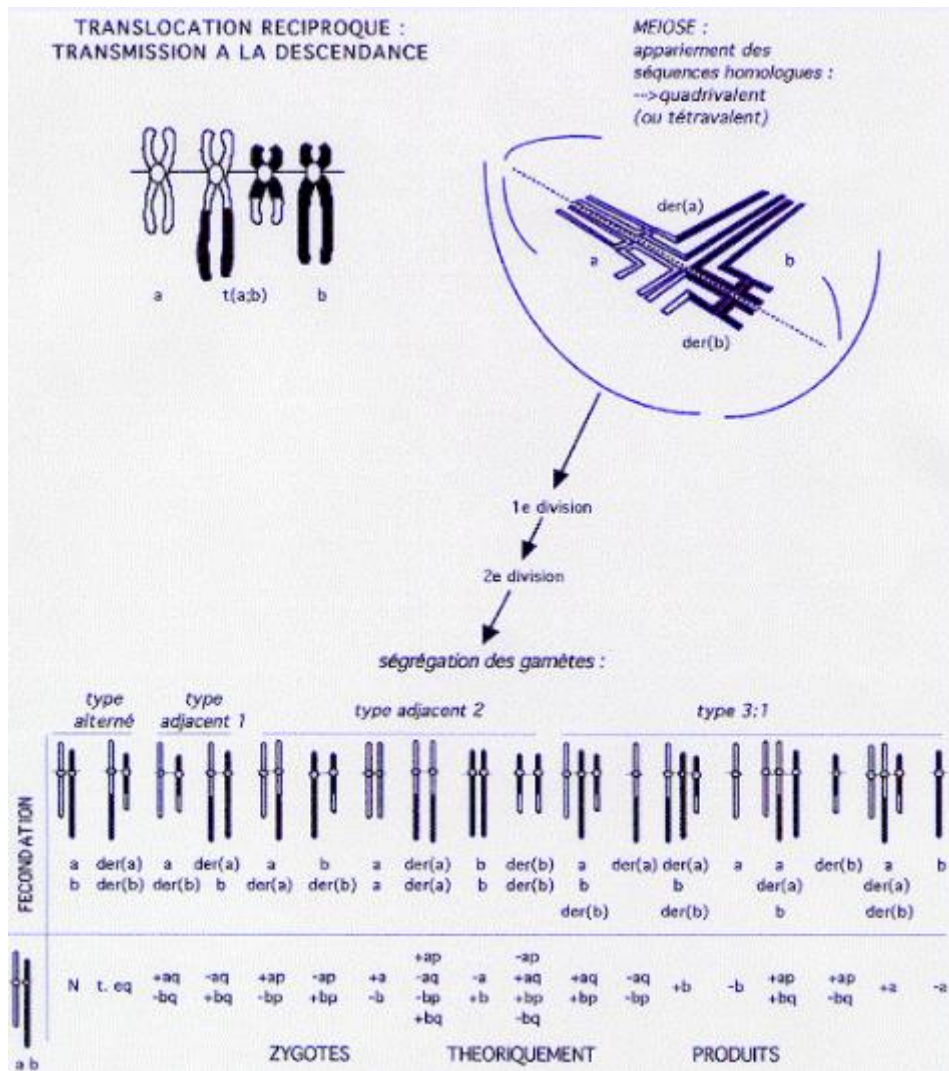


Figure 13 : Etude de la ségrégation d'une translocation réciproque (18).

Lorsque ces quatre chromosomes se séparent durant l'anaphase de la méiose, cinq possibilités se présentent (Figure 13) :

- Avec la ségrégation alterne, un gamète reçoit deux chromosomes normaux et dans un autre gamète, les deux chromosomes dérivés ou transloqués. Les gamètes obtenus sont normaux ou équilibrés (23 chromosomes).
- Avec la ségrégation non alterne mode adjacent I, les deux chromosomes sur la gauche du quadrivalent vont dans un gamète et les deux chromosomes sur la droite dans un autre. Les gamètes formés ont tous 23 chromosomes, mais ils ont soit une disomie partielle d'un chromosome associé à une nullisomie partielle de l'autre chromosome impliqué dans la translocation soit l'inverse.

- Avec la ségrégation non alterne mode adjacent II, les chromosomes supérieurs du quadrivalent vont dans un gamète, et les deux inférieurs dans l'autre. Les gamètes formés sont déséquilibrés au niveau du nombre de chromosomes avec 23 chromosomes. Ils ont une disomie partielle d'un des chromosomes impliqués dans la translocation avec une nullisomie partielle de l'autre.
- Avec le mode de ségrégation 3:1, il y aura la formation de gamètes constitués de 3 des 4 chromosomes formant le quadrivalent et de gamètes composés d'un seul des 4 chromosomes du quadrivalent. Ainsi, tous les gamètes formés sont déséquilibrés à 22 ou 24 chromosomes.
- Enfin, le mode de ségrégation 4:0, entraînera la formation de gamètes avec les 4 chromosomes constituant le quadrivalent ou bien aucun de ceux-ci. Tous, les gamètes sont déséquilibrés avec 21 ou 25 chromosomes.

1.1.1.3. Remaniement chromosomique complexe

Les réarrangements chromosomiques complexes sont définis historiquement par l'implication d'au moins 3 points de cassures avec un échange de matériel chromosomique entre au moins 2 chromosomes (22). Le réarrangement complexe a été décrit pour la première fois en 1970, en décrivant un réarrangement complexe familial impliquant trois chromosomes : 5, 9 et 13 (23). En 1988, Kausch et al. (24) suggère une classification structurale des réarrangements complexes répartis en trois groupes (Figure 14) :

- Type 1, "The three-way", ce type correspond à l'implication de trois points de cassure sur trois chromosomes, puis translocation et recollement de ces segments.
- Type 2, "The complex chromosome rearrangement", exceptionnels, ce type correspond à la présence dans un chromosome du réarrangement de plus qu'un point de cassure.
- Type 3, "The double two-way", ce type représente deux translocations réciproques indépendantes.

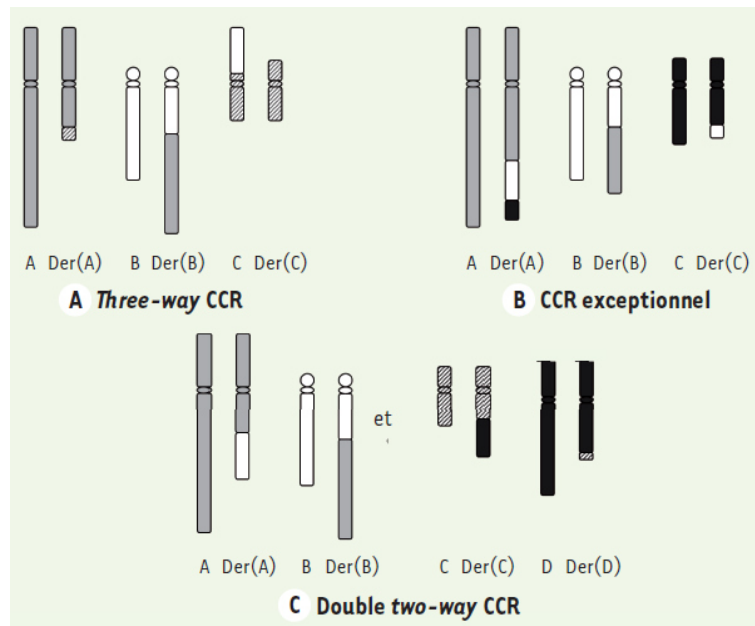


Figure 14 : Les trois types de réarrangements chromosomiques complexes (25).

Ces réarrangements sont rares et peuvent être familiaux ou bien *de novo*, équilibrés ou déséquilibrés, avec un phénotype normal ou anormal. Plus il y a de points de cassure, plus le risque d'un phénotype anormal est important. En 2011, le terme chromothripsis du grec *chromos* pour « chromosome » et *thrypsis* pour « pulvérisation » ou « éclatement en morceaux multiples » a été utilisé pour la première fois pour désigner un réarrangement chromosomique très complexe, chez un patient atteint d'une leucémie lymphoïde chronique avec 42 points de cassures sur le bras long du chromosome 4 (26). Ces remaniements ont été par la suite décrits en pathologie constitutionnelle et ont remis en question la définition de remaniement chromosomique complexe, qui sont à présent définis comme un remaniement dont il n'est pas possible de faire le conseil génétique (ISCN 2016).

Au début de l'identification du chromothripsis, une hypothèse émise pour sa génération était celle d'une pulvérisation d'une région d'un chromosome ou de plusieurs chromosomes suivie d'une reconstruction par simple recollement des fragments générés par le mécanisme NHEJ ou MMEJ (microhomology-mediated end joining,) et perte de certains d'entre eux (27) (Figure 15).

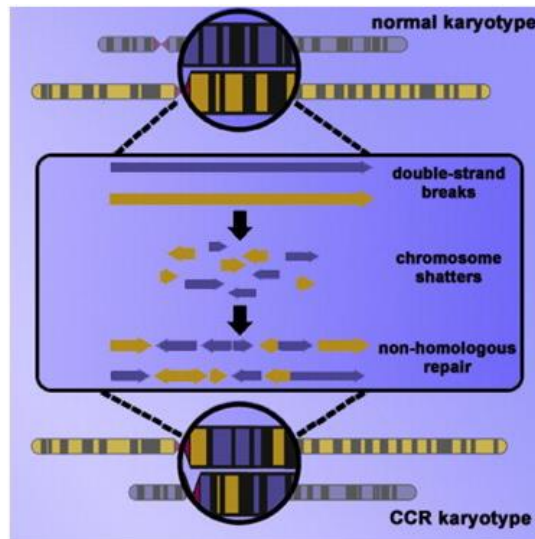


Figure 15 : Mécanisme du chromothripsis (28).

Un deuxième type de remaniements chromosomiques a été décrit, le « chromoanasythesis ». Les mécanismes de réparation des cassures sont différents de ceux du chromothripsis. Ce type de remaniements est basé sur le défaut de la machinerie de la réplication de l'ADN, et sur les deux mécanismes de réparation des cassures FoSTeS et MMBIR (Figure 16). Le « chromoanasythesis » est caractérisé par la présence de pertes et de gains de fragments chromosomiques sur un seul chromosome ou sur plusieurs (29). Ces réarrangements ont été décrits dans les anomalies acquises et constitutionnelles.

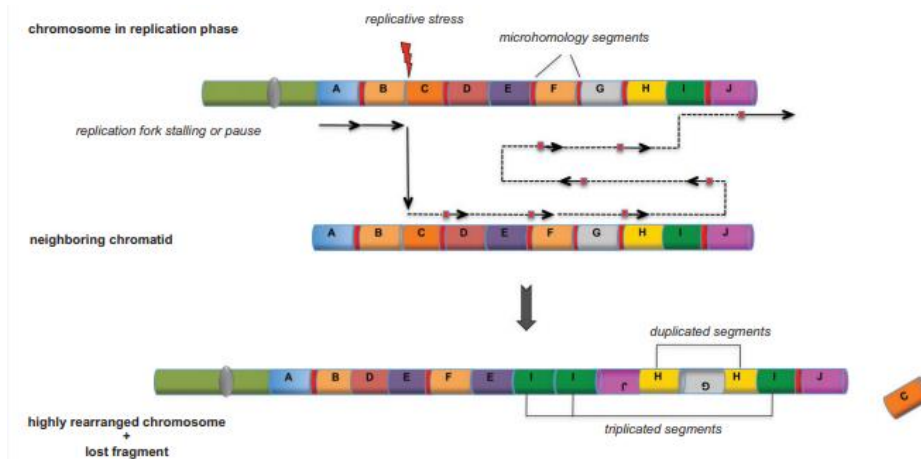


Figure 16 : Principe du processus de chromoanasythesis. Une fourche de réplication retardée va être insérée dans une deuxième fourche de réplication, grâce à la présence d'une microhomologie, par FoSTeS ou MMBIR. Le processus conduit à la formation de réarrangement complexe impliquant des duplications et des triplications (30).

Un troisième type de remaniements complexes, chromoplexie, a été décrit pour la première fois en 2013, dans le génome des cellules du cancer de la prostate (31). Ce type de remaniement suppose qu'un événement génère plusieurs points de cassure sur plusieurs brins d'ADN impliquant plusieurs chromosomes suivis d'une ligation les uns aux autres avec élimination de certains fragments (Figure 17). La chromoplexie provoque la formation de gènes de fusion comme dans le cancer de prostate (*TMPRSS2-ERG*) (31) ou dans le cas de sarcome d'Ewing (*EWS-ETS*) (32), ou bien l'inactivation des gènes suppresseurs de tumeur par exemple *PTEN* dans le cas du cancer de la prostate (31). Ce type de remaniements a été décrit dans plusieurs types de cancer, mais jusqu'à présent aucun cas dans les anomalies constitutionnelles.

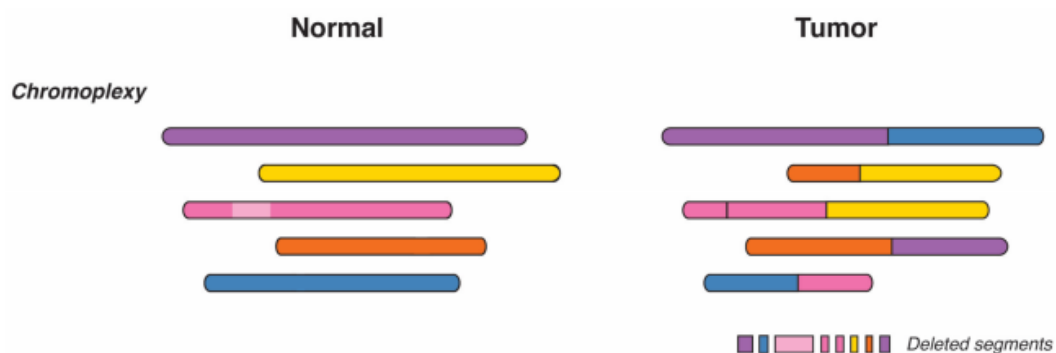


Figure 17 : Mécanisme de chromoplexie (33).

Ces trois types de remaniements chromosomiques complexes : chromothripsis, chromoanasythesis et chromoplexie, sont regroupés sous le terme de chromoanagenesis. Plusieurs facteurs conduisent à la pulvérisation des chromosomes, des causes exogènes comme les radiations ionisantes ou des drogues géno-toxiques. Des infections virales peuvent aussi provoquer des cassures multiples des chromosomes (34). Des causes endogènes, telles que l'altération du gène *TP53* engendre également ce type de remaniement. Chez une patiente atteinte du syndrome de Li-Fraumeni, un lien entre la présence de chromothripsis et mutation dans le gène *TP53* a été révélé. Une étude plus élargie a montré un lien entre la protéine P53 et ce type de réarrangement chez 11 des 13 patients présentant un chromothripsis. Cela peut être illustré par le rôle de *TP53* dans le contrôle de l'apoptose, ainsi une altération de *TP53* peut également faciliter la survie des cellules après une catastrophe comme le cas de chromothripsis. Le gène *TP53* joue un rôle dans le contrôle de la transition des cellules au niveau de la phase G2/M du cycle cellulaire ainsi les altérations de *TP53* dans le contrôle du cycle cellulaire contribuant à un compactage prématuré des chromosomes (35,36). La mort cellulaire

programmée (apoptose) pourrait être également responsable de ces réarrangements complexes. Les cellules de notre organisme sont soumises à divers stress qui peuvent induire une apoptose initiant la fragmentation de la chromatine. Certaines cellules peuvent ne pas terminer l'apoptose et ainsi auraient besoin de réparer l'ADN clivé conduisant à des réarrangements complexes (37). Cette hypothèse est appelée « apoptose abortive », qui expliquerait surtout que ces remaniements complexes toucheraient surtout les télomères (Figure 18), ce qui n'est pas le cas général de tous les chromothripsis.

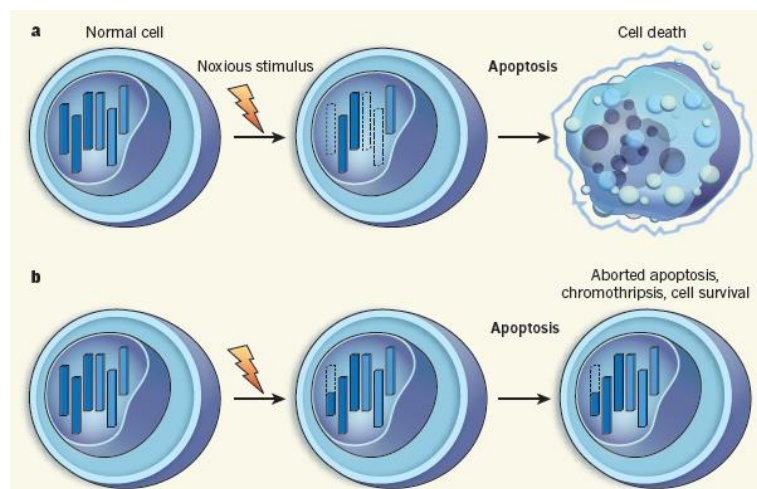


Figure 18 : Apoptose abortive. (a) Suite à un stress, la cellule rentre dans la mort cellulaire programmée. (b) Certaines cellules en apoptose l'interrompent et survivent en présentant des remaniements complexes (37).

Une autre hypothèse est « le modèle de micronoyau ». La présence de micronoyaux est une caractéristique de certaines cellules cancéreuses en culture (38). Même dans certains syndromes d'instabilité chromosomique, par exemple le syndrome de Bloom et l'anémie de FANCONI, une accumulation des micronoyaux dans les cellules est observée (39). La présence de micronoyau est également fréquente dans les cellules embryonnaires provoquant un développement embryonnaire anormal, qui est devenu une application clinique et un des critères de sélection des embryons pendant la fécondation *in vitro* (40). L'apparition d'un micronoyau résulte d'un défaut du fuseau achromatique à capter un centromère d'un chromosome (41). À l'intérieur de ce micronoyau, un ou plusieurs chromosomes subissent une anomalie de réparation des cassures, puisque ce micronoyau ne possède pas toutes les caractéristiques d'un noyau normal (42) ou peut présenter un défaut de la protéine p53. Ensuite

ce micronoyau contenant un ou plusieurs chromosomes remaniés est réincorporé dans le noyau principal de la cellule (38). En se basant sur cette hypothèse, Chatron et al. (43) suggère un mécanisme de formation de ces remaniements. Dans le micronoyau, une partie d'un chromosome se trouve dans un stade où la réplication de la chromatine est tardive tandis que les autres chromosomes vont rentrer dans la division cellulaire (Figure 19). A ce stade, un défaut du système de réparation de cassures engendre des chromoanasythesis (FoSTeS et MMBIR), ou des chromothripsis (NHEJ).

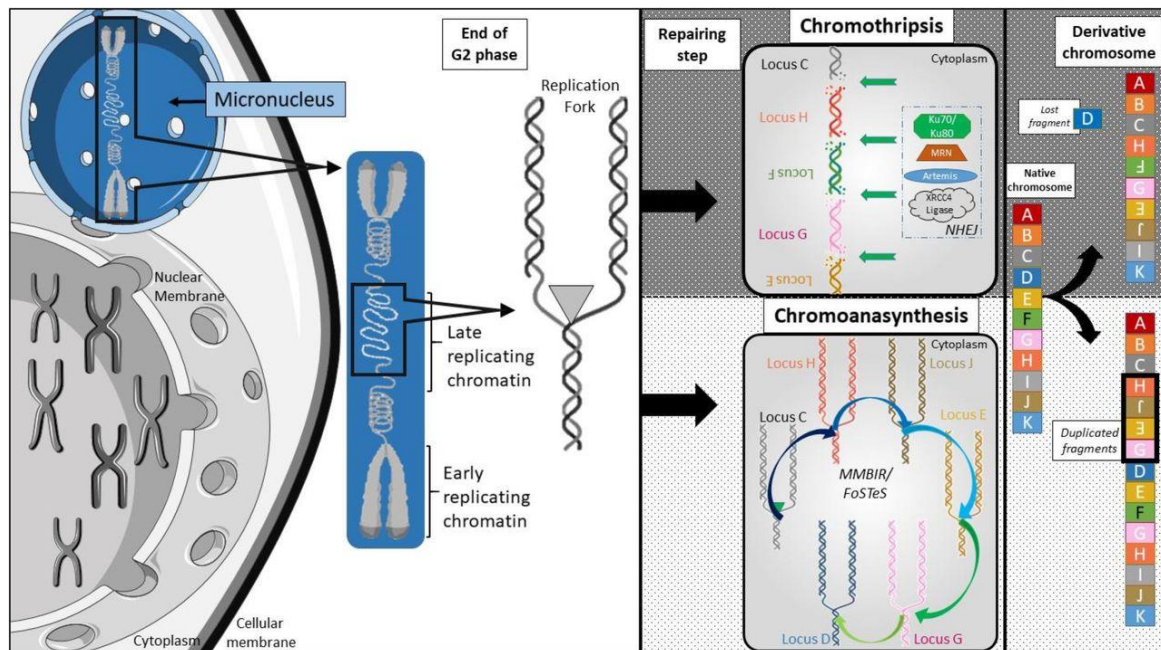


Figure 19: ce schéma illustre l'hypothèse de la formation des chromoanagenisis et chromothripsis dans le micronoyau (43).

1.1.1.4. Autres types d'anomalies chromosomiques

Instabilité chromosomique :

Les syndromes d'instabilité chromosomique regroupent un certain nombre de maladies et de syndromes caractérisés par une instabilité chromosomique. Cette instabilité est due à une altération de gènes impliqués dans la réparation de l'ADN et la stabilité du génome. Dans certains de ces syndromes, l'instabilité peut être visualisée sur un caryotype. L'anémie de FANCONI est caractérisée soit par un taux élevé de cassures chromosomiques ou chromatidiennes, soit par l'échange entre les chromatides sœurs révélant des images atypiques secondaires telles que des figures tri-ou tétra- radiales (44). Dans le syndrome de Bloom une augmentation du taux d'échange entre les chromatides sœurs peut-être visualisé après

traitement spécifique (incorporation dans l'ADN d'un analogue de base, le 5-bromodéoxyuridine (BrdU), en combinaison avec un colorant de l'ADN) (39).

Sites fragiles :

Les sites fragiles ont été définis à l'origine comme des lacunes non colorées visibles sur les deux chromatides d'un chromosome métaphasique. Les premiers décrits étaient le siège de cassures induites par l'emploi de différents traitements dans les cultures cellulaires *in vitro*. On observe chez certains individus une zone de fragilité constitutionnelle localisée sur un autosome donné. La présence/absence de certains de ces sites sur des homologues a été considérée comme ségréguant de façon mendélienne. Les autosomes et les sites les plus souvent observés sont : 2q11, 10q23, 10q25, 11q13, 16p12, 16q22, 17p12, 20p11. Le cas du chromosome X est particulier avec l'existence d'un site fragile en Xq27.3. Ce site fragile a été observé chez des garçons avec une déficience intellectuelle syndromique, d'où le nom du syndrome de l'X fragile. La transmission est celle d'une pathologie dominante liée à l'X (19).

Disomie uniparentale :

Ce phénomène est caractérisé par la présence chez un individu de deux chromosomes d'une même paire provenant d'un seul de ses parents. Il peut survenir dans le cadre d'une correction de monosomie ou de trisomie. Bien que la forme diploïde de l'allèle soit respectée, ce phénomène explique certaines pathologies. C'est un mode de transmission inhabituelle des maladies génétiques.

L'empreinte parentale, qui est un évènement physiologique, correspond à un mécanisme épigénétique qui entraîne une répression d'expression génique : expression d'une seule copie parentale d'un gène (soit la copie maternelle soit la copie paternelle). Dans le cas d'une disomie uniparentale, l'empreinte parentale est apposée sur les deux chromosomes de la même manière entraînant soit une perte d'expression des gènes ou à l'inverse, une surexpression génique. Ce déséquilibre d'expression de certains gènes est responsable d'un phénotype variable en fonction du chromosome concerné.

De plus, ce phénomène peut être l'origine de l'expression d'une maladie récessive. En effet, si un parent est porteur d'une mutation à l'état hétérozygote, celle-ci peut être transmise à l'état homozygote dans les cas d'une isodisomie uniparentale. Ce cas de figure a été décrit pour un patient atteint de la mucoviscidose du fait d'une microdélétion identifiée au niveau du chromosome 7 à l'état homozygote, héritée uniquement de la mère (45).

1.2. Mécanismes de formation des anomalies chromosomiques de structure

Les anomalies de structure surviennent suite à des cassures et leur réparation anormale. Ces réarrangements s'expliquent par différents mécanismes qui dépendent de l'architecture locale du génome. On peut différencier les réarrangements récurrents dont les points de cassures sont fixes, et surviennent dans des régions présentant une forte homologie, et les réarrangements non récurrents pour lesquels les points de cassures sont variables et surviennent à n'importe quel endroit du génome, et qui sont de taille variable avec des bordures distinctes. (46).

1.2.1. Les réarrangements récurrents

Dans la majorité des cas, la recombinaison homologe non allélique (NAHR) constitue le mécanisme le plus fréquent des réarrangements récurrents. Elle est liée à des séquences répétées appelées duplications segmentaires ou LCR (Low Copy Repeat), qui sont définies par des blocs d'ADN, de taille supérieure à 1 kb, situés en plusieurs localisations génomiques et présentant plus de 90% d'identité (Figure 20). Ces LCR représentent environ 10% du génome humain (47).

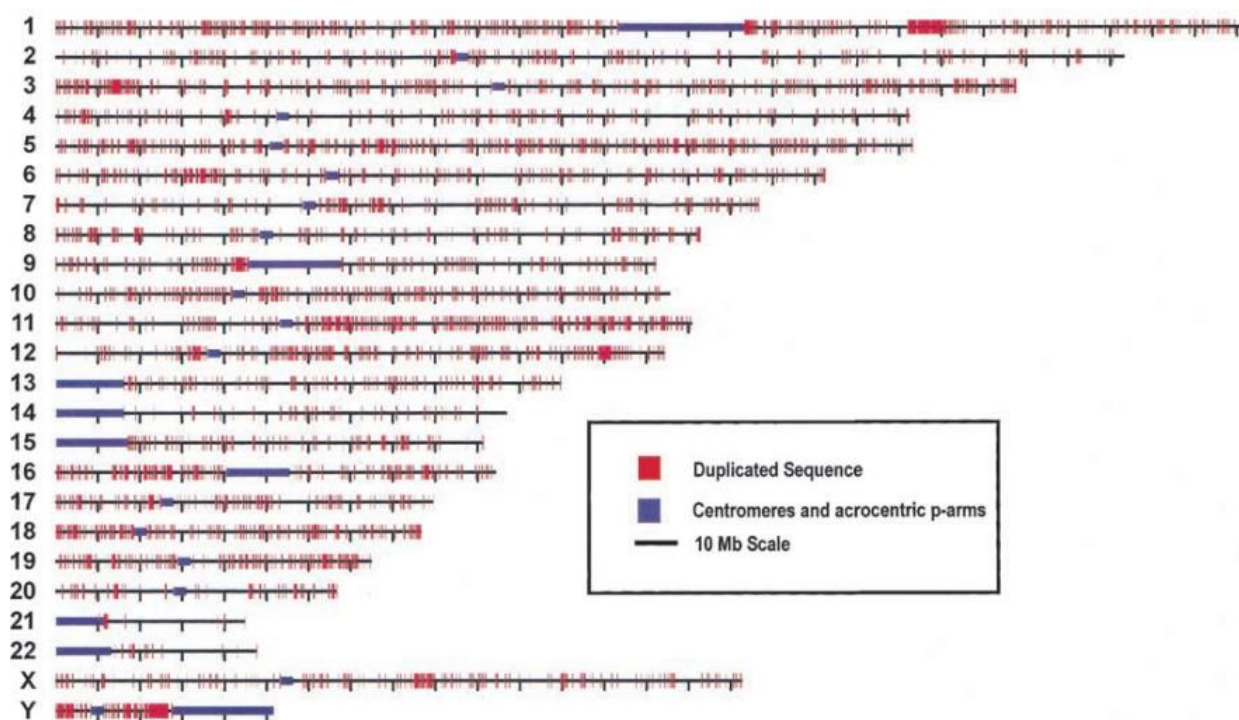


Figure 20 : Vue à l'échelle du génome des duplications segmentaires. *Les positions des alignements sont représentées en rouge pour chaque chromosome. Les barres violettes représentent les centromères ainsi que les bras courts des chromosomes acrocentriques. En raison de contraintes d'échelle, seules les duplications segmentaires qui représentent 98 à 100% d'homologie et celles qui ont une taille supérieure à 5kKb sont représentées (47).*

Au cours d'une méiose, une recombinaison homologue non allélique peut survenir du fait d'un mauvais alignement entre deux LCR. L'alignement s'est produit du fait de la très forte homologie de séquence, mais pour des régions génomiques non alléliques. Ces recombinaisons peuvent se produire au sein d'une même chromatide, entre les deux chromatides d'un même chromosome ou bien entre deux chromosomes homologues ou différents. Cette dernière possibilité peut expliquer les translocations réciproques. Un NAHR entre deux chromatides non sœurs peut entraîner la formation d'un isochromosome (48).

Un crossing-over entre deux LCR homologues inter ou intra-chromatidien orientés dans le même sens engendre des duplications ou des délétions (Figure 21). Si les deux LCR sont disposés en sens inverse, le NAHR génère une inversion. De nombreuses anomalies récurrentes sont la conséquence de ce type de recombinaison comme par exemple les microdélétions 22q11.2, 7q11.23 (syndrome de Williams et Beuren)...

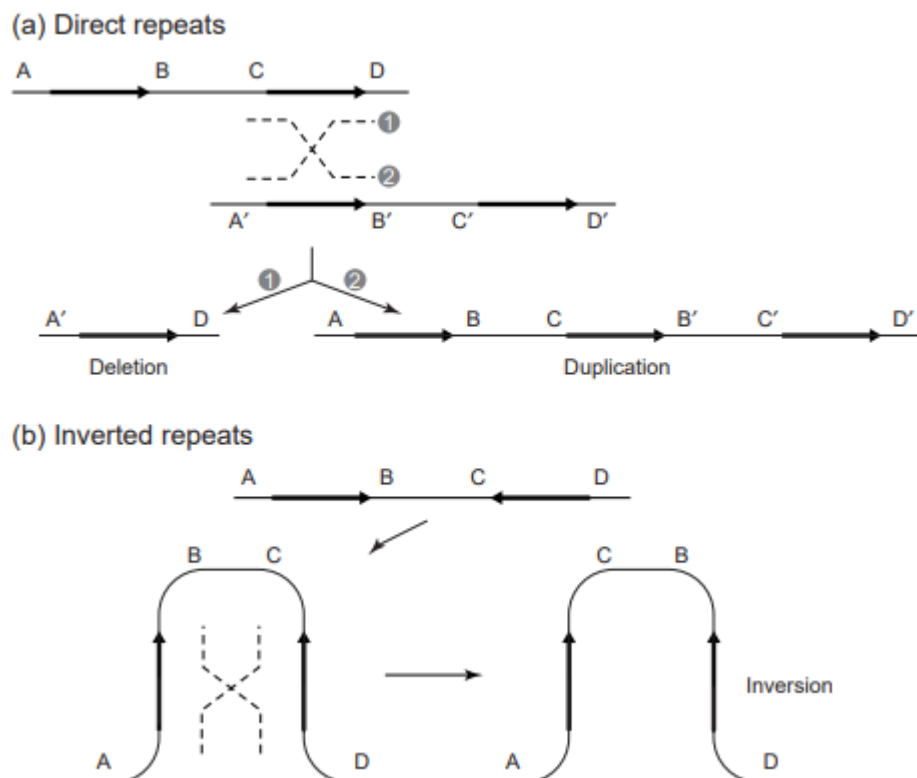


Figure 21 : Réarrangements génomiques résultant de la recombinaison homologue entre les duplications segmentaires. *Les séquences répétées sont représentées par des flèches en indiquant leur orientation. Les lignes pointillées représentent un crossing-over. (a) la recombinaison entre les répétitions du même sens entraîne une duplication et une délétion. (b) la recombinaison entre des répétitions en sens inverse entraîne une inversion (49).*

Des séquences répétées, de taille beaucoup plus petite que les LCR, présentant un haut degré d'homologie pourraient induire certaines recombinaisons homologues non alléliques rares responsables de certains remaniements non récurrents. Par exemple les séquences LINE (Long Interspersed Nuclear Elements), d'environ 7 kb et représentant 17% du génome ou encore les séquences Alu définies comme des séquences d'ADN d'origine virale, capables de se multiplier et de se déplacer de façon autonome, de 280 pb, représentant 10% du génome (50). De nombreux CNV sont la conséquence de ce type de recombinaison comme par exemple la leucodystrophie autosomique dominante de l'adulte (duplication au niveau du locus 5q23.2 impliquant le gène *LMNBI*), le syndrome de Miller-Dieker (délétion du locus 17p13.3)...

1.2.2. *Les réarrangements non récurrents*

La majorité des réarrangements non récurrents est expliquée par le mécanisme de ligature d'extrémités d'ADN non homologues (Non Homologous End Joining) qui ne nécessite pas de régions d'homologie. Son principe pour la réparation de cassure double brin de l'ADN se base sur trois voies enzymatiques : des nucléases qui vont éliminer l'ADN endommagé ; des polymérases pour réparer l'ADN et des ligase pour faire une ligation de la cassure (51) (Figure 22). L'une des caractéristiques de ce mécanisme est la délétion ou l'ajout de quelques nucléotides afin de réparer la cassure.

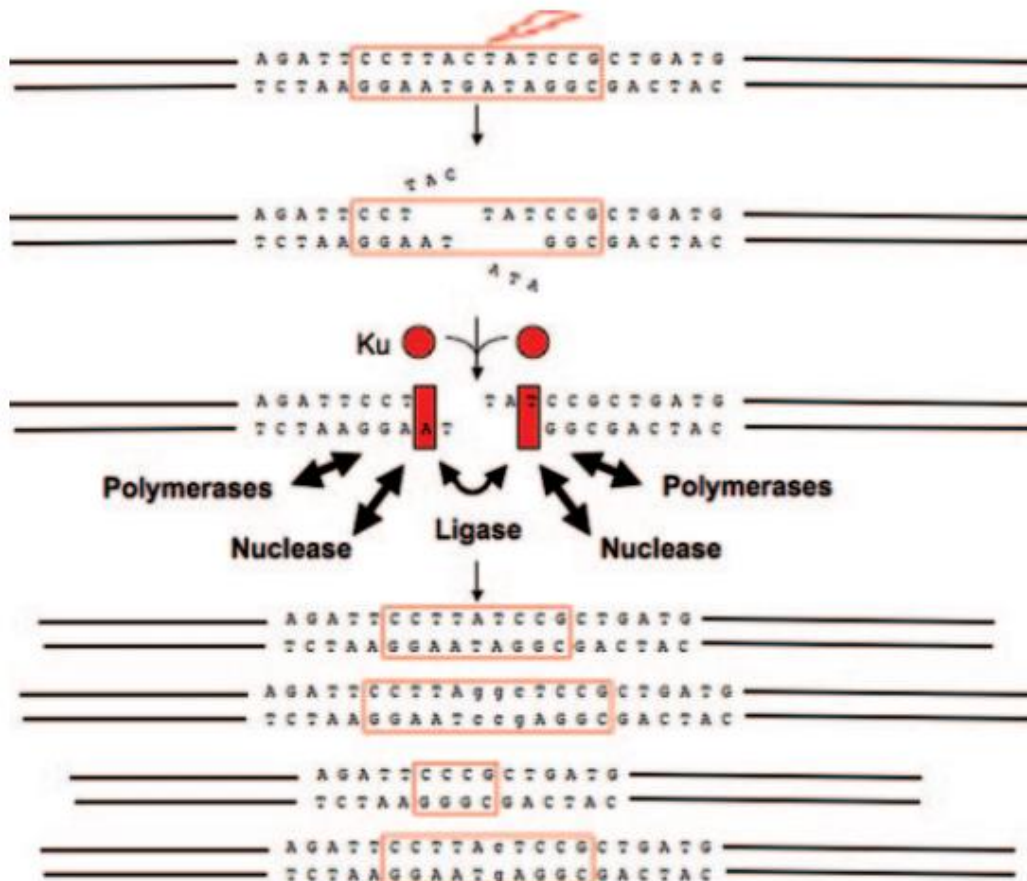


Figure 22 : Le mécanisme du NHEJ Identification des cassures double brins puis créations de ponts moléculaires entre les deux terminaisons cassées grâce au dimère Ku, un hétérodimère de Ku70 et Ku80 (également appelé Ku86) qui est une protéine de liaison à l'ADN qui peut glisser sur des extrémités d'ADN de diverses configurations (représentée sur le schéma par le changement des cercles rouges en des rectangles). Ensuite, une modification des terminaisons par ajout ou élimination de bases permet de les rendre compatibles et réparer la cassure (51).

1.2.2.1. Fork Stalling and Template Switching (FoSTeS)

Ce mécanisme a été décrit pour la première fois en 2007 par Lee et al., en étudiant la région génomique entourant le gène *PLP1* impliquée dans la maladie de Pelizaeus-Merzbacher. Il a observé des réarrangements non récurrents plus compliqués qu'une simple délétion ou duplication. En analysant les profils d'ACPA, il a remarqué que les duplications observées en FISH sont souvent interrompues par des fragments tripliqués ou délétés. La cartographie des points de cassures a révélé une complexité supplémentaire de ces réarrangements en montrant l'existence d'inversions et de translocations. Il a remarqué la présence de microhomologies de deux à cinq nucléotides à chaque jonction. Il a suggéré un nouveau mécanisme appelé FoSTeS basé sur les erreurs de la réplication d'ADN (52). Lorsque la fourche de réplication rencontre

une lésion d'ADN et s'arrête, le brin retardé va être inséré dans une deuxième fourche de réplication active à l'aide d'une microhomologie avec son extrémité 3' et permettre ainsi la poursuite de la synthèse d'ADN (Figure 23).

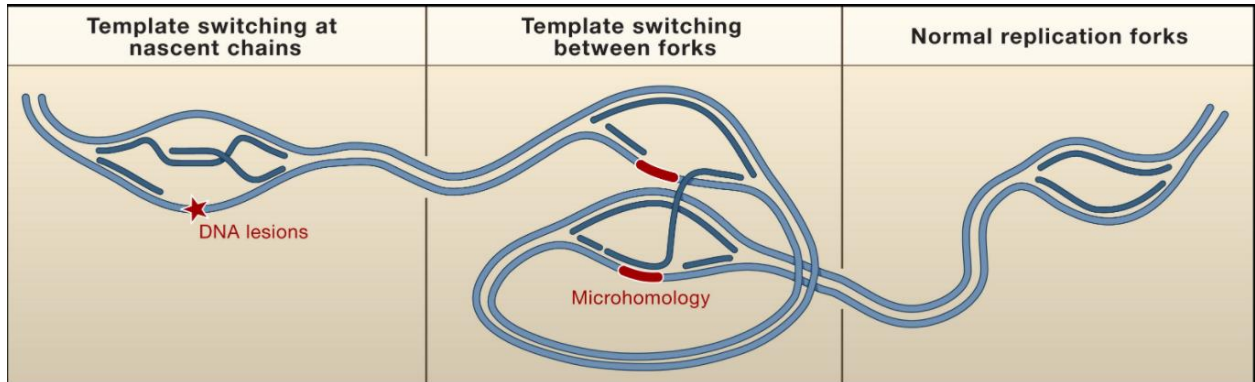


Figure 23 : Le mécanisme de FoSTeS : les fourches de réplication de l'ADN rencontrant des lésions (étoile rouge) sur le brin de matrice peuvent utiliser une autre fourche de réplication pour contourner la lésion d'ADN et générer ainsi des réarrangements chromosomiques. Ce processus nécessite des régions de microhomologie (barres rouges) (53).

1.2.2.2. Microhomology-mediated break-induced replication (MMBIR)

Un deuxième mécanisme de formation de CNV en lien avec les erreurs du mécanisme de réplication de l'ADN est le MMBIR. Lorsque la fourche de réplication rencontre une cassure du brin d'une seule extrémité, et qu'elle ne peut pas être réparée par une recombinaison homologe, l'extrémité 5' sera digérée. L'extrémité 3' libre va envahir une nouvelle fourche de réplication en s'appariant à une séquence présentant des micro-homologies (2-5pb), permettant une élongation. Ce phénomène se reproduit plusieurs fois générant des réarrangements complexes (Figure 24) (54).

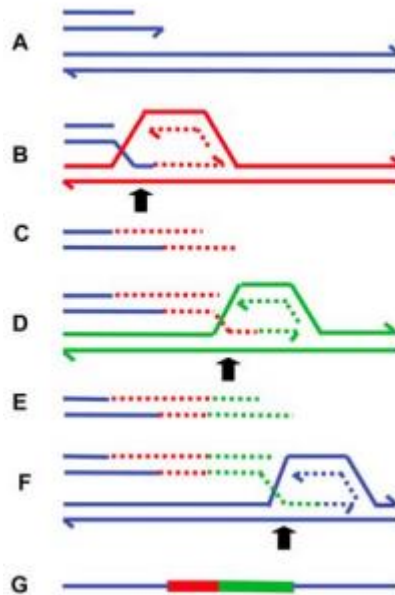


Figure 24 : Le mécanisme de MMBIR. Chaque ligne colorée présente un brin d'ADN et la demi-flèche indique l'extrémité 3'. (A) montre une cassure d'un brin avec une extrémité 3' libre et simple brin. (B) envahissement d'une nouvelle fourche de réplication. (C), (D) et (E) dans une nouvelle fourche de réplication. (F) répétition du mécanisme dans une fourche de réplication de la chromatide sœur du brin d'origine (couleur bleu) (G) Le produit final du brin réparé présentant des réarrangements complexes (54).

1.3. Méthodes d'étude des remaniements chromosomiques

1.3.1. Caryotype

Les chromosomes sont visibles sous forme de structures séparées seulement pendant la mitose. Ils correspondent à la forme la plus condensée de la chromatine, permettant la répartition de l'ADN entre les deux cellules filles lors de la division cellulaire. Le chromosome métaphasique se distingue par sa taille et sa forme. Chacun est constitué de deux molécules d'ADN représentées par deux chromatides sœurs. Les régions aux deux extrémités du chromosome sont les télomères. Chaque chromosome est caractérisé par un centromère, point d'attachement au fuseau mitotique lors de la division cellulaire. Le caryotype est une technique qui permet d'établir le nombre des chromosomes d'un individu et de préciser leur aspect, normal ou pathologique. Les chromosomes sont classés en fonction de leur taille décroissante, de l'indice centromérique et du patron de bandes. La réalisation d'un caryotype est basée sur la culture cellulaire. Le tissu le plus souvent utilisé est le sang (lymphocytes), mais d'autres tissus comme la peau (fibroblastes), la moelle (étude des hémopathies) ou encore d'autres types cellulaires comme les cellules trophoblastiques ou les amniocytes (caryotype fœtal) peuvent

être utilisés. Le sang ou les autres tissus sont cultivés dans un milieu favorable à leur division. Pour les lymphocytes, une lectine, la phytohémagglutinine, est ajoutée au milieu de culture, qui a un effet mitogène permettant une stimulation de la croissance des lymphocytes T. Après la phase de multiplication des cellules dans une étuve à 37 °C et à 5% de CO₂ (pour les amniocytes par exemple), celles-ci sont bloquées au stade de métaphase. C'est la phase durant laquelle les chromosomes sont le plus condensés permettant une analyse de leur nombre et de leur structure. Pour se faire, la colchicine est additionnée à la culture cellulaire, qui empêche la polymérisation de la tubuline et donc la formation du fuseau mitotique. Ensuite un choc hypotonique est effectué qui entraîne le gonflement des cellules par différence de pression osmotique. Les membranes cytoplasmique et nucléaire sont fragilisées et libèrent leur contenu. Les constituants cellulaires sont ensuite fixés afin de garder leur aspect condensé. Un culot chromosomique sera ainsi obtenu contenant les chromosomes de nombreuses cellules humaines. Une goutte de ce culot pourra être étalée sur une lame. A ce stade les chromosomes peuvent uniquement être classés en fonction de leur taille et leur indice centromérique. Des techniques de dénaturation appelées encore technique de marquage de bande, *banding*, permettent de distinguer une succession de bandes claires et sombres. Deux types de techniques de dénaturation sont le plus souvent utilisés : les bandes R obtenues par la dénaturation thermique et les bandes G obtenues après une dénaturation enzymatique par la trypsine (Figure 25). Ces deux marquages, en contretypage, sont complémentaires. Le marquage révèle l'euchromatine c'est à dire les régions chromosomiques où l'ADN est préférentiellement transcrit et l'hétérochromatine c'est à dire les régions chromosomiques où l'ADN est préférentiellement inactif. La dénaturation chromosomique est suivie par une coloration au Giemsa. Les lames sont ensuite observées en microscopie optique.

Le caryotype permet une analyse pangénomique du nombre et de la structure des chromosomes avec une résolution comprise entre 5 et 10 mégabases (Mb).

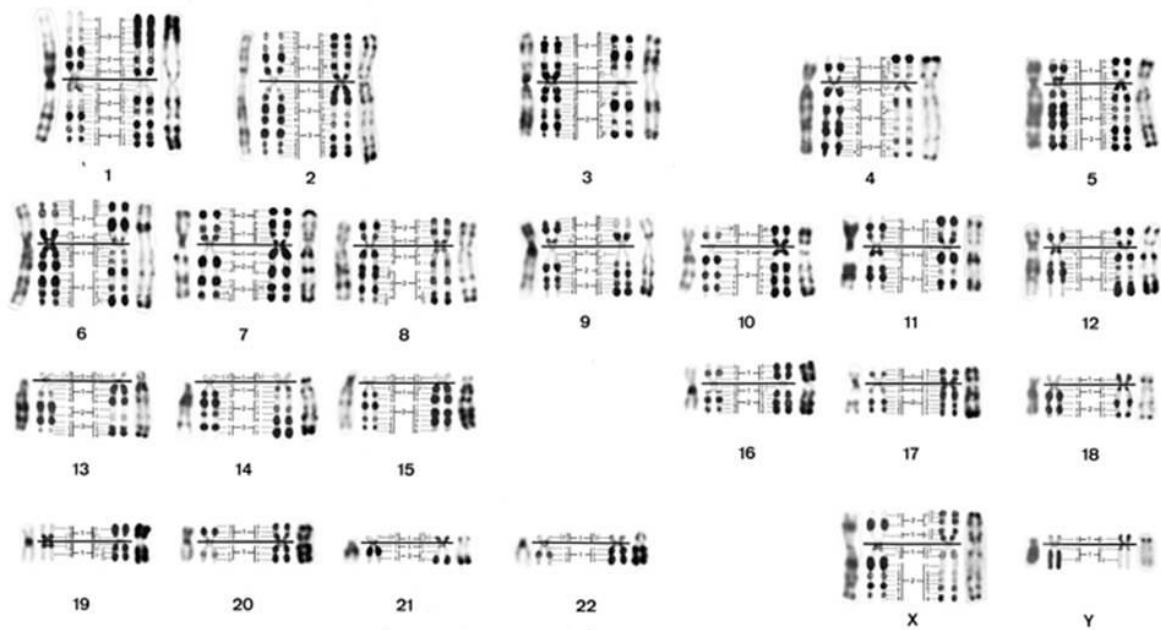


Figure 25 : Caryotype humain masculin pour chaque paire le chromosome de gauche est en bande G et celui de droite en bande R (ICSN).

1.3.2. *FISH*

Cette technique est basée sur les deux propriétés de l'ADN : la dénaturation et l'hybridation par complémentarité des bases nucléotidiques (Figure 26). Cette technique permet une étude ciblée de la structure des chromosomes à une résolution équivalente à celle de la sonde utilisée (environ 150-200 kilobases).

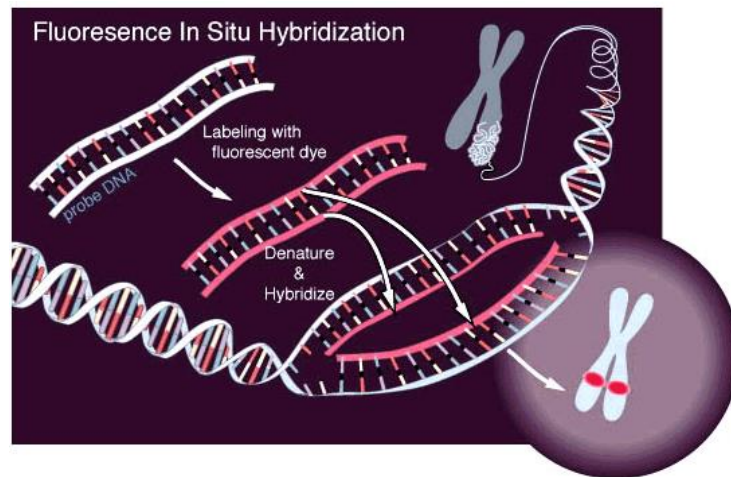


Figure 26 : Principe de la technique FISH : *Une sonde marquée par un fluorochrome est dénaturée pour obtenir un ADN simple brin. Elle s'hybride spécifiquement avec sa séquence cible complémentaire sur une préparation chromosomique dénaturée grâce à la complémentarité des bases nucléotidiques. L'observation sera effectuée grâce à un microscope à fluorescence (55).*

Cette technique est utilisée en diagnostic, notamment pour la recherche de microdélétions suspectées selon le phénotype du patient. Cette technique permet la recherche d'anomalies sur des noyaux interphasiques par exemple la recherche des trisomies 13, 18 et 21 en diagnostic prénatal. Elle est également utilisée pour confirmer un déséquilibre observé en ACPA.

Différents types de sondes sont utilisées :

- Les sondes chromosomes spécifiques (centromériques ou locus spécifiques), elles sont utilisées pour la détection des anomalies de nombre.
- Les sondes locus spécifiques (LSI) et subtélomériques, elles sont utilisées pour la recherche des syndromes microdélétionnels interstitiels ou subtélomériques.
- Les sondes de peinture chromosomique (WCP pour Whole Chromosome Painting). Ces sondes permettent essentiellement de caractériser une anomalie chromosomique mise en évidence par une autre technique.
- La M-FISH ou caryotype multi couleur, contrairement à la FISH, permet une analyse globale du génome. Cette technique permet l'identification de chromosomes dérivés ou marqueurs et de remaniements complexes (56).

La plupart des sondes locus-spécifiques sont fabriquées à partir de BAC (Bacterial Artificial Chromosome). Ces BAC sont choisis dans les banques de données du génome humain

(<http://genome.ucsc.edu/>) et reçus sous la forme de cultures de bactéries recombinantes contenant la séquence ADN d'intérêt qui sera extraite et marquée par un fluorochrome avec par exemple la technique de nick-translation. Ces BACs présentent une approche efficace pour identifier les points de cassure d'un remaniement chromosomique par la « marche sur le chromosome » (57) (figure 27). Cette stratégie nécessite la réalisation d'un caryotype, afin d'identifier au mieux les régions chromosomiques contenant les points de cassure, puis des techniques de FISH itératives. Les sondes sont choisies de part et d'autre des points de cassure théoriques. A chaque hybridation, la position des sondes utilisées est qualifiée sur le dérivé chromosomique et le chromosome normal permettant ainsi de choisir de nouvelles sondes jusqu'à encadrer, voire chevaucher, le point de cassure à déterminer (Figure 27).

Bien que cette approche soit abordable tant sur la faisabilité au laboratoire que sur le plan financier, elle pose plusieurs problèmes :

- de nombreuses sondes doivent être utilisées avant d'identifier la sonde près du point de cassure, d'où l'emploi principalement de sondes non commerciales et la nécessité de disposer d'une banque de sondes couvrant tout le génome dans le laboratoire,
- cette technique peut prendre plusieurs mois (fabrication des sondes, FISH itératives...),
- la quantité de matériel chromosomique du patient doit être suffisante, la méthode se faisant séquentiellement ; dans les cas limites, il est parfois nécessaire de déshybrider des lames pour tester de nouvelles sondes. L'existence d'une lignée cellulaire immortalisée du patient permet de pallier ce problème.

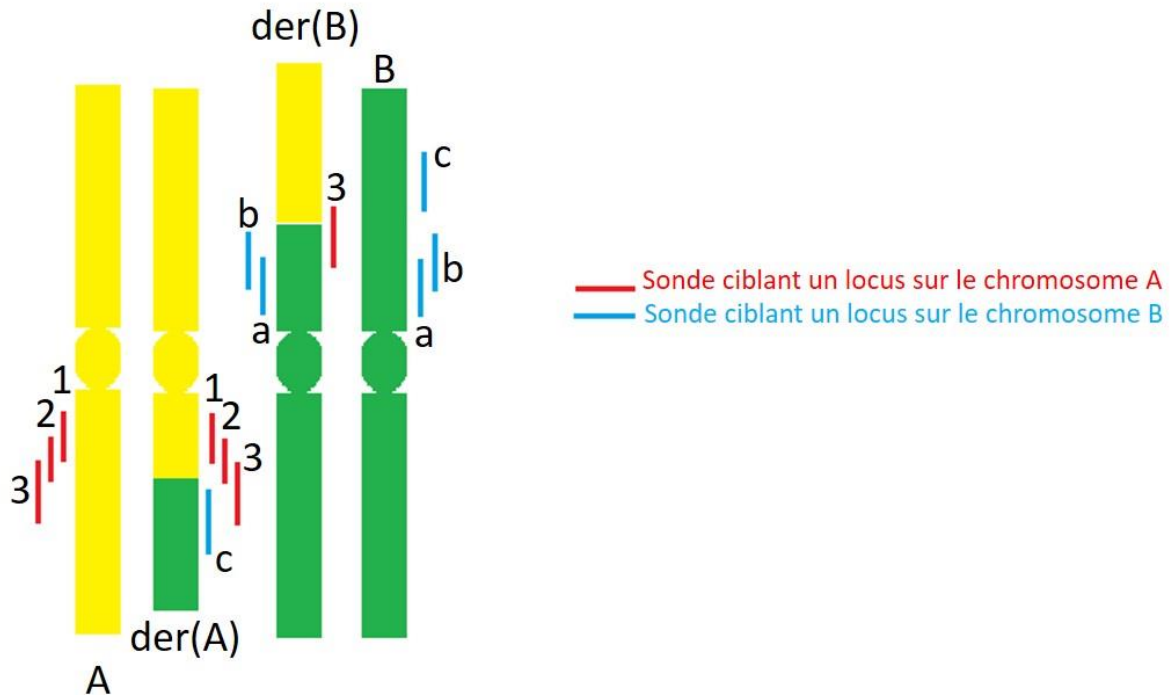


Figure 27 : Marche sur chromosome par FISH. Caractérisation du point de cassure par FISH en analysant la position relative de chaque sonde sur le chromosome normal et les chromosomes dérivés en déduisant pour chaque sonde si le point de cassure se situe en amont (cas des sondes 1, 2, a et b), chevauche le point de cassure (sonde 3) ou en aval de la sonde hybridée (sonde c).

1.3.3. *a*-CGH ou ACPA

Inspirée de la technique FISH, l'hybridation génomique comparative (CGH, *Comparative Genomic Hybridization*) permet une observation pangénomique du génome et d'identifier des réarrangements déséquilibrés, tels que les duplications et délétions, mais n'est pas capable de détecter les réarrangements équilibrés tels que les inversions et translocations. Au début cette technique consiste à hybrider deux ADN, un de référence et l'autre d'intérêt, marqués avec des fluorochromes différents, sur une préparation de chromosomes en métaphase provenant d'un individu sans remaniement chromosomique. En comparant l'intensité de la fluorescence, on peut identifier les gains et les pertes le long des chromosomes (Figure 28) (58).

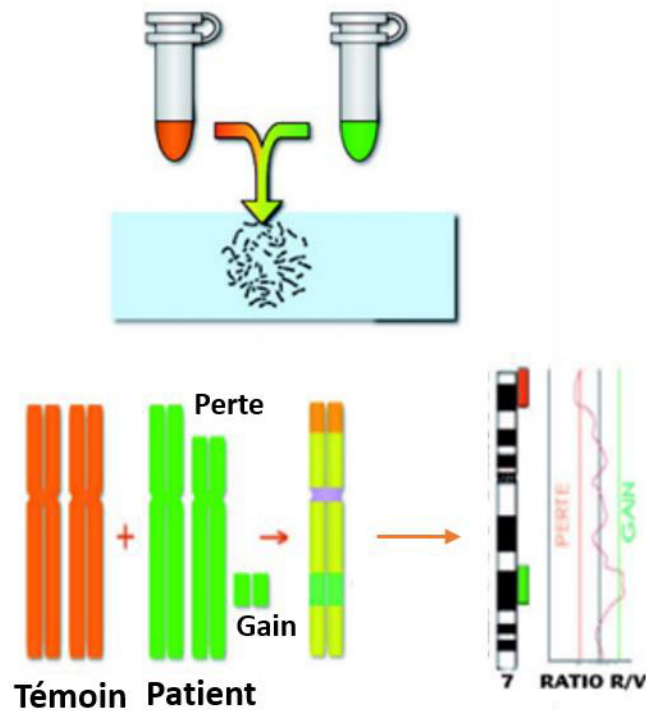


Figure 28 : Principe de la CGH sur chromosome. L'ADN du témoin est marqué en rouge, celui du patient en vert, après hybridation sur une lame chromosomique d'un témoin l'analyse de la CGH sur chromosome est basée sur la mesure de l'intensité des 2 fluorochromes : couleur jaune indique la présence de la même quantité, le vert indique du gain de matériel de cette région chez le patient et le rouge montre une perte. Cette intensité est représentée par un ratio en échelle logarithmique (\log_2), la valeur zéro indique l'égalité entre les 2 échantillons, <0.8 pour une perte et >1.25 pour un gain (58).

Comme la résolution de cette technique est limitée par la qualité des chromosomes utilisés, et afin d'améliorer sa résolution, les chromosomes ont été remplacés par des clones génomiques type BAC, YAC, cosmides puis par des fragments d'ADN, oligonucléotides, appelées sondes, réparties sur l'ensemble du génome et fixées sur une lame le plus souvent en verre ou en silice (Figure 29). Cette technique est appelée CGH-array (Comparative Genomique Hybridization on micro-array) ou Analyse Chromosomique sur Puce à ADN (ACPA). De ce fait, la résolution de l'ACPA est déterminée par la densité des sondes et la distance génomique les séparant ou par la longueur des segments d'ADN clonés (59). En augmentant le nombre des sondes, la résolution de la puce sera plus grande et des anomalies de plus petites tailles pourront être identifiées.

Il existe 2 types de sondes oligonucléotidiques :

- Des sondes monomorphes qui évaluent uniquement le nombre de copies. Ils sont potentiellement très nombreux et régulièrement espacés le long du génome, permettant une étude du génome en haute résolution.

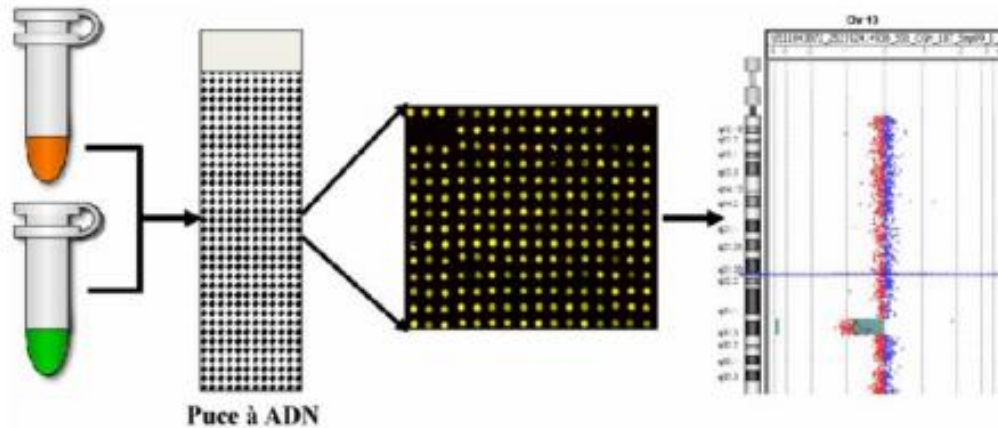


Figure 29 : Principe a-CGH. L'ADN du patient et celui du témoin sont marqués par deux fluorochromes différents et sont hybridés sur une lame présentant des oligonucléotides. L'analyse est basée sur le rapport de l'intensité de fluorescence reflétant les CNV (60).

- Les sondes génétiques bi-alléliques, Single Nucleotide Polymorphism (SNP), portant une variation sur une paire de base. Après l'établissement de la carte des SNP produite par le projet HapMap (61), ces SNP sont utilisés afin d'identifier des CNV et de déterminer l'origine parentale d'un remaniement (avec l'étude des parents). Ces sondes permettent ainsi d'identifier également des pertes d'hétérozygotie et de faire des analyses de liaison. Son principe (Figure 30) ne comporte pas d'hybridation compétitive comme la CGH-array, mais consiste à hybrider l'ADN génomique du patient seul sur les sondes fixées sur une puce.

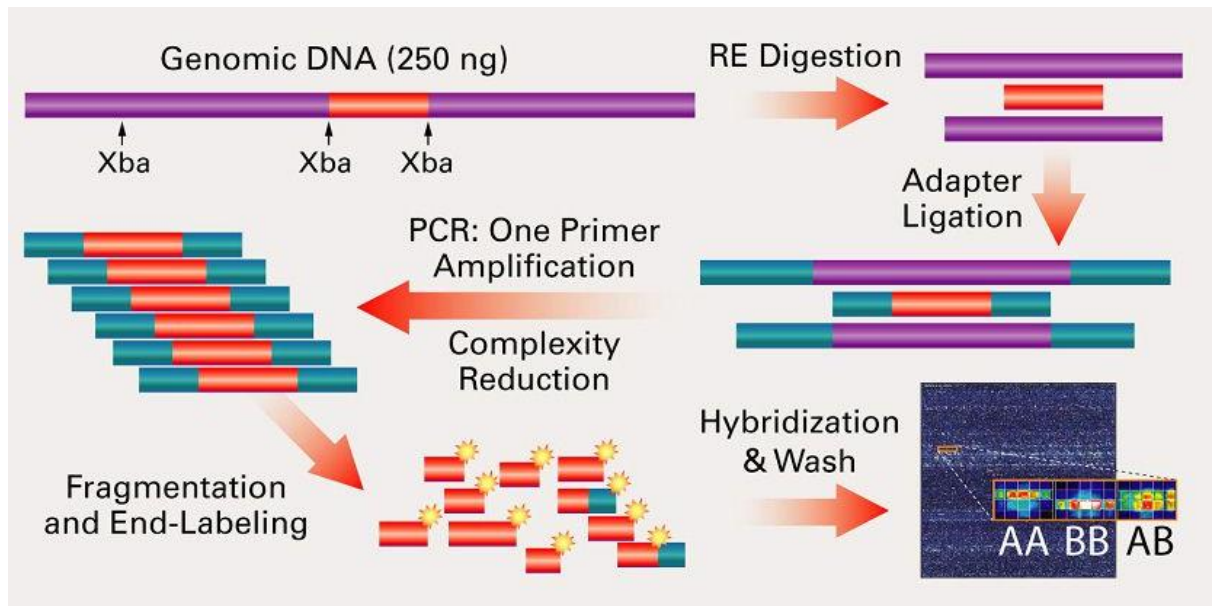


Figure 30 : Principe de SNP-array. L'ADN génomique du patient est fractionné grâce à des enzymes de digestions. Ensuite des adaptateurs sont ajoutés aux fragments d'ADN afin de l'amplifier. Cette amplification des fragments d'ADN est suivie d'une fragmentation et d'un marquage par un fluorochrome. Enfin, les fragments d'ADN marqués sont hybridés sur la puce sur laquelle sont fixés les marqueurs SNP (62).

L'analyse de l'ACPA permet de définir les variations du nombre de copies (CNV, *copy number variation*) et l'interprétation de ces CNV repose sur plusieurs critères (63,64) :

- La nature du CNV : perte ou, gain, d'une ou plusieurs copies,
- Sa taille : généralement, un CNV pathogène est significativement plus grand qu'un CNV bénin. Pourtant, certains CNV de très petites tailles peuvent être pathologiques s'ils concernent quelques exons d'un gène pathologique,
- Son contenu génique : un CNV dans une région dont la densité génique est élevée est en faveur de son caractère pathogène, de même si le CNV concerne un gène décrit en pathologie humaine et qu'il existe une corrélation avec le phénotype en tenant compte des conséquences fonctionnelles d'un CNV impliquant un gène donné (dominant, récessive, lié à l'X)
- Sa présence dans la base de données de population contrôle (DGV) ou bien dans les bases de données des patients (Decipher, ClinVar),
- Données de ségrégation familiale : un CNV *de novo* est plutôt en faveur d'un caractère pathogène alors qu'un CNV hérité est plutôt en faveur d'un caractère bénin. Il existe des exceptions comme la pénétrance incomplète, l'expressivité variable, une délétion héritée

révélant une maladie récessive, la transmission liée à l'X, les régions soumises à empreinte, un parent porteur du CNV à l'état de mosaïque ou bien un CNV pathogène familial.

Les CNVs sont classés en 5 catégories (65,66) : Les CNV dits bénins qui sont des polymorphismes retrouvés fréquemment dans la population générale, les CNVs probablement bénins, qui correspondent à des variants plus rares sans conséquence phénotypique, les CNV dits pathogènes et probablement pathogènes responsables du phénotype du patient ; enfin les variants de signification inconnue ou VOUS (Variant Of Uncertain clinical Significance) dont la pathogénicité potentielle et l'implication dans le phénotype du patient sont incertaines ou inconnues, avec les connaissances actuelles.

1.4. Les conséquences des anomalies chromosomiques

Les anomalies chromosomiques sont impliquées dans les anomalies du développement et dans les troubles de la reproduction.

Pour les anomalies du développement, et dans la majorité des cas, les anomalies de nombres sont accidentelles, donc non récurrentes dans la fratrie, à l'exception des trisomies résultant des translocations robertsoniennes parentales. Seules trois trisomies autosomiques homogènes viables sont : les trisomies 13 (syndrome de Patau), 18 (syndrome d'Edwards) et 21 (syndrome de Down). Aucune monosomie autosomique homogène n'est viable. Les autres trisomies autosomiques homogènes sont létales et provoquent des fausses couches. Elles ne sont viables qu'en cas de mosaïque ou de trisomie partielle (par exemple en présence d'un marqueur chromosomique surnuméraire). Les anomalies de structures déséquilibrées sont principalement responsables des anomalies congénitales. Les dérivés de translocations (monosomie partielle d'une région chromosomique et trisomie partielle d'une autre) ou des inversions sont également à l'origine d'anomalies congénitales.

Dans le cadre des anomalies de la reproduction masculine, la première cause de l'infertilité est le syndrome de Klinefelter, caractérisé par la présence d'un chromosome X surnuméraire. Dans ce cadre, la probabilité de trouver une anomalie chromosomique de structure est de 2 à 20 fois plus élevée que dans la population générale (67). En cas d'anomalies de structure déséquilibrée, le phénotype dépend de la région impliquée et de son contenu génique, dans ce cas l'infertilité masculine est associée avec d'autres manifestations cliniques (à l'exception de la microdélétion du chromosome Y). Pour cela les anomalies de structure équilibrées sont les plus rencontrées dans le cadre d'infertilité. Ces anomalies peuvent altérer

la spermatogenèse. La malségrégation chromosomique dans les gamètes a pour conséquence des fausses couches spontanées ou la naissance d'un enfant porteur d'un déséquilibre chromosomique. Comme les points de cassure de ces remaniements sont aléatoires, ils peuvent interrompre un gène impliqué dans la spermatogenèse, ou altérer son expression par effet de position.

Les réarrangements chromosomiques conduisent à des conséquences dépendant du matériel chromosomique impliqué dans l'anomalie. En général, les réarrangements déséquilibrés se traduisent par un phénotype anormal et leur conséquence phénotypique dépend des gènes impliqués dans le réarrangement. Les remaniements chromosomiques apparemment équilibrés (RCAE) sont associés à un phénotype normal, car en principe la totalité de l'information génétique est présente et leurs risques consistent à la production de gamètes déséquilibrés. Cependant très rarement on observe un phénotype anormal. Différentes hypothèses peuvent expliquer leur phénotype :

- **Présence d'un microremaniement déséquilibré identifié par l'utilisation de techniques plus résolutive.**

Pendant longtemps la translocation réciproque a été considérée comme équilibrée, car elle se base sur le principe qu'il n'y a ni perte ni gain de matériel génétique en particulier car le caryotype a une résolution de 5 à 10 Mb. Avec le progrès des techniques, et l'augmentation de la résolution, le terme équilibré a été remplacé par apparemment équilibrée. La FISH et la CGH-array ont permis de diagnostiquer des déséquilibres cryptiques non décelables par le caryotype, associés à des RCAE chez des patients à phénotype anormal. Ainsi de Gregori M et al. rapportent une délétion chez 40% des patients ayant une translocation réciproque (68), donnée similaire rapportée par Schluth-Bolard C et al., en 2009, qui a observé des déséquilibres chez 20 patients parmi 47 avec une déficience intellectuelle et/ou des malformations congénitales multiples présentant un RCAE. Les délétions étaient soit présentes au niveau d'un point de cassure, ou bien en dehors des points de cassures. Pour un cas, les deux délétions étaient localisées au niveau et à distance des points de cassures (69).

- **Interruption d'un gène au niveau des points de cassure**

L'interruption d'un gène par un point de cassure a pour conséquence d'altérer la séquence codante, entraînant un mécanisme d'haplo-insuffisance, ou la quantité de la protéine produite

est insuffisante à la fonction normale. Plusieurs équipes ont associé l'implication d'un gène interrompu au phénotype d'un patient (16). De plus, l'interruption d'un gène peut révéler une maladie récessive. C'est le cas des translocations X-autosome avec expression d'une maladie récessive liée à l'X chez les filles, qui ont permis d'identifier de nombreux gènes. Par exemple, l'identification du gène *KLF8*, impliqué dans la déficience intellectuelle chez une patiente présentant une $t(X;21)$ (70).

- **Formation d'un gène de fusion**

Une autre conséquence des RCAE est la formation de gènes de fusion. La plupart des cas de gènes de fusion ont été décrits dans le domaine de l'oncologie. L'exemple le plus connu est le chromosome Philadelphie résultant de la translocation entre le chromosome 9 et 22 responsable d'un gène de fusion entre les gènes *BCR* et *ABL*. La formation de gène de fusion a été rapportée très rarement en cytogénétique constitutionnelle, c'est le cas entre les 2 gènes *PTK2-THOC2* chez une patiente porteuse d'une translocation $t(X;8)$ et présentant une forme non progressive d'ataxie congénitale, des troubles cognitifs et une hypoplasie cérébelleuse (71).

- **Effet de position**

Les RCAE peuvent entraîner une modification des régions régulatrices dues aux changements de l'environnement chromatinien des gènes (72). Ces modifications s'expliquent par plusieurs mécanismes (figure 31) : le point de cassure sépare le gène et sa région promotrice ce qui pourrait entraîner des effets répresseurs (figure 31A). A l'inverse, le point de cassure peut rapprocher un gène d'un élément régulateur d'un autre gène et ainsi possiblement modifier son expression (figure 31B). Le point de cassure peut également déplacer un gène à proximité d'un autre gène, ces deux gènes entrants alors en compétition dans la régulation avec les éléments régulateurs (figure 31D). Enfin, le remaniement peut placer le gène dans un environnement hétérochromatinien qui peut empêcher ou diminuer la transcription du gène par propagation de l'effet répresseur de ces régions appelée PEV (Position Effect Variegation) (Figure 31E).

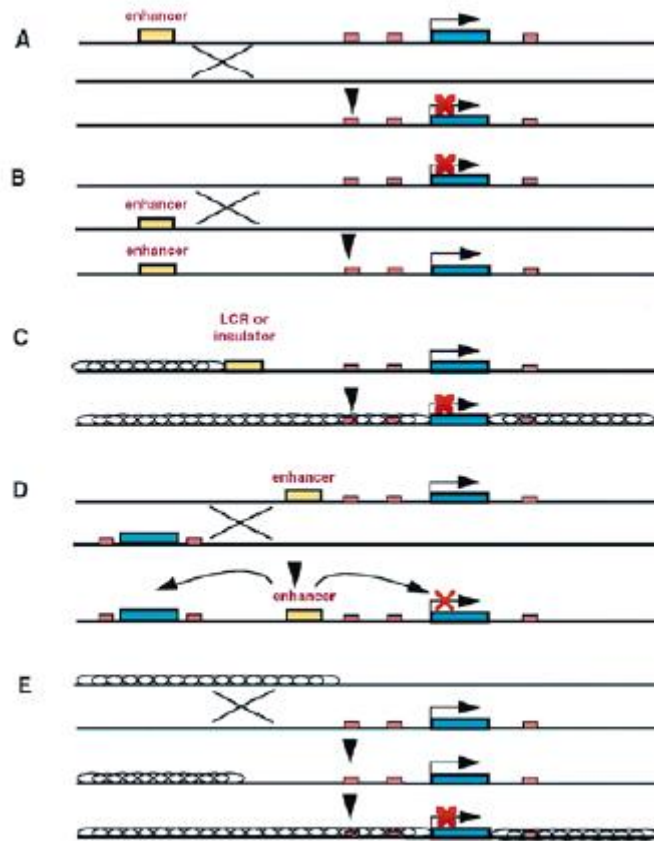


Figure 31 : Les différents mécanismes expliquant un effet de position. (A) Le point de cassure se situe entre l'unité de transcription et le promoteur. (B) Le rapprochement d'un gène avec un élément régulateur d'un autre. (C) Elimination d'un LCR ou d'un insulateur entraîne une suppression d'une séquence régulatrice. (D) Concurrence entre deux gènes avec les éléments régulateurs de l'un des deux. (E) PEV (72).

- **Association fortuite entre un RCAE et un phénotype anormal : une autre étiologie génétique**

Comme le suggère Schluth-Bolard C en 2009 (69), il faut éliminer une autre cause génétique responsable du phénotype chez le patient. Une étude récente (73) a montré que 5% des patients présentent au moins deux maladies mendéliennes.

- **Cause non génétique**

L'explication d'un phénotype anormal chez un patient présentant un RCAE peut ne pas être liée à une étiologie génétique. Ce phénotype anormal peut être secondaire à une cause acquise comme une origine toxique, infectieuse...

1.5. Troubles de la reproduction : infertilité masculine

1.5.1. Généralité

L'appareil génital masculin comprend différents organes et structures permettant la production, la maturation, le stockage et le transport des spermatozoïdes jusqu'à l'appareil génital féminin. Il est constitué (Figure 32) :

- De deux testicules (ou gonades mâles) qui assurent deux fonctions essentielles endocrine, la synthèse et la sécrétion d'hormones stéroïdes principalement la testostérone et exocrine, la production des spermatozoïdes ;
- Du tractus génital formé des voies spermatiques intra- testiculaires (tubes droits et *rete testis*) et extra-testiculaires (canaux efférents, épидидyme, canal déférent et canal éjaculateur) qui assurent le transport des spermatozoïdes ;
- Des glandes annexes (vésicules séminales, prostate et glandes de Cowper), produisant la majeure partie du sperme, qui sert de liquide de transport et de nutrition pour les spermatozoïdes ;
- Du tractus uro-génital représenté par l'urètre (prostatique, périnéal et pénien), s'ouvrant à l'extérieur par le méat urinaire.

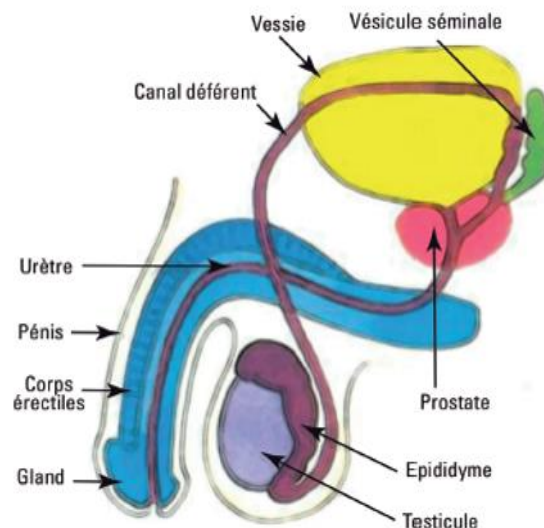


Figure 32 : Illustration schématique de l'appareil génital masculin (74).

L'intérieur du testicule contient des tubes séminifères et du tissu entre les tubules appelé espace interstitiel. Les tubes séminifères sont le lieu de la production des gamètes mâles, et donc de la

spermatogénèse. La paroi de chaque tube est constituée par l'épithélium séminifère, où sont associées les cellules somatiques de Sertoli et les cellules de la lignée germinale, reposant sur une enveloppe périvitulaire ou lamina propria. Les cellules de Sertoli forment une couche unistratifiée qui s'étend de la membrane basale à la lumière des tubes. Elles entourent les différents types de cellules germinales plus ou moins complètement avec leurs prolongements cellulaires. La spermatogénèse s'effectue ainsi au contact étroit des cellules de Sertoli qui n'ont pas seulement une fonction de soutien et de nutrition, mais aussi une activité sécrétrice et phagocytaire. Elles sont reliées les unes aux autres un peu au-dessus de la membrane basale par des jonctions serrées (tight junction), il en résulte deux zones dans l'épithélium : une zone basale dans laquelle sont alignées les spermatogonies et une zone adluminale où se développent les cellules de tous les autres stades de la spermatogénèse. Les jonctions serrées permettent d'établir une barrière hémato-testiculaire dans le tube séminifère, barrière importante, car les cellules haploïdes présentent des caractéristiques antigéniques particulières et doivent de ce fait être séparées du reste de l'organisme.

1.5.2. *La spermatogénèse*

La spermatogénèse est un processus continu permettant à partir des cellules souches diploïdes ($2n$), les spermatogonies, d'obtenir des cellules haploïdes ($1n$), les spermatides qui vont se différencier en spermatozoïde à travers deux mécanismes : la spermiogénèse et la spermiogénèse (75) (Figure 33).

La spermiogénèse comprend la prolifération des spermatogonies, la différenciation en spermatocytes et la méiose (division méiotique des spermatocytes et production des spermatides). La spermiogénèse correspond à la différenciation des spermatides ronds en spermatozoïdes.

Ce processus permet la production massive et continue de spermatozoïdes au cours de la vie reproductive de l'individu. Le temps nécessaire pour qu'une spermatogonie se transforme en spermatozoïde est de 74 jours (75).

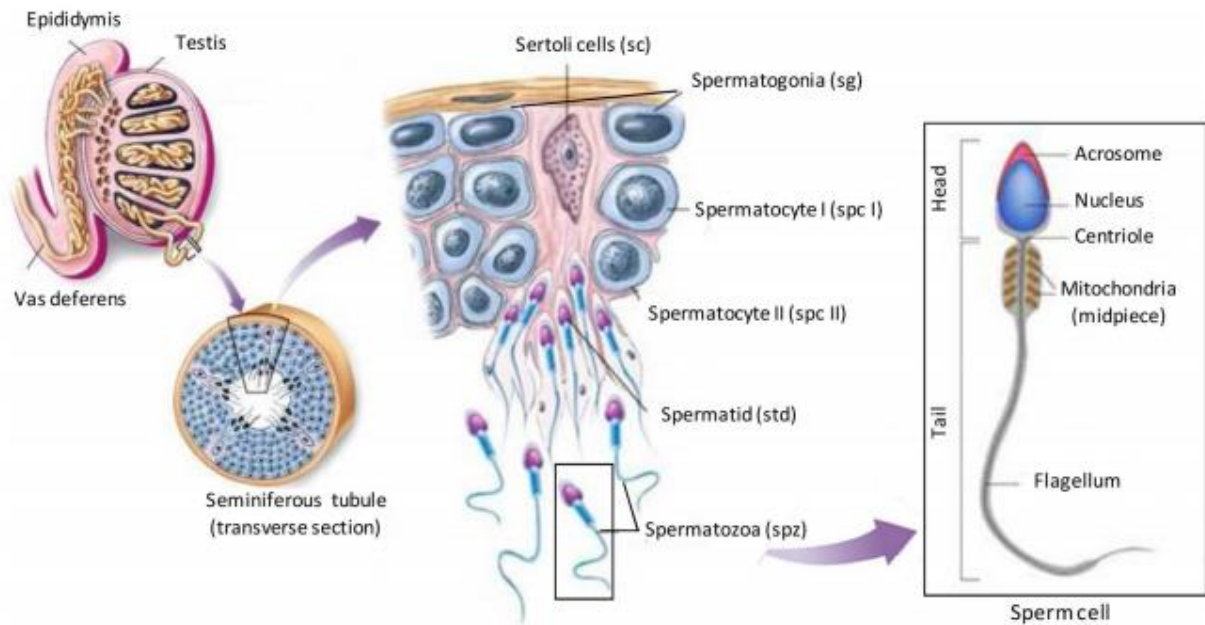


Figure 33 : Les différentes étapes de la spermatogénèse (76).

1.5.3. *La spermocytogénèse*

1.5.3.1. *Phase de multiplication*

C'est la première phase reposant sur la mitose des spermatogonies. Il existe trois types de spermatogonies (77): les spermatogonies de type Ad (d=dark), les spermatogonies de type Ap (p=pâle) et les spermatogonies de type B.

Les spermatogonies Ad sont les cellules souches de la spermatogénèse. Chaque cellule se divise par mitose en une spermatogonie Ap et une spermatogonie Ad. La spermatogonie Ad assure la réserve du stock des spermatogonies tandis que la spermatogonie Ap se divise par mitose en deux spermatogonies B. Les spermatogonies B sont à l'origine de la formation des spermatocytes I.

1.5.3.2. *Phase d'accroissement*

Cette phase correspond à l'interphase et au début de prophase I. Chaque division cellulaire commence par une phase de croissance (G1) suivie d'une phase de synthèse de l'ADN (S, les spermatocytes I doublent leur ADN ($2n$ chromosomes, $4n$ ADN) et une phase G2 où la cellule croît afin d'avoir suffisamment d'organites pour sa descendance. Les spermatocytes I, initialement localisés dans le compartiment externe du tube séminifère, passent dans le compartiment interne.

1.5.3.3. Phase de maturation

La méiose est un processus propre de la cellule germinale pour la production des gamètes. Elle est composée de deux divisions successives, la méiose I et la méiose II précédée par une phase unique de réplication de l'ADN. Elle va permettre la division des spermatocytes I diploïdes en spermatocytes II puis en quatre spermatoïdes haploïdes, mais portant des combinaisons d'allèles différents (Figure 34). Donc, elle diffère de la division mitotique. Premièrement, les chromosomes homologues s'apparient. Deuxièmement, des échanges entre les chromosomes homologues se produisent (crossing-over), ce qui aboutit à des segments de chromosomes avec de nouvelles constitutions. Finalement, le lot chromosomique est réduit de moitié pendant la première division.

La première division méiotique (méiose I) ou division réductionnelle comporte une prophase I longue qui est une étape cruciale du processus méiotique (78). La prophase est elle-même divisée en 5 stades distincts : le leptotène, le zygotène, le pachytène, le diplotène et la diacinèse.

1. Le stade leptotène (du grec, *leptos* = fin) : Pendant ce stade, la chromatine se condense en chromosomes bien individualisés qui s'apparient en chromatides sœurs. A ce stade, une structure protéique appelée élément axial se forme le long de chaque paire de chromatides sœurs pour former une structure continue. Les chromosomes homologues commencent à se rapprocher les uns des autres. Les centrioles se sont dupliqués et débutent leur migration aux pôles opposés de la cellule.
2. Le stade zygotène (du grec, *zygos* = par paires) : Les chromosomes homologues s'unissent grâce à un processus de reconnaissance et d'appariement (79).
3. Le stade pachytène (du grec, *packhus* = épais) : c'est le stade le plus long de la prophase I, caractérisé par l'appariement complet des chromosomes homologues autosomes. En ce qui concerne la paire de chromosomes sexuels X et Y, les gonosomes, sont réunis dans une structure particulière, une vésicule. Les régions d'homologie entre les gonosomes sont situées aux extrémités de ceux-ci : ce sont les régions pseudo autosomiques du bras court (PAR1) et du bras long (PAR2). En effet, c'est normalement uniquement par ces deux petites régions que les chromosomes X et Y peuvent s'apparier. La réparation des cassures double brins s'effectue par un *crossing-over* entre chromatides non-sœurs. A la fin de ce stade, toutes les cassures double brin sont réparées (79).

4. Le stade diplotène (du grec, *diploos* = double) : Les chromosomes homologues se séparent pour la plupart, mais restent attachés les uns aux autres aux endroits où ont eu lieu les recombinaisons, appelées chiasma.
5. Le stade diacinèse (du grec, *dia* = à travers et *kinesis* = mouvement) : est le dernier stade de la prophase I, caractérisé par une condensation maximale des chromosomes. La membrane nucléaire disparaît. La cellule rentre en métaphase I.

La première division de méiose se poursuit par :

- La métaphase I : les chromosomes vont s'aligner à l'équateur de la cellule pour former la plaque équatoriale.
- L'anaphase I : les chromosomes homologues migrent chacun vers un pôle de la cellule.
- La télophase I : qui est l'étape finale de la première division, les chromosomes se décondensent et l'enveloppe nucléaire se reforme autour des chromosomes. La cellule mère se sépare alors en deux cellules filles appelées spermatocytes secondaires.

Ces spermatocytes II subissent la seconde division méiotique (méiose II) ou division équationnelle, qui peut s'apparenter à une mitose classique, mais sans réelle prophase au début de cette division. Elle correspond à la disjonction des chromatides des chromosomes ainsi les chromatides sœurs se répartissent entre les deux cellules filles pour donner les spermatides rondes haploïdes.

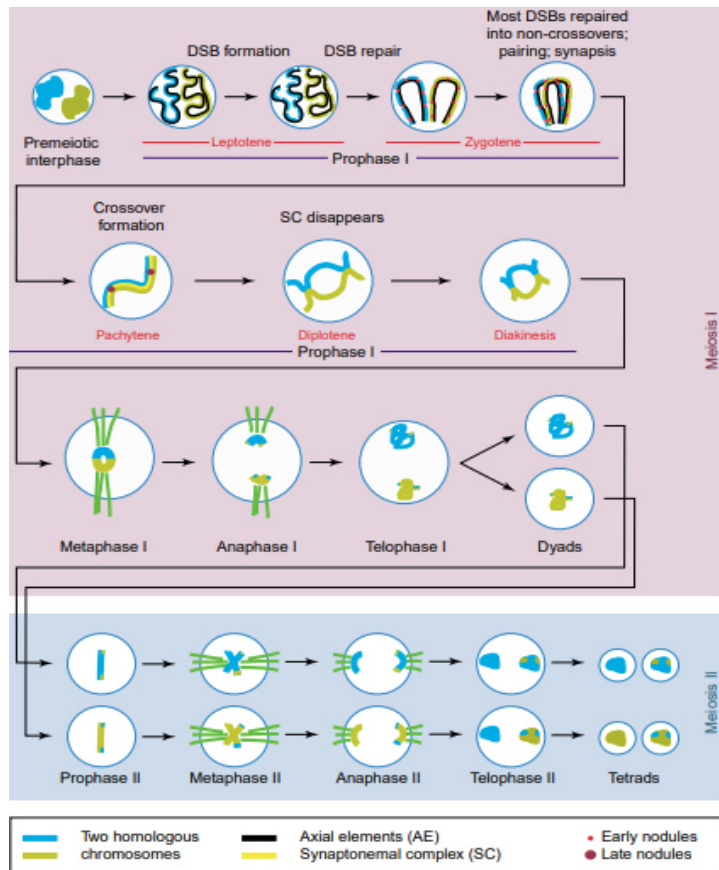


Figure 34 : Les principales étapes de la méiose. Une seule paire de chromosomes est représentée et chaque homologue est représenté dans une couleur différente. Les événements majeurs au cours de la prophase I sont représentés, y compris la formation et la réparation de la rupture double brin de l'ADN (DSB), la formation de croisements, l'appariement des chromosomes homologues, le synapsis et le crossing-over. Les chromosomes homologues se séparent en anaphase I. Pendant la deuxième division de la méiose (méiose II), les chromatides se séparent, un peu comme lors d'une division mitotique normale (78).

1.5.4. La spermiogenèse

Ce processus assure la transformation d'une cellule arrondie en une cellule effilée à cytoplasme réduit avec une réorganisation du noyau et le développement de l'acrosome. Il comprend plusieurs étapes qui peuvent se dérouler d'une manière synchrone (Figure 35).

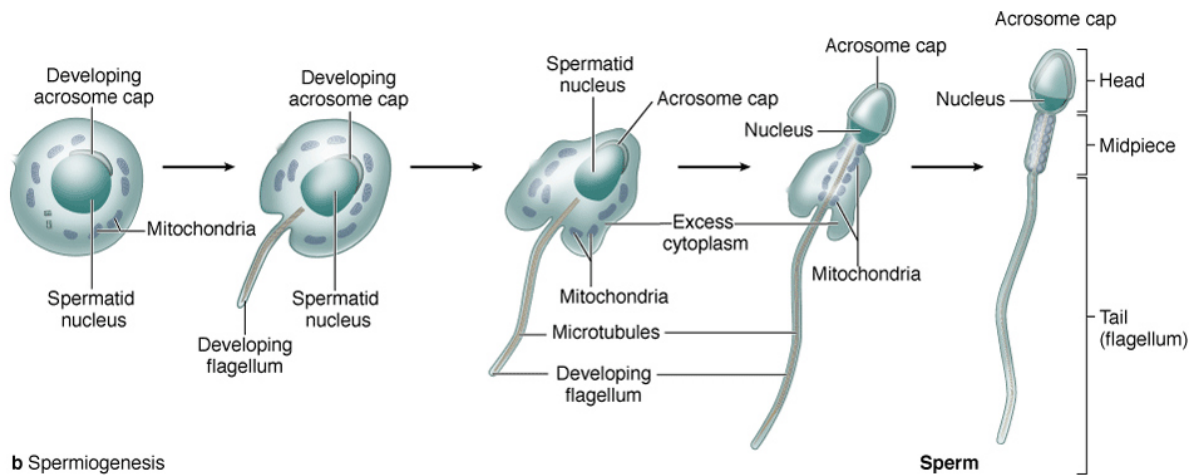


Figure 35 : Spermiogénèse. *La maturation des spermatides impliquant la condensation du noyau, développement de l'acrosome, du flagelle et l'excrétion du cytoplasme (80).*

Durant la spermiogénèse, la taille du noyau diminue avec la condensation de la chromatine en remplaçant les histones par des protéines de transitions, TNP1 et TNP2, remplacés ensuite par les protamines, PRM1 et PRM2 (81). Les protamines présentent plusieurs domaines de liaison à l'ADN appelé « *enchroaching domains* » riches en arginine (acide aminé chargé positivement), qui se lie fortement au phosphosquelette de l'ADN (chargé négativement). Les protamines sont aussi riches en cystéine. La formation de ponts disulfures intra et inter moléculaires entre ces résidus favorise la compaction et la stabilisation de la chromatine (82) (Figure 36).

Ce réarrangement de la chromatine est nécessaire pour protéger le matériel génétique du gamète mâle contre les agressions physiques et chimiques durant son transit dans les voies génitales mâles et femelles. La réduction de la taille du noyau qui résulte de la condensation de sa chromatine rend le spermatozoïde plus hydrodynamique et favorise sa mobilité.

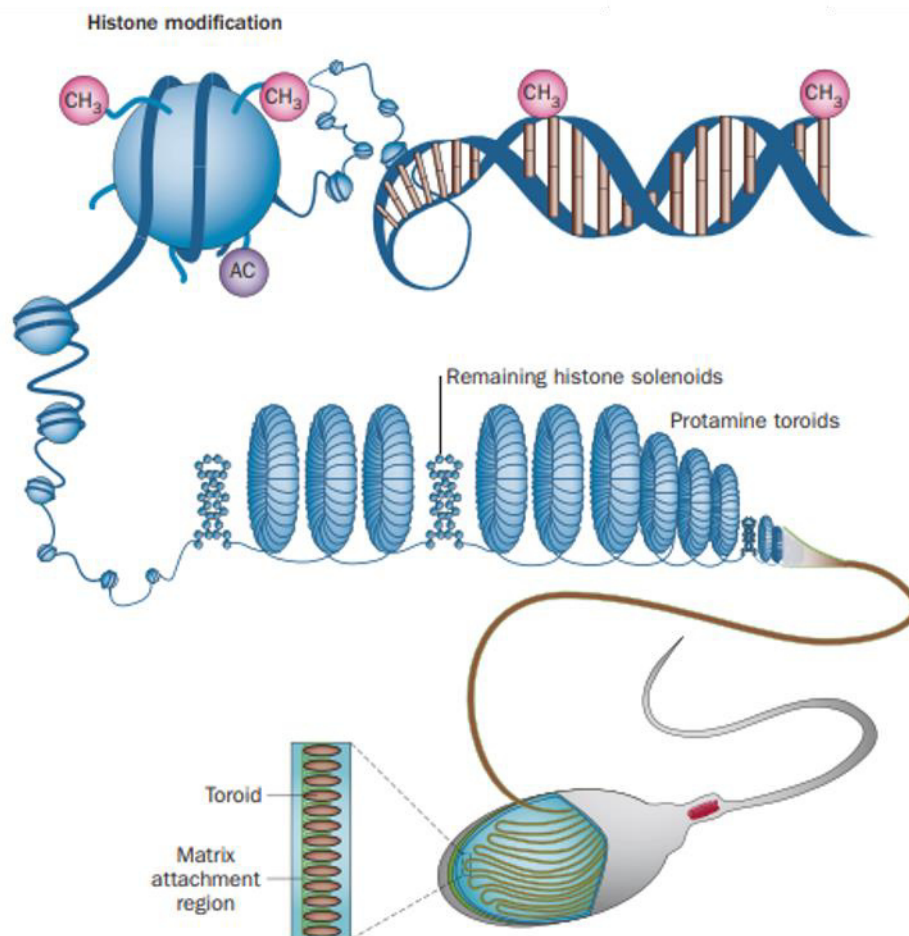


Figure 36 : Condensation de la chromatine. Au cours de la spermiogénèse, les histones sont remplacées par des protamines, condensant l'ADN en toroïdes étroitement emballés (83).

1.5.4.1. La formation de l'acrosome

La fusion des vésicules de l'appareil de Golgi, forme une vacuole acrosomiale, l'acrosome, qui se dispose sur un pôle du noyau. L'acrosome recouvre presque deux tiers de la surface du noyau. Il joue un rôle important lors de la fécondation, cette vésicule contient des enzymes lysosomiales qui permettent aux spermatozoïdes de pénétrer la zone pellucide de l'ovocyte (84).

1.5.4.2. La formation du flagelle

Il commence à se développer à partir du centriole distal du centrosome qui a migré vers le pôle postérieur de la cellule. Il assure la mobilité du spermatozoïde qui est générée par l'axonème. Il est constitué de neuf doublets de microtubules périphériques et de deux doublets internes. Les doublets externes sont également reliés entre eux par les complexes protéiques

qui forment les dynéines externes et internes. Ces dynéines permettent la mobilité du spermatozoïde en exerçant une contraction alternée (85). Le flagelle est composé de trois pièces : la pièce intermédiaire, la pièce principale et la pièce terminale.

- La pièce intermédiaire est constituée d'une gaine de mitochondries regroupées en anneaux autour de l'axonème, produisant l'énergie pour les mouvements du flagelle ;
- La pièce principale est relativement longue, comporte une gaine fibreuse dense autour de l'axonème ;
- La pièce terminale ne contient que l'axonème (80).

Le cytoplasme est éliminé par phagocytose par les cellules de Sertoli.

1.5.5. *Contrôle neuro-endocrinien de la fonction testiculaire*

La spermatogenèse dépend à la fois de l'hypothalamus, de l'hypophyse et des testicules. L'hypothalamus sécrète la GnRH (Gonadotrophine Releasing Hormone) qui agit sur l'hypophyse en sécrétant deux gonadotrophines : la FSH (Follicule Stimulating Hormone) et la LH (Luteinizing Hormone). La LH agit sur les cellules de Leydig en stimulant la production de la testostérone qui assure le maintien de la spermatogenèse. La FSH agit sur les tubes séminifères par l'intermédiaire des cellules de Sertoli pour assurer l'initiation et le maintien de la spermatogenèse. Des systèmes de rétrocontrôle négatif agissent sur le système hypothalamo-hypophysaire par le biais de la production de testostérone et d'inhibine par les testicules (Figure 37).

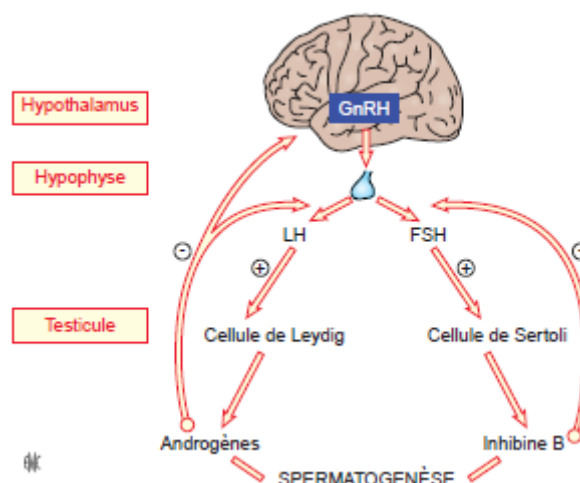


Figure 37 : Régulation hormonale de la spermatogenèse. *LH : luteinizing hormone ; FSH follicle stimulating hormone ; GnRH : gonadotrophin releasing hormone (86).*

1.5.6. *Régulation thermique*

Une hypothermie physiologique, la température des testicules doit être plus basse que la température corporelle, 33-35°C, permet un fonctionnement optimal de la spermatogenèse. Cette thermorégulation est assurée grâce au système vasculaire, qui permet par des échanges thermiques à contre-courant, un refroidissement du sang artériel arrivant au testicule et grâce au système scrotal, qui possède des récepteurs thermiques et qui va permettre les échanges entre le scrotum et le milieu extérieur. Toute élévation faible va entraîner des anomalies de la spermatogénèse et donc une altération fonctionnelle et morphologique des spermatozoïdes (87).

1.6. *L'infertilité masculine*

D'après l'Organisation Mondiale de la Santé, l'infertilité se définit par l'incapacité d'un couple à procréer après 1 an d'essais de conception sans moyen de contraception, malgré des rapports sexuels fréquents. L'infertilité est un problème majeur de santé publique affectant 15% des couples à l'échelle mondiale (88). Dans la moitié des cas, la cause est d'origine masculine caractérisée par un déficit de quantité, de mobilité, de morphologie ou de fonctionnement des spermatozoïdes. Les anomalies spermatiques sont multifactorielles, peuvent avoir des causes multiples, génétiques ou non génétiques. Comme la spermatogenèse est une succession de processus complexes qui s'effectue de manière coordonnée, la moindre altération affectant une seule de ces étapes est susceptible d'entraîner un phénotype d'infertilité.

1.6.1. *Les causes de l'infertilité masculine*

Les causes d'infertilité masculine peuvent être séparées en deux catégories : obstructives ou excrétoires correspondant à une obturation des canaux permettant l'extériorisation des spermatozoïdes malgré une production normale de spermatozoïdes par les testicules et non-obstructives ou sécrétoires, correspondant à une anomalie de production des spermatozoïdes par les testicules.

L'infertilité masculine obstructive est classée en deux groupes : congénitale et acquise. Les causes congénitales, sont dues à l'absence des canaux déférents ou des vésicules séminales, comme par exemple, l'atrésie bilatérale des canaux déférents par mutation du gène *CFTR*, ou le syndrome de Young. Les causes acquises peuvent être dues à des lésions entraînant des

obstructions au niveau l'épididyme et/ou canaux déférents (tumeurs testiculaires, séquelles interventionnelles ou inflammatoires, des rares volumineux kystes prostatiques comprimant les canaux éjaculateurs), des troubles de l'éjaculation (paraplégie, troubles sexuels, éjaculation rétrograde), des causes infectieuses qui provoquent des lésions inflammatoires de l'épididyme ou bien par l'effet de certains médicaments qui provoque des éjaculations rétrogrades.

L'infertilité masculine non obstructive peut être pré-testiculaire ou testiculaire. L'infertilité non obstructive pré-testiculaire correspond à un déficit de stimulation des testicules par les gonadotrophines (FSH, LH) entraînant un hypogonadisme hypogonadotrope (89). Elle peut être congénitale liée à une anomalie constitutionnelle, par exemple le syndrome de Kallman, ou acquise, par exemple une tumeur de la région hypothalamo-hypophysaire, un traumatisme crânien, une hyperprolactinémie ou une cause médicamenteuse (74).

L'infertilité non obstructive testiculaire correspond à une atteinte fonctionnelle du testicule, liée à une anomalie qui peut être congénitale, secondaire à des anomalies chromosomiques ou géniques, autosomiques ou gonosomiques. Ces anomalies génétiques peuvent survenir *de novo*, ou peuvent être familiales. Les causes acquises comportent la varicocèle, la cryptorchidie, des maladies infectieuses (oreillons), le cancer du testicule ou la prise de certains médicaments.

A ces deux catégories, s'ajoute la cause idiopathique qui englobe toutes les causes d'infertilité inconnue, posant la question de l'impact du mode de vie et de l'environnement.

1.6.2. *Le diagnostic de l'infertilité masculine*

L'exploration de l'infertilité masculine débute par un interrogatoire et un examen clinique qui ont pour but d'orienter l'origine de l'infertilité. L'existence d'antécédents familiaux peut orienter vers une cause génétique. Des antécédents chirurgicaux ou médicaux, expositions professionnelles à des agents toxiques seront recherchés. L'examen physique permet la recherche de varicocèle et d'examiner la prostate et les vésicules séminales.

L'examen en première intention est le spermogramme – spermocytogramme. D'autres tests spécialisés permettent d'améliorer l'évaluation de l'infertilité : un bilan hormonal (dosage de la FSH, LH, testostérone, inhibine), un dosage des marqueurs du liquide séminal, une spermoculture, l'échographie testiculaire, un bilan immunologique lors de la suspicion de la présence d'anticorps anti-spermatozoïdes, et des examens génétiques : caryotype, recherche de microdélétion du chromosome Y, recherche de la mutation au niveau du gène *CFTR*.

1.6.3. *Spermogramme et spermocytogramme*

En 1951, MacLeod étudia la relation qui pouvait exister entre la fertilité masculine et les caractéristiques du sperme (90). Depuis, le spermogramme constitue la première investigation pour explorer l'infertilité masculine. Il vise à connaître les altérations de trois paramètres : le nombre, la mobilité et la morphologie des spermatozoïdes.

L'examen macroscopique de l'éjaculat permet de qualifier l'aspect et la viscosité de l'échantillon et de mesurer le volume et le pH. Normalement, l'échantillon est d'aspect laiteux (gris opalescent). Il peut être plus translucide s'il contient peu de spermatozoïdes. Si l'échantillon contient des érythrocytes, il peut être de couleur rouge brunâtre. Il peut également avoir une coloration plus jaunâtre si le patient souffre d'ictère ou prend certains suppléments de vitamines ou médicaments. En général, le sperme se coagule dès l'émission et se liquéfie dans un délai de 30 à 60 minutes grâce aux enzymes protéolytiques d'origine prostatiques. S'il y a un problème de liquéfaction, son origine peut être prostatique. Le volume spermatique est normalement compris entre 1,5 et 6 ml pour trois jours d'abstinence sexuelle. Une aspermie désigne l'absence totale d'éjaculat ou un volume de sperme inférieur à 0,5 ml. Elle peut être due à une éjaculation rétrograde. L'hypospermie est définie par un volume entre 1,5 ml et 0,5 ml. Elle peut être due soit à une abstinence très courte, un déficit de sécrétion glandulaire, ou à un hypogonadisme (91). Une hyperspermie, définie par un volume total de l'éjaculat supérieur à 6 ml, est observée en présence de lésions infectieuses des glandes annexes, des vésicules séminales, ou bien suite à une abstinence trop longue (92). Le pH spermatique est compris entre 7,2 et 7,8.

Un pH acide, inférieur à 7,2 évoque une atteinte des vésicules séminales, tandis qu'un pH basique, supérieur à 7,8 évoque une insuffisance prostatique (86).

L'analyse microscopique permet d'évaluer :

- Numération des spermatozoïdes : La numération spermatique normale doit être supérieure à 15 millions de spermatozoïdes par ml d'éjaculat. Une azoospermie désigne l'absence totale de spermatozoïdes dans l'éjaculat. Une oligospermie correspond à une concentration spermatique inférieure à 15 millions par ml d'éjaculat. Une cryptozoospermie indique la présence de très rares spermatozoïdes, inférieurs à 100 000 dans la totalité de l'éjaculat.
- L'agglutination de spermatozoïdes : Il s'agit d'amas de spermatozoïdes liés les uns aux autres. Elle est évocatrice de la présence d'anticorps anti-spermatozoïdes, sans toutefois la confirmer (93).

- Cellules rondes : indique la présence dans l'éjaculat de spermatozoïdes immatures et de leucocytes. Un éjaculat devrait normalement en contenir moins de 1 million par ml. Une leucospermie, une concentration supérieure à 1 million par ml, peut-être la traduction d'une infection génitale, mais il faut bien sûr la confirmer (93).
- Erythrocytes : Normalement, le sperme ne devrait pas contenir d'érythrocytes. Leur présence est appelée hémospermie et révèle une infection ou un traumatisme.
- Mobilité : une mobilité est dite progressive si les spermatozoïdes se déplacent avec une vitesse peu importe et dans différentes directions. Elle est dite non progressive si les spermatozoïdes bougent, mais sans déplacement. Dans un éjaculat normal la proportion de spermatozoïdes mobiles progressifs ou non à une heure doit être supérieure à 50%, dans le cas contraire on parle d'asthénospermie.
- Vitalité : L'OMS propose de l'évaluer lorsque la motilité des spermatozoïdes progressifs est inférieure à 40%. La vitalité spermatique normale doit être égale ou supérieure à 58% de l'ensemble des spermatozoïdes. Une nécrospermie désigne l'absence totale de spermatozoïdes vivants dans l'éjaculat.

Les caractères normaux selon les critères de l'OMS, 2010 et la classification de David modifiée sont précisés dans le tableau ci-dessous :

Tableau I : Paramètres spermatiques définis selon les critères de l'OMS.

Paramètres	Normes	Définition de l'anomalie
Volume	1,5ml - 6ml	< 1,5ml = hypospermie > 6ml = hyperspermie
Numération	15 millions - 200 millions/ml ≥ 39 millions (au total)	0 = azoospermie < 15 millions/ml = oligozoospermie > 200 millions/ml = polyzoospermie
Mobilité	Mobilité progressive de type (a+b) : ≥ 32% Mobilité de type (a+b+c) : ≥ 40%	(a+b) < 32% = asthénozoospermie (a+b+c) < 40% = asthénozoospermie
Vitalité	≥ 58%	< 58% = nécrozoospermie
Leucocytes	≤ 1 million/ml	> 1 million/ml = leucospermie
Formes normales classification David modifiée	≥ 20%	< 15% = tératozoospermie

Le spermocytogramme étudie la morphologie des spermatozoïdes, et évalue le pourcentage des formes typiques. Il porte sur la forme de la tête, la pièce intermédiaire et le flagelle. Deux

principales classifications sont actuellement utilisées : la classification de David modifiée et la classification de Kruger.

- La classification de David modifiée :

La méthode de David a été proposée pour la première fois en 1972 lors d'un colloque sur la fécondité et la stérilité masculine et a été publiée en 1975. Elle a été modifiée et améliorée au début des années 1990 par l'équipe de Jouannet puis par Auger & Eustache en 2000 (94). Le spermocytogramme est normal selon cette classification si le pourcentage des formes typiques est supérieur à 50%. La classification de David modifiée comporte 15 anomalies morphologiques (figure 38) :

- Sept anomalies de la tête : têtes allongées, amincies, microcéphales, macrocéphales, multiples, base anormale et acrosome anormal ou absent.
- Trois anomalies de la pièce intermédiaire : reste cytoplasmique, pièce intermédiaire amincie et pièce intermédiaire angulée.
- Cinq anomalies du flagelle : absent, écourté, de calibre irrégulier, enroulé et multiple.

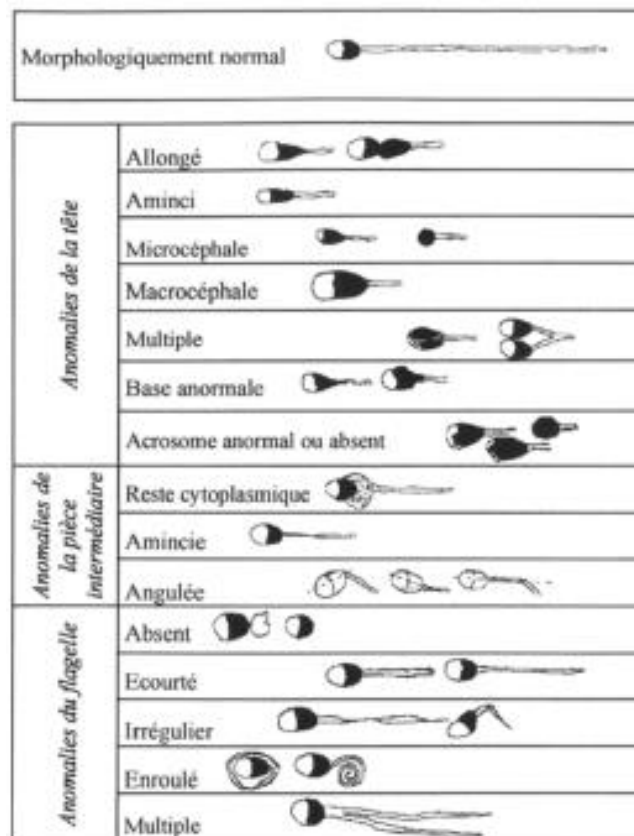


Figure 38 : Types d'anomalies selon la classification de David modifiée (94).

Elle permet également le calcul d'un facteur prédictif d'infertilité, c'est l'index d'anomalies multiples (IAM). L'IAM est un indicateur du nombre moyen d'anomalies associées par

spermatozoïde anormal et correspond au rapport du nombre total d'anomalies recensées au nombre total de spermatozoïdes anormaux. Sa valeur est normalement inférieure à 1,6 (95).

- La Classification de Kruger :

La classification de Kruger est décrite comme étant plus stricts que les autres classifications avec une meilleure valeur prédictive positive sur les résultats de fécondation *in vitro* (FIV) (95). Elle recense 4 classes d'anomalies par ordre d'importance (Figure 39): les anomalies concernant l'acrosome et la tête, celles du cou et de la pièce intermédiaire, celles du flagelle et la présence d'un cytoplasme non expulsé. Il suffit d'une anomalie pour que le spermatozoïde soit classé en forme atypique. L'absence totale d'anomalie définit le spermatozoïde typique (96). Cependant un spermogramme anormal nécessite toujours un deuxième prélèvement de contrôle trois mois plus tard, la durée d'un nouveau cycle de spermatogenèse.

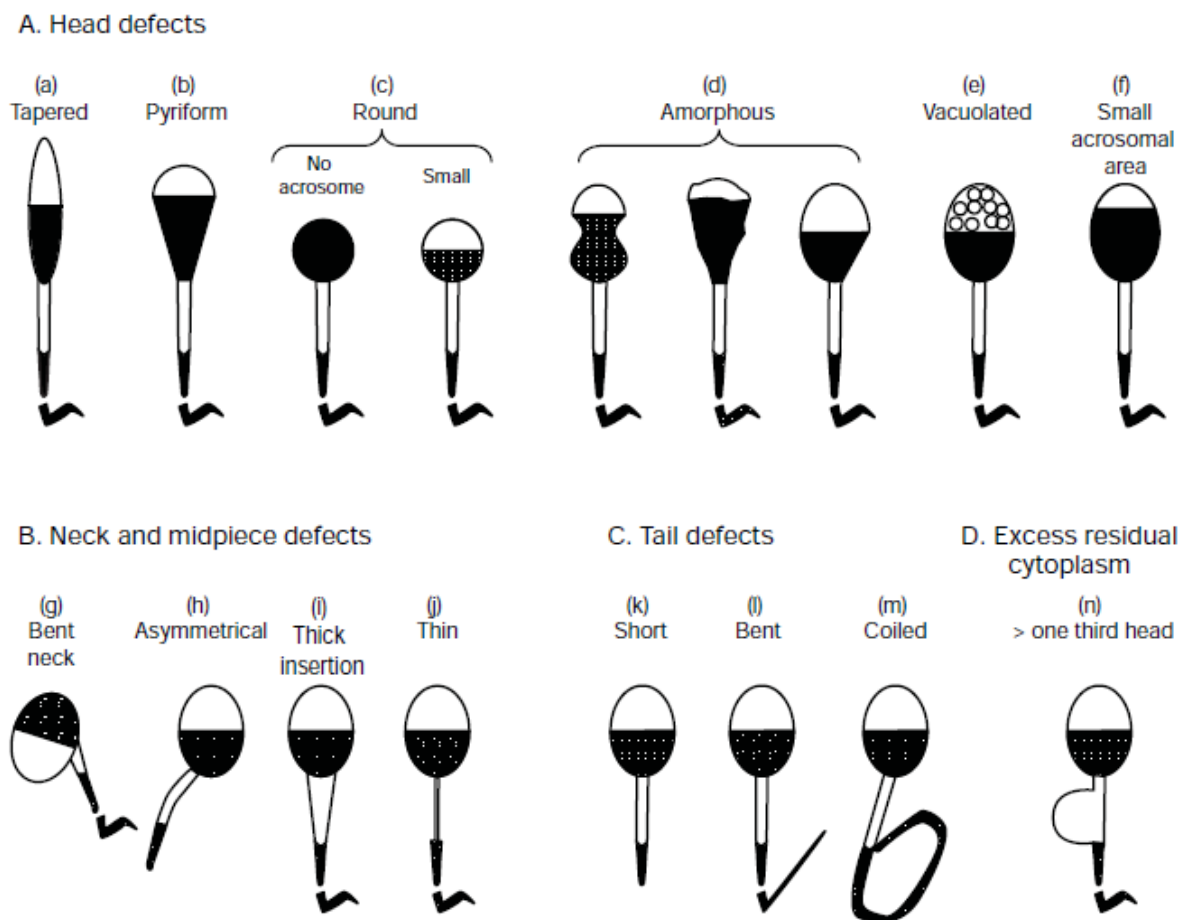


Figure 39 : Les anomalies morphologiques des spermatozoïdes selon la classification de Kruger (A) Anomalie de la tête : a- Allongée b- en forme de poire c- tête ronde : absence ou la présence d'un petit acrosome d- Amorphe e- vacuolée f- petite zone acrosomale (B) défaut du cou et de la pièce intermédiaire : g- cou plié h- insertion asymétrique i- épais j- mince (C)

défauts du flagelle : *k*-court *l*-plié *m*-enroulé (**D**) excès de résidus cytoplasmique *n*- la taille de l'excès de résidus cytoplasmiques est supérieur au 1/3 de la taille tête (93).

1.6.4. **Génétique de l'infertilité masculine**

Les causes génétiques représentent 15% des cas d'infertilité masculine (97). Ces anomalies génétiques peuvent être chromosomiques, anomalies de nombre ou de structure, impliquant les gonosomes ou les autosomes, des CNVs, des mutations géniques. Trois arguments sont en faveur de l'implication d'une composante génétique dans l'infertilité :

- Dans la population générale, la fréquence des anomalies chromosomiques est estimée entre 0,8 et 1%. Tandis qu'elle est de 15% chez les patients azoospermiques et 2% chez les patients oligospermiques (98). Ces réarrangements pourraient être responsables d'un défaut de la méiose, mais l'interruption d'un gène impliqué dans la spermatogenèse pourrait aussi jouer un rôle.
- La présence de cas familiaux d'infertilité ou d'individus infertiles issus d'un mariage consanguin renforce l'hypothèse de l'implication de gènes récessifs dans la spermatogenèse.
- La mutation génique dans un modèle animal qui se traduit phénotypiquement en une infertilité fortifie l'hypothèse de l'implication de ce gène dans l'infertilité (86).

1.6.4.1. **Les anomalies chromosomiques**

De nombreuses études ont rapporté l'association d'anomalies chromosomiques à des troubles de la spermatogenèse et à l'infertilité masculine.

1.6.4.2. **Anomalies chromosomiques de nombre**

Syndrome de Klinefelter

En 1942, Klinefelter décrit pour la première fois le syndrome qui porte aujourd'hui son nom l'observation de neuf sujets masculins infertiles âgés de 17 à 33 ans. En 1957, Ferguson-Smith *et al* a identifié des corps de Barr chez 10 des 91 hommes étudiés pour une azoospermie ou une oligozoospermie sévère et il a suspecté l'existence d'anomalies chromosomiques chez les patients présentant une infertilité masculine (99). Deux ans plus tard, Jacobs et Strong ont identifié la présence d'un chromosome X surnuméraire chez les hommes présentant le syndrome de Klinefelter (100). Ce syndrome représente plus de 60% des anomalies de nombre touchant les gonosomes (101). Sa prévalence est de 11% chez les azoospermiques (102) et 5%

chez les hommes oligozoospermiques sévères (103). Les signes cliniques sont très variables d'un patient à l'autre. Avant la puberté, les patients n'ont pas de problème particulier. Il peut être observé plus fréquemment un décalage des acquisitions motrices, des difficultés pour l'apprentissage du langage et de la lecture un retard de développement moteur, une faiblesse musculaire. A l'adolescence, les patients atteints du syndrome de Klinefelter ont une taille plus grande que la moyenne, les jambes et les bras sont plus longs (macroskelie), le torse plus court et les hanches sont plus larges par rapport aux autres enfants. Après la puberté, ces patients ont des petits testicules fermes, une gynécomastie et une diminution des poils du visage et du corps. Du point de vue neuropsychologique, ces patients possèdent un quotient intellectuel dans les limites de la normale et sont décrits comme plutôt timides et immatures.

Le syndrome de Klinefelter provoque généralement l'arrêt de la spermatogenèse au stade de pachytène. La présence d'un chromosome X surnuméraire dans la vésicule sexuelle perturberait et entraînerait une détérioration de la cellule (104). Dans les cas de mosaïque, ces patients produisent un nombre variable de spermatozoïdes présents dans une biopsie testiculaire, mais rarement dans l'éjaculat (103).

Syndrome 47,XYY

Il a été décrit plus la première fois par Sandberg en 1961 (105). Ce syndrome est secondaire, comme pour toute aneuploïdie, par une non-disjonction chromosomique lors de la méiose II, mais seulement chez l'homme puisqu'intéressant le chromosome Y. Les cas de mosaïque sont dus à une non-disjonction mitotique au début du développement embryonnaire. Le plus souvent, cette formule gonosomique n'a pas de conséquences phénotypiques. Une grande taille peut être observée, et dans de rares cas des difficultés d'apprentissage sont rapportées. En général, les cellules germinales XYY sont capables d'échapper au contrôle du stade pachytène et continuer la méiose. Ainsi, les patients sont souvent fertiles avec une perturbation au niveau des paramètres spermatiques plus ou moins sévères. Cette dysgonosomie est néanmoins quatre fois plus fréquente chez les sujets infertiles que dans la population générale (106).

1.6.4.3. Anomalies chromosomiques de structures

Les translocations robertsoniennes

C'est l'anomalie chromosomique de structure la plus fréquente chez les hommes infertiles, en particuliers chez les oligospermiques, sa fréquence est 9 fois supérieure à celle des nouveau-nés (107).

Les translocations réciproques équilibrées

Cette anomalie fréquemment est trouvée chez des hommes azospermiques (1%) ou oligospermiques (0.6%). La fréquence de ce réarrangement chromosomique chez les hommes infertiles est de 5 à 6 fois plus élevée que chez les nouveau-nés (107).

Les autres anomalies de structures, les inversion paracentrique et péricentrique, les réarrangements chromosomiques complexes et les marqueurs chromosomiques peuvent altérer la synapsis chromosomique provoquant l'arrêt de la spermatogenèse. Très peu d'études de ces cas associés à des infertilités ont été rapportées dans la littérature. Ainsi, il est difficile d'estimer la proportion de ces anomalies associées à des troubles du spermogramme.

Homme 46,XX

Dans la grande majorité de ces cas, le gène *SRY* est présent, suite à un crossing over inégal, sur le chromosome X ou, plus rarement, avec un autre chromosome. On parle du syndrome du mâle XX ou syndrome de la Chapelle. Dans quelques rares cas, si le gène *SRY* est absent, ce syndrome résulte de l'atteinte d'un autre gène, impliqué dans la cascade du déterminisme sexuel entraînant la formation du testicule (*SOX9*, *SOX3*, *WNT4*...). Ces patients présentent des signes cliniques proches de ceux présentés par les patients ayant un syndrome de Klinefelter à l'exception de la dysmorphologie (108).

Le gène *SOX9*, est un facteur de transcription, essentiel pour le développement sexuel et squelettique. Ce gène peut être responsable de la dysplasie campomélique. Il est localisé au niveau du bras long du chromosome 17, dans un désert génique où il y a l'absence de séquence codante. Une variation génique ponctuelle au niveau de ce gène ainsi qu'une délétion ou une duplication en amont (109) ou bien une translocation impliquant cette région peut être responsable de la maladie. C'est un exemple de maladie par effet de position. Par exemple il a été rapporté dans une famille trois filles avec une anomalie du déterminisme sexuel et une délétion en amont du gène *SOX9* de 349 Kb (110). Des microdélétions localisées à 1,5 Mb de *SOX9* et des points de cassure due à des translocations localisées à 1 et 1,2 Mb de *SOX9* ont été rapportées comme responsable d'une séquence de Pierre Robin chez des patients (111).

1.6.4.4. Microremaniements cryptiques et infertilité

A ce jour, à l'exception de la microdélétion du chromosome Y, de rares CNV ont été décrits comme responsables d'infertilités masculines. Krausz *et al.* (2012) a étudié spécifiquement les CNV liés au chromosome X, qui contiennent un nombre élevé de gènes impliqués dans la spermatogenèse. Sur une cohorte de 96 hommes infertiles et 103 hommes fertiles, plusieurs CNV et gènes candidats ont pu être identifiés (112). Une autre étude a pu caractériser un CNV englobant le gène *E2F1*, localisé sur le chromosome 20q11.22. Ce CNV présent en gain ou perte de copie est impliqué dans l'infertilité. Des études sur le modèle murin ont révélé que une sous ou sur expression de la protéine E2F1, responsable d'une atrophie testiculaire causée par une anomalie de la spermatogenèse (113). Une délétion à l'état homozygote du gène *DPY19L2*, provoque une globozoospermie. Cette délétion est due à un mécanisme de NAHR, due à la présence de deux régions répétées homologues de 28 kb, *LCR*, localisées aux deux extrémités de ce gène (114).

Le chromosome Y est l'un des petits chromosomes atteignant une taille d'environ 53 Mb. Sur les parties distales sont localisées les régions pseudo-autosomiques communes avec le chromosome X. Dans la partie distale du bras long, se trouve une zone d'hétérochromatine constitutive de taille variable. Entre la région PAR1 et la région d'hétérochromatine s'étale une région spécifique au chromosome Y, appelée MSY (Male Specific Y chromosome).

En 1976, Tiepolo et Zuffardi ont mis en évidence une association entre la délétion du bras long du chromosome Y et l'azoospermie (115). Cette région délétée a été nommée AZF (AZoospermia Factor). Les progrès des outils moléculaires ont permis de mieux caractériser cette région est identifier trois loci, AZFa, AZFb et AZFc, responsable d'infertilité masculine. Ces loci comportent de nombreuses régions répétées (éléments rétroviraux, séquences palindromiques de répétitions en tandem, plusieurs gènes en multiples copies) favorisant ainsi la survenue de microdélétions récurrentes par le mécanisme de NAHR. Ces microdélétions représentent une cause génétique fréquente impliquée dans l'azoospermie et l'oligospermie sévère (Figure 40) (86).

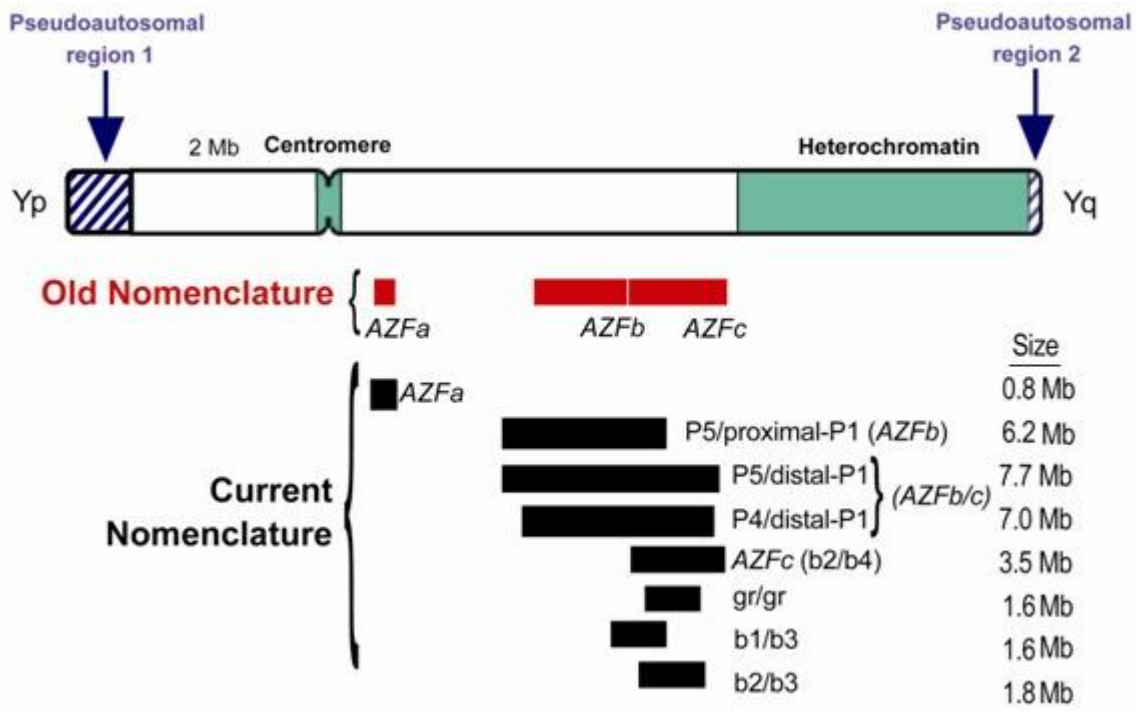


Figure 40 : Schéma du chromosome Y indiquant la position des régions précédemment définies AZFa, AZFb et AZFc et la position des délétions récurrentes actuellement rapportées basées sur les répétitions palindromiques. Yp bras court, Yq bras long (116).

La microdélétion du locus AZFa entraîne une azospermie avec l'absence de cellules germinales (SCOS : sertoli cell-only syndrome).

La microdélétion du locus AZFb provoque une azospermie avec un arrêt de la spermatogenèse.

La microdélétion d'AZFc provoque une atteinte variable de la spermatogenèse avec une traduction clinique qui varie d'une azospermie à une oligospermie (117).

1.6.4.5. Gènes et infertilité masculine

De nombreuses variations géniques associées à des syndromes pathologiques peuvent être également liées à une infertilité. Par exemple, le syndrome de Kallmann, due à des défauts de la sécrétion ou de l'action de GnRH entraînant un hypogonadisme chez l'homme (74).

Les mutations au niveau du gène *CFTR*, localisé sur le chromosome 7q31.2 responsables d'une atrésie bilatérale des canaux déférents (ABCD) comme mentionné auparavant(118). Dans

certains cas, des mutations du gène *CFTR* ont été trouvées dans des formes isolées d'ABCD (119).

La dyskinésie ciliaire primaire, maladie autosomique récessive présentant une hétérogénéité génique, est responsable d'une maladie respiratoire primitive et d'une anomalie de la mobilité des spermatozoïdes (120).

Les cas familiaux dont plusieurs frères présentent le même phénotype d'infertilité a permis l'identification des gènes impliqués dans l'infertilité masculine non syndromique. Les nouvelles technologies dont le séquençage à haut débit a permis d'identifier plusieurs gènes. Le tableau II résume les gènes identifiés chez des patients ayant une infertilité masculine non syndromique.

Tableau II : Liste des gènes identifiés chez des patients ayant une infertilité masculine non syndromique.

Phénotype	Gène	Locus	Mode de transmission	OMIM	Référence (PMID)
Teratozoospermie	ZPBP	7p12.2	AR	608498	21911476
Globozoospermie	DPY19L2	12q14.2	AR	613893	30333325
	SPATA16	3q26.31	AR	609856	17847006
Macrozoospermie	AURKC	19q13.43	AR	603495	28187507
Spermatozoïde acéphalique	BRDT	1p22.1	AR	602144	28199965
	PMFBP1	16q22.2	AR	618085	30298696
	SUN5	20q11.21	AR	613942	29331481
Spermatozoïde à petite tête	RNF220	1p34.1	AR	616136	30500349
Anomalies morphologiques multiples des flagelles de sperme	AK7	14q32.2	AR	615364	29365104
	AKAP4	Xp11.22	XL	300185	21255775
	CEP135	4q12	AR	611423	28866084
	CFAP43	10q25.1	AR	617558	30251428
	CFAP44	3q13.2	AR	617559	30251428
	CFAP69	7q21.13	AR	617949	30415212
	DNAH1	3p21.1	AR	603332	29449551
	FSIP2	2q32.1	AR	615796	30137358
	WDR66	12q24.31	AR	618146	30310178
	Asthénozoospermie	CATSPER1	11q13.1	AR	606389
GALNTL5		7q36.1	AD	615133	24398516
MNS1		15q21.3	AR	610766	30148830
SPAG17		1p12	AR	616554	28548327
PLK4		4q28.1	AD	605031	26452337
Azoospermie	PSMC3IP	17q21.2	AR	608665	29240891
	RBMXL2	11p15	AD	605444	14996998
Azoospermie ou oligozoospermie	REC8	14q12	AD	608193	18570052
	SOX8	16p13.3	AD	605923	29373757
	SYCP2	20q13.33	AR	604105	31866047
Oligozoospermie	USP26	Xq26.2	XL	300309	27726449
	FANCM	14q21.2	AR	609644	30075111
	FATE1	Xq28	XL	300450	12811541
	HAUS7	Xq28	XL	300540	29017965
	KDM3A	2p11.2	AD	611512	30053768
	KLHL10	17q21.2	AD	608778	17047026
	UBE2B	5q31.1	AR	179095	23378580
Oligoasthénoteratozoospermie	TDRD6	6p12.3	AR	611200	29551503
	AR	Xq12	XL	313700	30251955
Azoospermie non obstructive	CCDC155	19q13.33	AR	618125	29790874
	DMC1	22q13.1	AR	602721	29331980
	DMRT1	9p24.3	AD	602424	24934491
	DNMT1	19p13.2	AD	126375	25739334
	DNMT3B	20q11.21	AD	602900	25739334
	E2F1	20q11.22	AD	189971	25439843
	GTF2H3	12q24.31	AR	601750	24398516
	HSF2	6q22.31	AD	140581	23064888
	MAGEB4	Xp21.2	XL	300153	28401488
	MEI1	22q13.2	AR	608797	29659827
	MEIOB	16p13.3	AR	617670	28206990
	NANOS2	19q13.32	AR	608228	29790874
	SPINK2	4q12	AR	605753	28554943
	SPO11	20q13.31	AR	605114	29790874
	STX2	12q24.33	AR	132350	29570232
	SYCE1	10q26.3	AR	611486	25899990
	SYCP3	12q23.2	AD	604759	22670862
	TAF4B	18q11.1	AR	601689	24431330
	TDRD9	14q32.33	AR	617963	28536242
	TEX11	Xp11	XL	300311	29661171
	TEX14	17q22	AR	605792	29790874
	TEX15	8p12	AR	605795	28355598
	UBR2	6p21.1	AD	609134	25739334
	XRCC2	7q36.1	AR	600375	30489636
	ZMYND15	17p13.2	AR	614312	24431330

1.7. Les anomalies du développement humain

1.7.1. Introduction

Les anomalies du développement résultent de la morphogenèse anormale d'un tissu ou d'un organe aboutissant à une anomalie. Ces anomalies qui peuvent être structurelles ou fonctionnelles sont désignées congénitales ou constitutionnelles, car elles sont présentes à la naissance. Ces anomalies peuvent être décelables avant la naissance, à la naissance ou bien plus tard dans la vie. Elles peuvent être majeures et avoir des conséquences médicales, sociales ou esthétiques graves pour la personne concernée et elles peuvent être même une cause de surmortalité. Certaines sont mineures, avec des conséquences limitées, même parfois méconnues, et leur prise en charge est simple. Ces anomalies congénitales dérivent de quatre mécanismes de défaut de morphogenèse : malformation, déformation, dysplasie et disruption.

La malformation congénitale est une absence ou la formation d'une manière anormale ou incomplète d'un organe ou d'un tissu suite à un processus anormal intrinsèque (122). Ce processus intrinsèque est le plus souvent dû à un facteur génétique. Ces malformations sont appelées encore primaires ou vraies.

Les autres mécanismes sont appelés secondaires et sont dus à des facteurs extrinsèques, qui perturbent le processus normal du développement. La disruption est secondaire à un phénomène extrinsèque qui interfère avec, voire arrête, le processus de développement. La déformation, secondaire à une force mécanique, modifie les formes des organes développés. La dysplasie désigne une anomalie de l'histogenèse ou d'un tissu, localisée ou généralisée (122,123). Ces mécanismes sont résumés dans la figure 42.

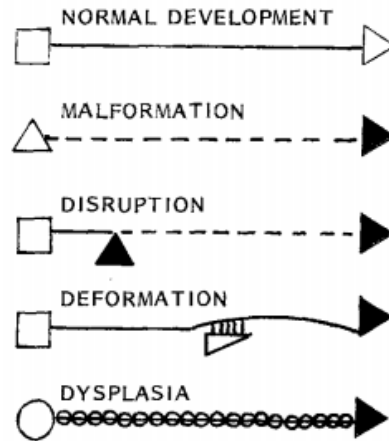


Figure 42 : Les différents mécanismes provoquant une anomalie de morphogénèse (123).

Ces malformations peuvent être localisées ou multiples. La majorité des malformations congénitales majeures sont isolées. L'association de deux ou plusieurs anomalies congénitales majeures sont dénommées poly malformations. Ces poly malformations peuvent être classées en séquence, association ou syndrome (Figure 43).

Une séquence malformative présente une succession d'anomalies, en cascade, qui résultent d'une seule cause ou un facteur (122). Par exemple la séquence de Pierre-Robin se caractérise par la présence de trois anomalies bucco-faciales qui sont les conséquences les unes des autres : une hypoplasie de la mâchoire inférieure, un rétrognatisme qui a pour conséquence que la langue chute en arrière, provoquant un obstacle à la fermeture du palais responsable d'une fente palatine (124). Une association malformative est la survenue de malformations sans lien fonctionnel entre eux, plus fréquente que le voudrait le hasard. L'association VACTERL (Vertebral anomalies, Anal atresia, Cardiac malformations, Tracheo Esophageal fistula, Renal anomalies and Limb defects) en est un exemple. Un syndrome malformatif est défini par une combinaison spécifique de plusieurs anomalies secondaires à une seule étiologie, comme par exemple, le syndrome de Down (trisomie 21).

Dans certains cas, la présence de deux ou plusieurs malformations non liées à une séquence ou un syndrome, ou bien présente par hasard et n'appartenant pas à une association connue sont classées comme anomalie congénitale multiple.

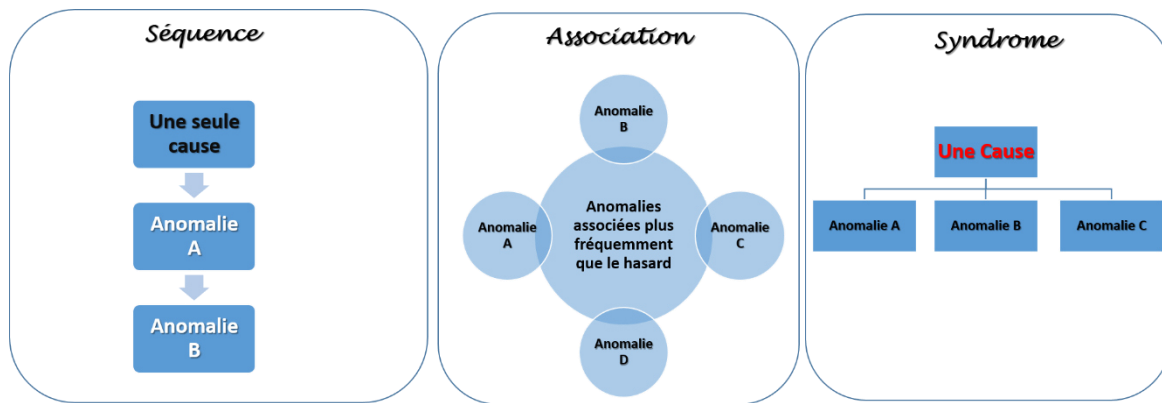


Figure 43 : Les trois classes de polymalformations : Séquence, Association et Syndrome.

1.7.2. *Les étiologies des anomalies congénitales*

D'après le rapport de l'Organisation Mondiale de la Santé, trois millions de fœtus et de nouveau-nés présentent une anomalie congénitale majeure chaque année. Des enquêtes ont montré que la prévalence des anomalies congénitales à la naissance varie d'un pays à un autre. Cette variation est due à des influences sociales (par exemple le taux de mariages consanguins), écologiques et économiques (125).

Les anomalies congénitales les plus courantes sont : les malformations cardiaques congénitales, les anomalies du tube neural et le syndrome de Down. Dans 40 à 60% des cas, l'étiologie des anomalies congénitales est inconnue. L'étiologie génétique correspond 30 à 40% des cas. Parmi les facteurs de risque d'étiologie génétique ont été rapportés l'augmentation de l'âge maternel : augmentation des non-disjonctions chromosomiques et l'augmentation de l'âge paternel : augmentation des mutations *de novo*. Les facteurs environnementaux (les agents tératogènes tels que les agents infectieux, des médicaments ou carences en certains nutriments, exposition à des rayonnements...) expliqueraient 5 à 10% des cas (126).

1.7.3. *Les causes génétiques*

Les causes génétiques impliquées dans les anomalies congénitales peuvent être chromosomiques (anomalies de nombre ou de structure), des CNV, géniques (mutations *de novo*, récessive, dominant, autosomique, liée au chromosome X), des anomalies liées à l'empreinte parentale ou des mutations dans l'ADN mitochondrial.

1.7.3.1. Les syndromes microdélétionnels et microduplicationnels

Plusieurs microdélétions ou microduplications récurrentes (mécanisme de NAHR) sont associées à un phénotype spécifique chez les patients atteints (figure 44). Par exemple, la microdélétion récurrente au niveau du locus 7q11.23 provoquant le syndrome de Williams-Beuren impliquant le gène *ELN* et la microduplication au même locus est responsable du syndrome de microduplication 7q11.23 (127).

De nombreuses autres microdélétions ou microduplications non récurrentes ont été décrites (Figure 44). Les points de cassure peuvent survenir dans n'importe quelle région du génome. Le phénotype est dépendant du contenu génique de la région impliquée. Même si les points de cassure ne sont pas identiques, les patients peuvent présenter le même phénotype du fait de l'altération d'une région minimale critique commune.

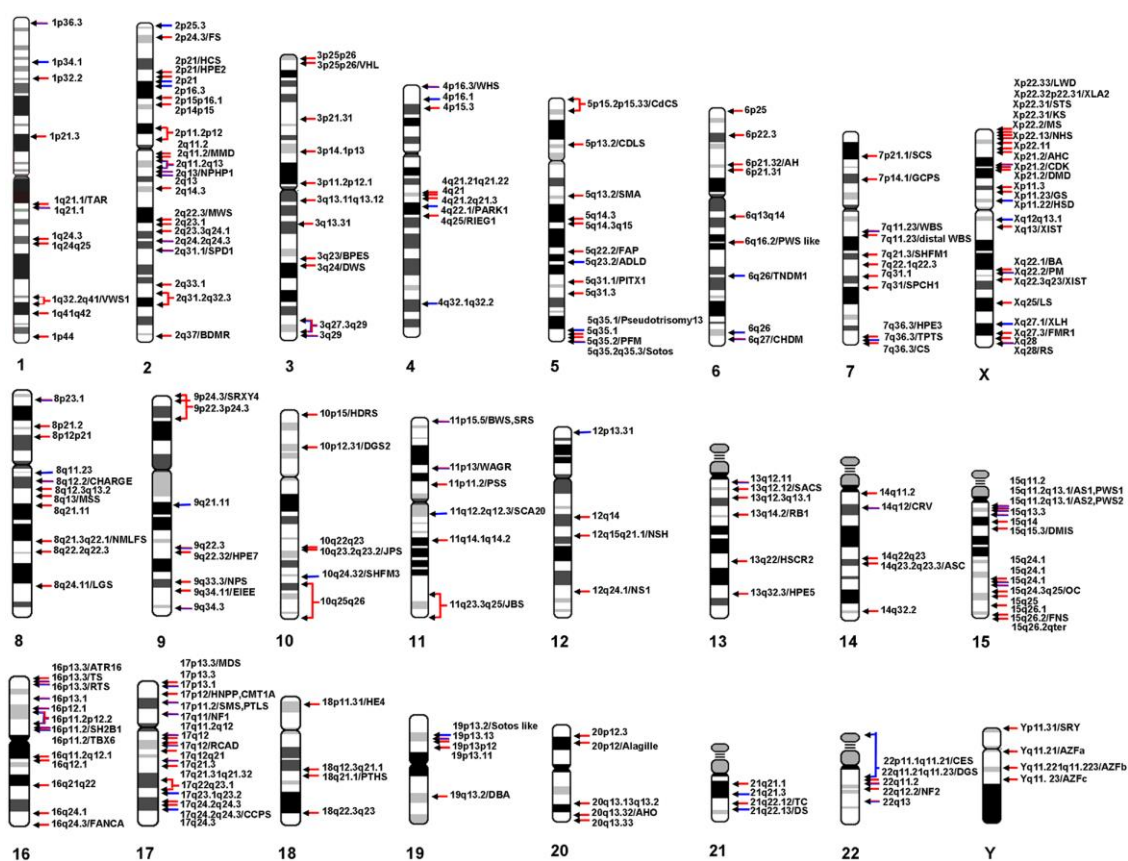


Figure 44 : Vue d'ensemble de la majorité des microdélétions et microduplications génomiques citées au moins deux fois dans la littérature (année 2012). Les flèches rouges indiquent une microdélétion, les flèches bleues indiquent une microduplication (128).

1.7.3.2. Les causes géniques

Une variation nucléotidique est définie comme une variation de la séquence d'ADN d'une personne par rapport à une séquence dite de référence, c'est-à-dire une séquence consensuelle du génome humain obtenue par compilation de la séquence de plusieurs individus réputés sains. Une variation nucléotidique pathogène dans un gène sous forme d'une délétion, insertion, duplication et substitution peut provoquer une anomalie congénitale. Elle peut être récessive, dominante, autosomique ou lié à l'X. À noter que certaines maladies présentent une grande hétérogénéité génique et génétique. Jusqu'à présent plus que 7000 maladies monogéniques rares distinctes sont impliquées dans les anomalies congénitales (129). Depuis une dizaine d'années, ce nombre augmente d'une façon exceptionnelle, date de l'utilisation du séquençage massif en parallèle.

1.7.3.3. Les facteurs extrinsèques

Les facteurs extrinsèques constituent 5 à 10 % des anomalies congénitales. Ils peuvent être d'origine maternelle ou bien être extérieurs à la mère. Des maladies chroniques de la mère, comme le diabète, l'hypertension artérielle, la phénylcétonurie, l'addiction à l'alcool (responsable du syndrome d'alcoolisme fœtal), au tabac et aux drogues peuvent être l'origine d'anomalies congénitales. Certaines professions maternelles, pour lesquels il y a exposition à des radiations ou produits radioactifs, peuvent être à l'origine d'anomalies congénitales. La carence en certains nutriments et vitamines telle que la vitamine B9 ou l'acide folique sont également responsables d'anomalies congénitales.

Une infection virale (cytomégalovirus, rubéole, varicelle...), parasitaire (toxoplasmose), bactérienne (syphilis) durant la grossesse peut être à l'origine d'anomalies congénitales (130,131). L'hyperthermie maternelle augmente aussi le risque d'anomalies congénitales. Certains médicaments ou substances comme le plomb, mercure, arsenic... peuvent entraîner l'apparition d'anomalies congénitales (132)(133).

Une force mécanique exercée sur le fœtus, comme un oligoamnios (diminution du liquide amniotique) peut entraîner des anomalies congénitales (séquence de Potter).

1.8. Apport du NGS dans l'étude des remaniements chromosomiques

1.8.1. Bref historique du séquençage

Le séquençage est une méthode qui détermine l'ordre exact dans lequel se succèdent les nucléotides (Adénine, Thymine, Cytosine, Guanine) dans une molécule d'ADN donnée. Les premières techniques de séquençage ont été développées par les Américains Allan Maxam et Walter Gilbert et l'anglais Frederick Sanger. Ils ont mis au point deux méthodes très différentes permettant d'accéder à la lecture de la séquence de l'ADN. Maxam et Gilbert ont exploité des stratégies de dégradation chimique sélective. Quant à Sanger, il a choisi une stratégie de synthèse enzymatique sélective. La méthode de Sanger (Figure 45) a rapidement dépassé la méthode de Maxam-Gilbert. Elle est considérée comme la première génération de séquençage. Celle-ci se base sur l'utilisation d'une ADN polymérase qui va incorporer des nucléotides terminateurs de chaînes marqués (ddNTP). La séparation électrophorétique par taille croissante des fragments obtenus permet de décoder, grâce à 4 marquages fluorescents différents correspondant aux 4 nucléotides, la séquence du fragment d'ADN étudié (Figure 45). En général, cette technique permet d'obtenir des séquences de longueur comprise entre 400 et 850 pb ce qui présente une certaine limite, comme par exemple dans le cas de maladies qui présentent une hétérogénéité génique ou dans le cas des gènes de grandes de tailles (134).

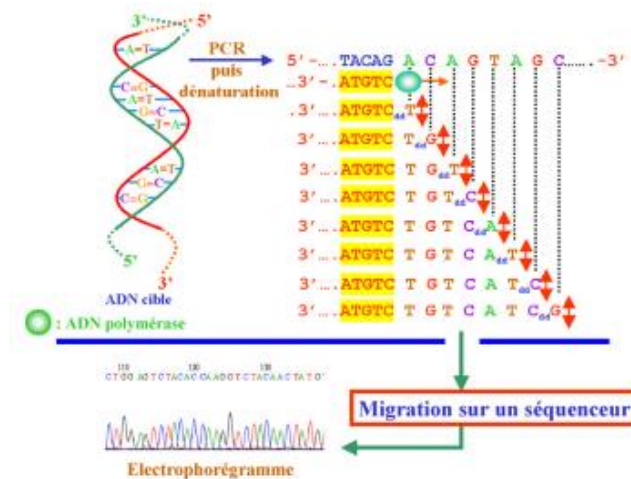


Figure 45 : Principe du séquençage Sanger. Après avoir amplifié et purifié une région spécifique de l'ADN délimité par deux amorces, une réaction de séquence est réalisée. Son principe est basé comme une PCR asymétrique avec l'utilisation uniquement d'une seule amorce et l'utilisation de dNTP et de ddNTP marqué chacun par un fluorophore. Lorsqu'un ddNTP est incorporé, l'extension s'arrête. Ensuite, après une purification, cette réaction de séquence est migrée sur le séquenceur par électrophorèse capillaire et l'électrophorégramme permet d'obtenir la séquence d'ADN (134).

Le séquençage de nouvelle génération (NGS), que l'on appelle aussi séquençage massif en parallèle de deuxième génération ou séquençage à haut débit, a été mis au point en 2005 (134). Sa caractéristique principale est sa capacité à produire une quantité très importante de données génétiques à un coût raisonnable. Les principales technologies de NGS actuellement sont (Figure 46) :

- Le pyroséquençage : La spécificité de cette technologie repose sur une PCR en émulsion pour l'amplification des fragments à séquencer : la PCR a lieu dans une microgoutte renfermant une microbille d'agarose en phase aqueuse, séparée des autres billes (plusieurs millions) par de l'huile. On obtient ainsi des copies d'un seul fragment d'ADN par bille. Chacune des billes est ensuite déposée dans un des 1,6 million de puits d'un support solide appelé PTP (PicoTiter Plate). Des réactions de pyroséquençage ont alors lieu dans chaque puits, basé sur l'émission d'un signal photoluminescent secondaire à la libération d'un pyrophosphate lors de l'incorporation de nucléotides (technologie développée en particulier par la société Roche) (135).

- Le séquençage par nano-mesure de pH : Elle repose sur la libération d'un ion H⁺ à chaque intégration d'un nucléotide au cours de l'élongation. Ce phénomène crée une modification de

pH pouvant être détectée par une puce de silicium semi-conductrice composée de plusieurs millions de transistors (technologie développée par la société Thermo-Fisher-Ion torrent) (136).

- Le séquençage par chimie de terminateurs réversibles : basé sur l'incorporation de nucléotides modifiés fluorescents (technologie développée en particulier par la société Illumina).

Concernant ces différentes technologies, le principe général reste le même malgré des spécificités propres à chacune. Une librairie est préparée qui consiste à la fragmentation de l'ADN. Puis des adaptateurs sont liés, ils servent d'amorces universelles pour l'amplification clonale des fragments à séquencer. Lors du séquençage, à chaque incorporation d'un nucléotide, un signal est détecté (un signal lumineux dans la technologie d'Illumina, une variation du pH dans la technologie Thermo-Fisher-Ion Torrent).

La technologie utilisée pour cette étude est la technologie Illumina de séquençage par synthèse qui est actuellement la technologie de NGS dominante sur le marché. Nous expliquerons les différentes étapes de séquençage en prenant comme exemple cette technologie.

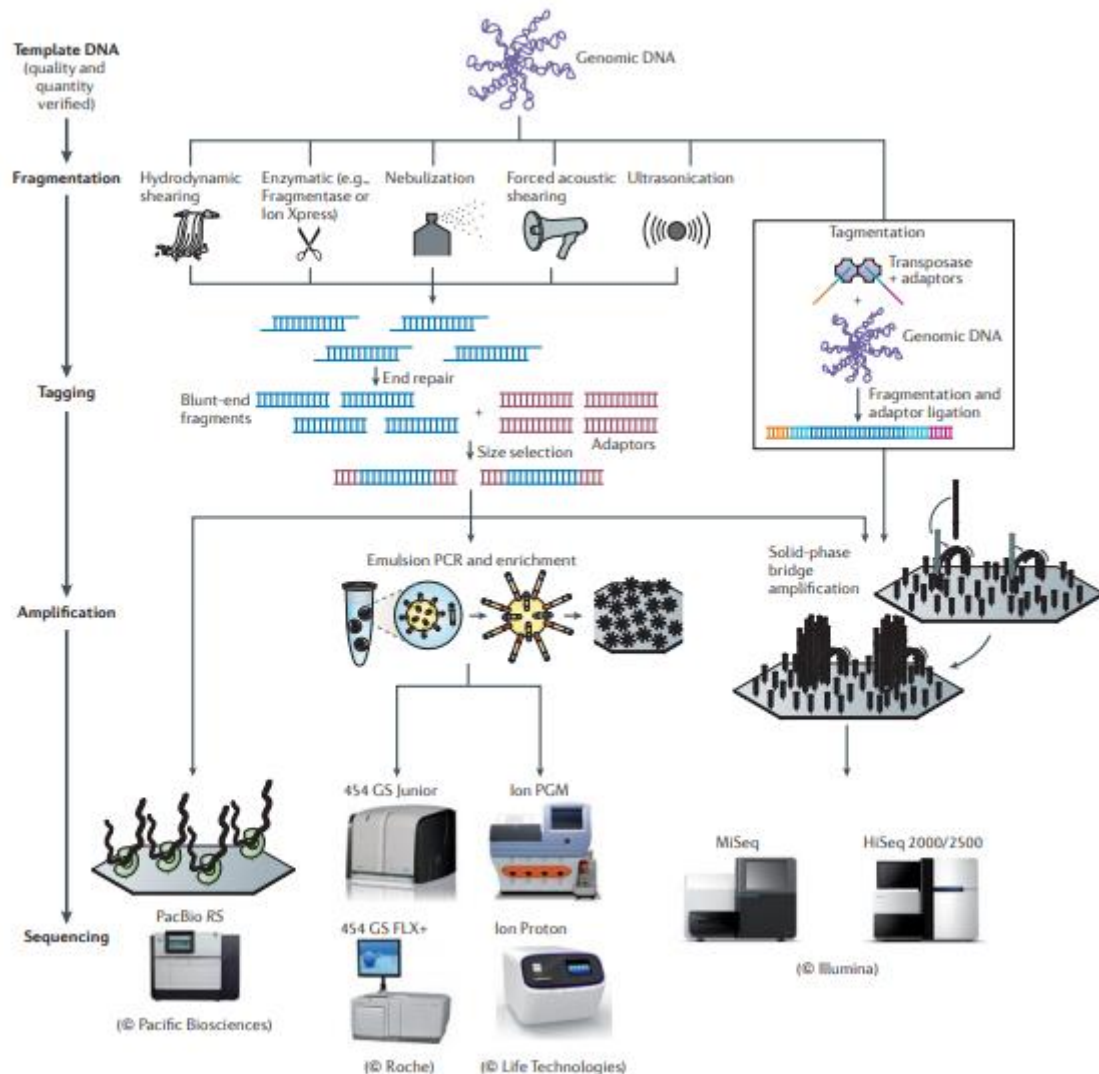


Figure 46 : Les principales technologies de séquençage à haut débit. *Après l'extraction de l'ADN, une fragmentation mécanique ou enzymatique est réalisée suivie de l'addition des adaptateurs puis une amplification par émulsion ou sur un support solide sera réalisée. Enfin une étape de séquençage sera effectuée propre à chaque technologie (137).*

1.8.2. Principe et technique du séquençage du génome entier de la technologie Illumina

Le principe du séquençage du génome entier (WGS) et du séquençage de l'exome (WES) est identique. La principale différence se situe au niveau de la préparation de la librairie (Figure 47).

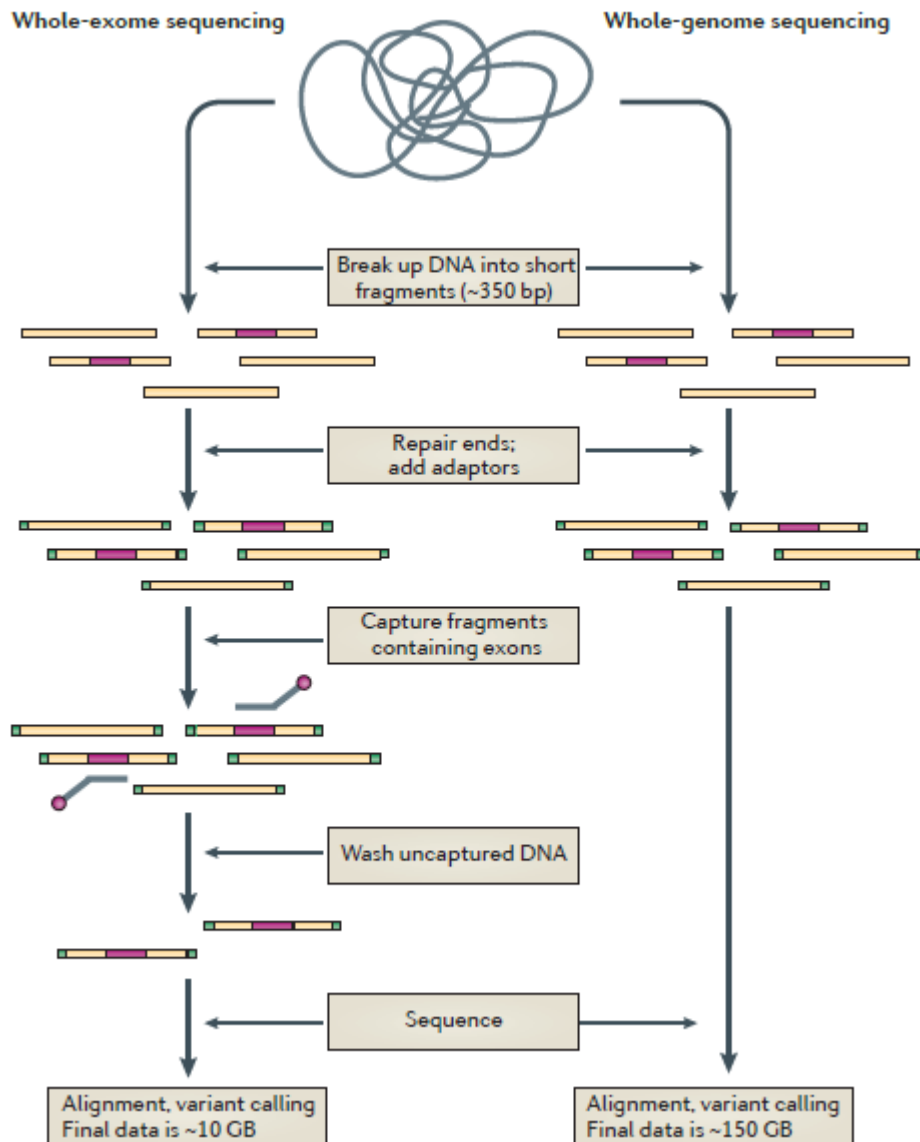


Figure 47 : Les étapes du WES et WGS. Fragmentation de l'ADN suivi de l'ajout des adaptateurs. Pour le WES il y a une étape de capture des exons (138).

1.8.2.1. *Le Whole Genome Sequencing*

Préparation des bibliothèques :

Pour chaque patient, l'ADN génomique est extrait à partir des leucocytes sanguins et fragmentés. Des adaptateurs permettant le séquençage et des index (codes-barres) permettant le multiplexage de l'ADN de plusieurs individus sont ligués au fragment d'ADN, avant une purification finale (Figures 48 et 49).

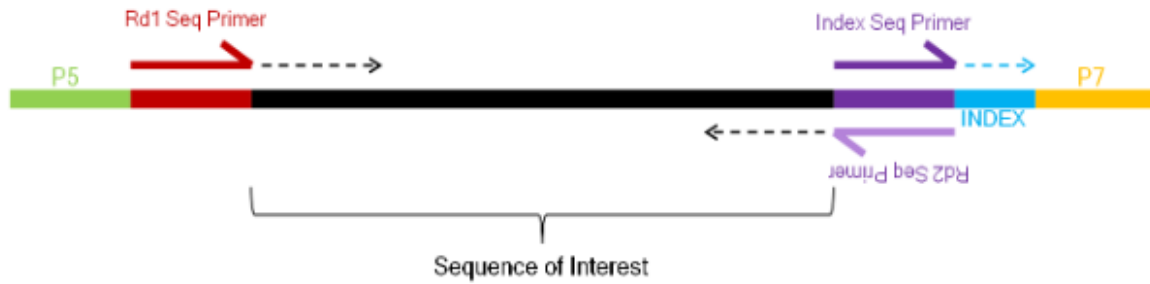


Figure 48 : Préparation de la librairie. Après fragmentation de l'ADN, à chaque fragment d'ADN (en noir) sera ajouté un index (bleu) correspondant à un code-barre spécifique pour chaque patient. Les adaptateurs P5 (vert) et P7 (orange) sont complémentaires à ceux présent sur la flow cell, les Rd1 SP (rouge) et Rd2 SP (mauve) servent comme point de départ à la synthèse des brins forward et reverse (Illumina®).

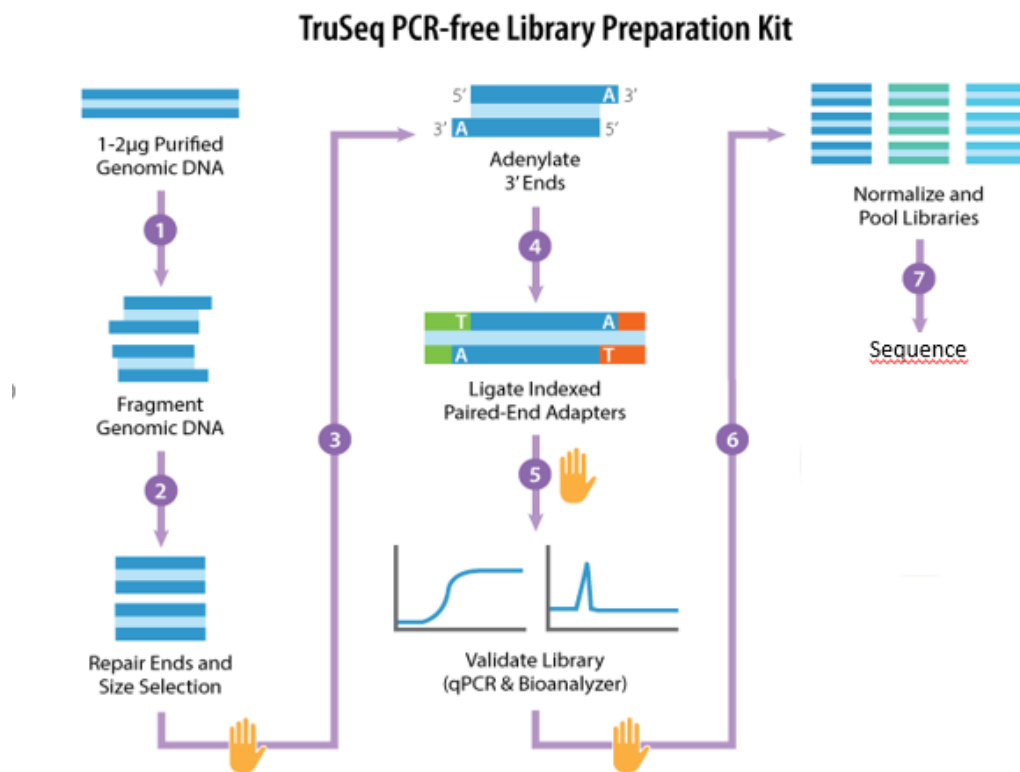


Figure 49 : Les différentes étapes pour la préparation de la librairie selon le kit TruSeq PCR Free (139).

Amplification clonale :

La spécificité de cette technologie est qu'elle a lieu sur un support solide, une surface de verre appelée *flow cell* (FC) et repose sur une amplification en pont (bridge PCR) des

fragments à séquencer. Les fragments de la librairie à séquencer possèdent des adaptateurs (P5 et P7) à leurs extrémités qui vont leur permettre de se fixer sur la FC de façon aléatoire, par hybridation sur les amorces qui sont complémentaires aux P5 et P7 et qui en couvrent la surface. Un nouveau brin est alors synthétisé par une polymérase. Le brin d'origine est éliminé par dénaturation et l'extrémité libre du brin restant s'hybride à une amorce adjacente pour former un pont. La polymérase synthétise à nouveau le brin complémentaire pour former un pont d'ADN double brins puis les deux copies sont libérées par dénaturation. Ce cycle d'amplification en pont recommence pour former un regroupement d'ADN clonal en une zone appelée cluster (Figure 50). Cette amplification en fragments d'ADN clonaux est nécessaire, car lors de l'étape de séquençage elle permettra d'amplifier le signal pour que la caméra de l'automate permette sa détection. Les brins anti-sens sont éliminés par clivage : c'est la linéarisation.

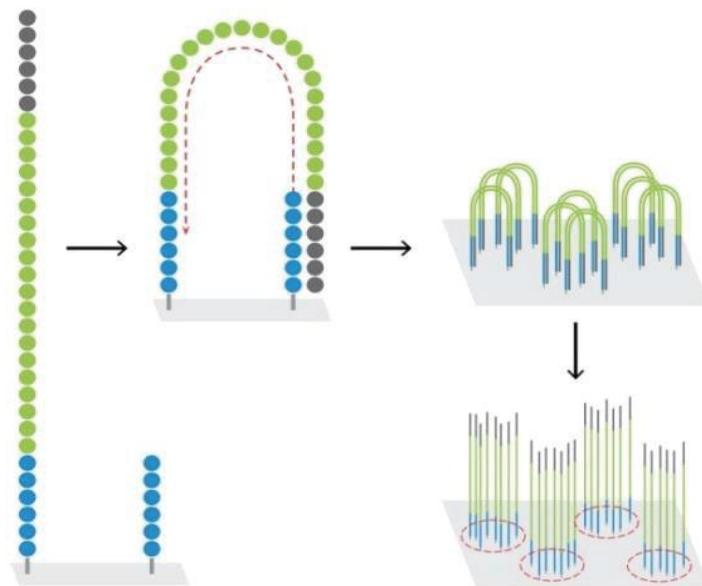


Figure 50 : Amplification clonale par Bridge PCR. *Le fragment d'ADN se fixe sur la flow cell par un brin. Par la suite, l'autre brin libre se fixe en formant un pont, il y a synthèse d'un nouveau brin complémentaire et ainsi de suite permettant la formation d'un cluster (Illumina®).*

Séquençage par synthèse :

Cette technologie se base sur la détection par un système de caméra de l'incorporation de nucléotides marqués lors de la synthèse du brin complémentaire. Les brins *forward* (Read 1) sont séquencés dans un premier temps. Le NextSeq 500 (illumina®) possède deux « Channels » et utilise seulement deux fluorochromes pour permettre de distinguer les nucléotides, T marqué en vert, C marqué en rouge, A marqué par les deux fluorochromes rouge et vert donnant une couleur jaune et le nucléotide G sans couleur.

Les différents nucléotides sont tous ajoutés dans le milieu, et le nucléotide complémentaire à la séquence est incorporé par une ADN polymérase. Après une étape de lavage des nucléotides non utilisés, et suite à l'excitation par un laser, la base ajoutée est détectée par l'émission de sa fluorescence. Afin d'éviter la lecture de plusieurs bases à la fois, le nucléotide inséré possède un bloqueur. Le cycle suivant, le bloqueur est éliminé et un deuxième nucléotide est inséré. Ces cycles de *Wash and clean*, appelés séquençage par synthèse (Figure 51), permettent la synthèse base par base du brin complémentaire et donc la détection de la séquence exacte de ce brin.

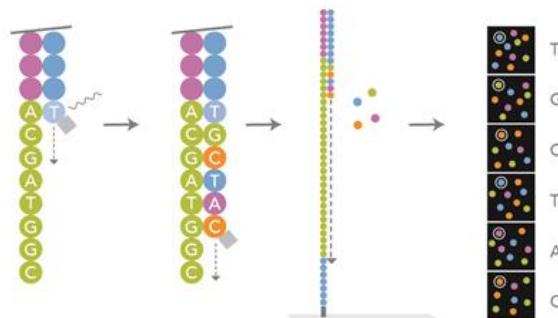


Figure 51 : Séquençage par synthèse. Le séquençage s'effectue grâce à une chimie de terminateurs réversibles : des nucléotides bloqués marqués par fluorescence. Quand l'un d'entre eux est incorporé, la fluorescence émise est relevée puis le fluorophore et le bloqueur sont clivés permettant l'ajout d'un nouveau nucléotide. A chaque cycle d'incorporation, une base peut être déterminée. (Illumina®)

Le séquençage en *paired-end* permet de séquencer les deux extrémités d'un fragment d'ADN. Dans ce processus, après avoir séquencé les n bases (en général entre 75 et 150) à une extrémité du fragment, la molécule d'ADN est retournée et c'est l'autre extrémité qui est séquencée selon le même principe (Figure 52). Ce type de séquençage facilite la détection des réarrangements structuraux.

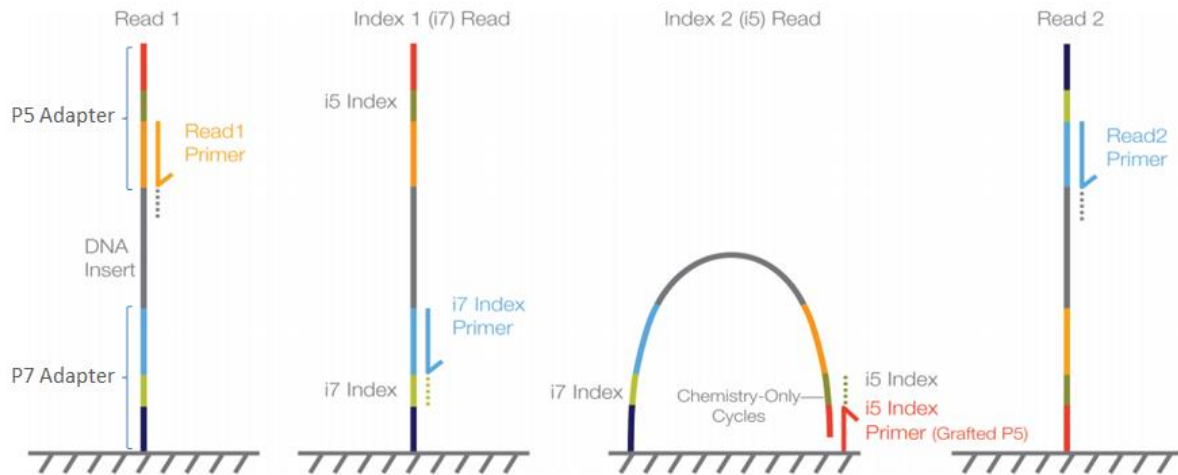


Figure 52 : Séquençage Paire-end. Après avoir séquençé la première extrémité, le fragment d'ADN se retourne et se fixe sur la flow cell afin de séquençer l'autre extrémité (Illumina®).

Deux notions sont alors importantes pour évaluer la qualité du séquençage réalisé. Il s'agit de la profondeur (*Depth*) qui correspond au nombre de reads qui couvrent une base, ainsi que la couverture (*Coverage*), qui représente la proportion de régions couvertes pour une profondeur donnée.

Trois approches ont été décrites pour analyser les diverses variations de structure :

- La méthode dite Read-depth, en se basant sur le fait que la quantité de séquences produites pour un locus donné reflète la quantité d'ADN de départ, permet de détecter des CNV et donne des informations similaires à celles obtenues par ACPA (Figure 53) (140). Cette méthode ne détecte pas les anomalies équilibrées.

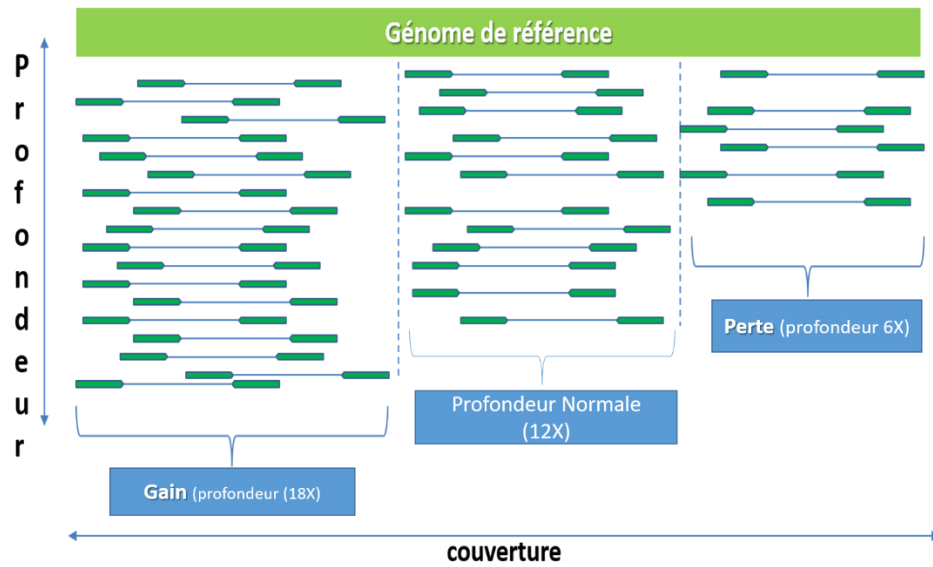


Figure 53 : Schématisation de la méthode de Read-depth. L'augmentation de la profondeur des reads indique la présence d'un gain, la diminution, une perte.

- La méthode dite Read-pair se base sur l'alignement et l'orientation des reads pour identifier les réarrangements chromosomiques (141). En général, cette méthode détecte les différents types d'anomalies de structure (Figure 54).

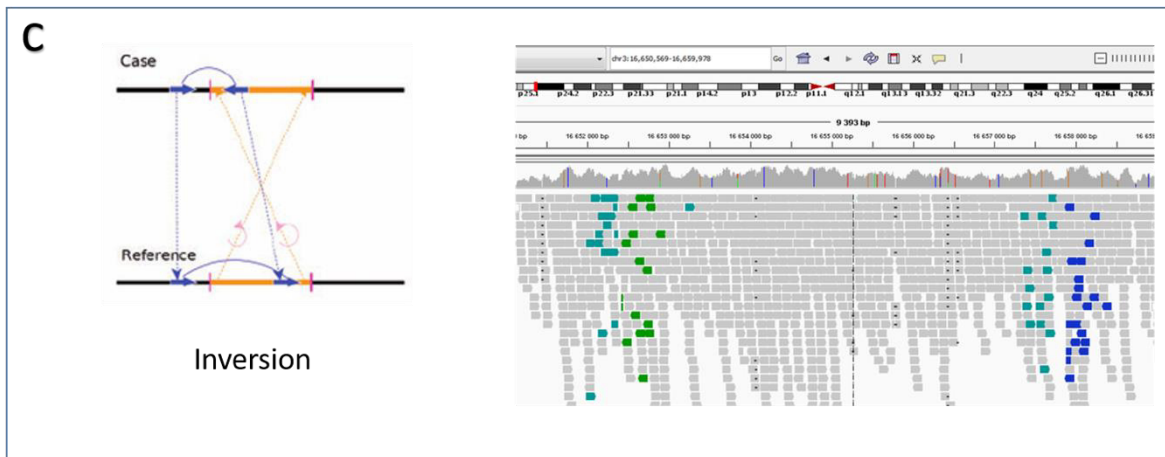
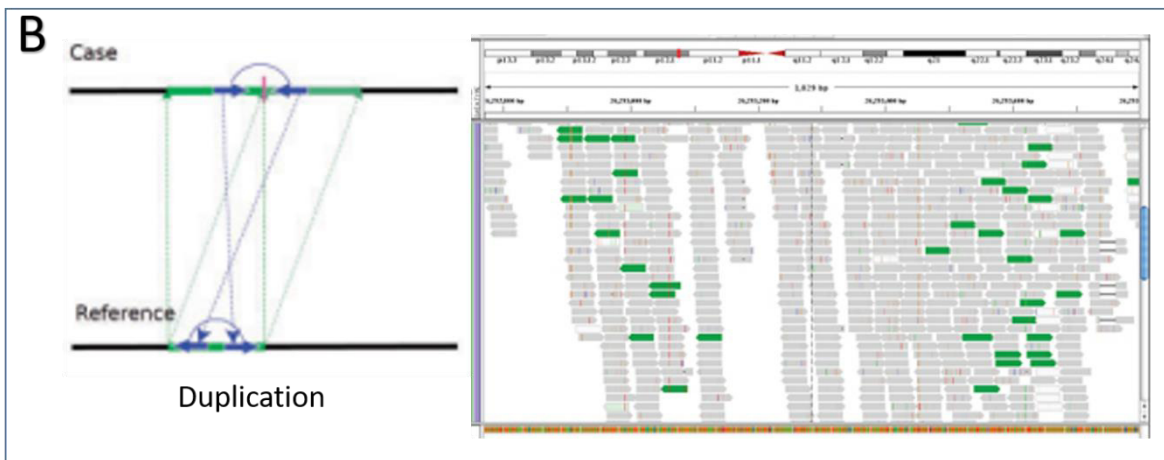
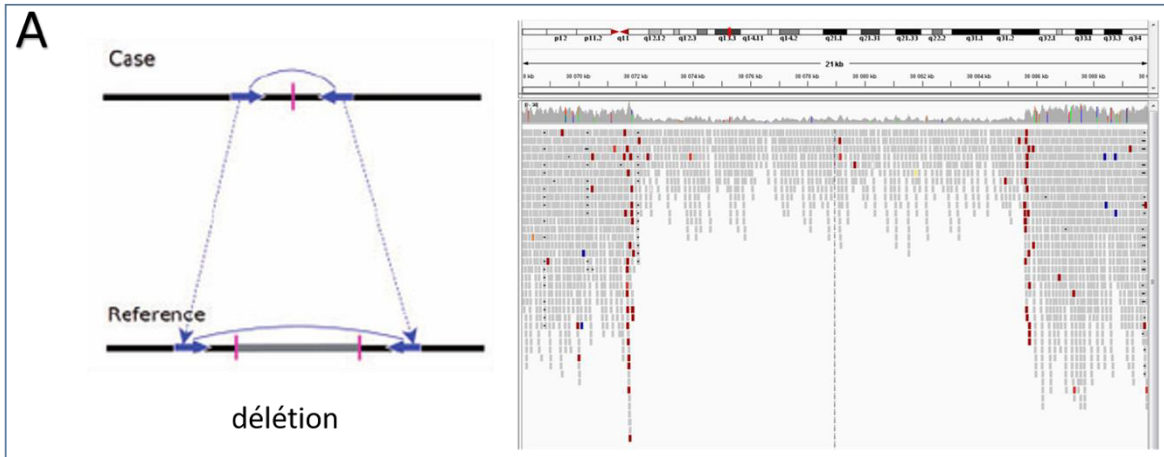




Figure 54 : Méthode Read-pair. Schématisation des différentes anomalies de structure avec la visualisation sur le logiciel IGV, schéma modifié de Ruibin et al. (142).

Dans le cas normal, chaque paire de reads est alignée sur le génome de référence dans le bon ordre, d'une orientation correcte et la taille de l'insert est celle prévue.

A – dans le cas d'une délétion, les paires de reads alignés sur un génome de référence s'étendent plus loin qu'attendu.

B- dans le cas d'une duplication, la distinction entre une duplication en tandem ou inversée se fait à partir de l'orientation et l'alignement des brins.

C- dans le cas d'une inversion, les reads de part et d'autre du point de cassure s'alignent dans une orientation anormale

D- dans le cas d'une translocation, les reads de part et d'autre du point de cassure s'alignent sur 2 chromosomes différents.

- La méthode dite Split-read, utilise les lectures ou reads qui chevauchent directement le point de cassure chromosomique. Dans ce cas, une des deux reads s'aligne sur les deux loci partenaires. Le point de cassure est situé entre ces jonctions (Figure 55).

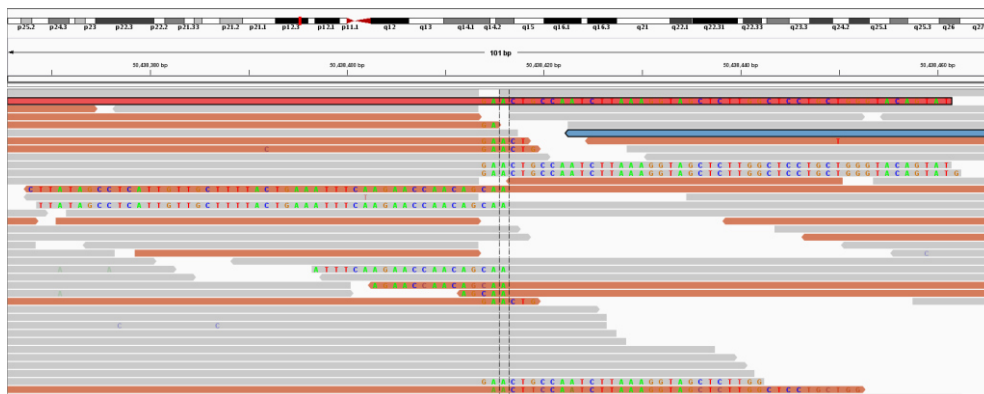


Figure 55 : Méthode Split-read. Visualisation sur le logiciel IGV de split read au niveau du chromosome 6 chez un patient (143)

Les RCAE sont des anomalies de structure regroupant les translocations, les inversions, les insertions et les remaniements complexes. Elles sont révélées par le caryotype, mais en raison de sa résolution limitée, ni les pertes ni les gains éventuels de matériel génétique au niveau de ces réarrangements ne seront décelés. En général, les sujets porteurs d'un RCAE sont asymptomatiques et touchent 1,54‰ des nouveau-nés (144). Dans des cas rares, ces RCAE sont associés à un phénotype anormal. Comme déjà discuté dans le paragraphe 1.4, ce phénotype anormal peut résulter d'un déséquilibre cryptique (145), d'une pathologie du point de cassure (146) où il peut s'agir d'une association fortuite. L'identification des points de cassure peut être réalisée grâce aux techniques de FISH, d'array painting, de PCR « long range », ou de Southern blot. Ces techniques étant longues, difficiles à réaliser et coûteuses (147) l'exploration de ces patients est jusqu'à présent réalisée dans un cadre de recherche.

Le WGS est une technique pangénomique, utilisée dans plusieurs situations diagnostiques. Elle permet d'identifier les variations nucléotiques, mais aussi les réarrangements structuraux comme expliqué supra. La démocratisation des techniques de type NGS et la réduction de coûts a permis d'identifier rapidement de nombreux points de cassures par étude génome entier de patients porteurs d'un RCAE et un phénotype (146,148,149). Ces résultats ont permis le transfert de la caractérisation des points de cassure de la recherche vers le diagnostic. Cette technique a été utilisée dans le cadre de ce travail, pour caractériser les points de cassures chez des patients présentant une anomalie du développement et/ou une déficience intellectuelle.

2. Matériels et Méthodes

2.1. Hybridation in situ en fluorescence (FISH)

Les sondes sont marquées à partir de BAC par nick-translation : 1 µg d'ADN est incubé avec 0,01 mM de dTTP, 0,01 M de DTT, 0,04 mM de dNTP, 0,01 mM de dUTP marqué soit à la fluoréscéine (vert), soit à la rhodamine (rouge), 0,08 ng/µL de DNase et 0,14 U/µL de DNA Polymérase (Promega) dans le tampon approprié pour un volume final de 50 µL, pendant 3h à 15°C puis la réaction est inactivée 10 min à 72°C. La sonde est précipitée en présence de d'ADN Cot-1 et reprise dans du tampon d'hybridation. L'hybridation des sondes se fait par dépôt sur des chromosomes en métaphase étalés à partir du culot issu de la culture des lymphocytes sanguins du patient. Après les lavages, les lames sont contre-colorées au DAPI et lues au microscope à épifluorescence AxioImager Z2 (Zeiss) à l'aide du logiciel Isis v4.0 (Métasystem).

2.2. Analyse chromosomique sur puce à ADN de la technologie Affymetrix

L'ADN génomique des patients est digéré par une enzyme de restriction (NSPI), suivi d'une ligation des adaptateurs afin de l'amplifier. Ensuite une fragmentation du produit amplifié est effectuée suivie par un marquage de l'ADN du patient. Enfin une hybridation sur la puce d'ADN est réalisée. Une puce de type SNP Affymetrix 750K a été utilisée. Cette puce contient des marqueurs non polymorphes permettant de détecter des CNV et des marqueurs de types SNP permettant d'identifier des régions d'homozygotie. Cette puce est scannée par le scanner « Gene Chip, Affymetrix » qui va générer un fichier de type *Cel File*, qui sera analysé par le logiciel ChAs (Chromosome Analysis Suite software).

2.3. Caractérisation des points de cassure par WGS

Pour chaque patient, l'ADN génomique est extrait à partir des leucocytes sanguins par le kit QIAamp Blood Midi Kit (QIAGEN). Un dosage par spectrophotométrie sur Nanodrop permet de vérifier l'efficacité de l'étape d'extraction (le rapport 260 sur 280 nm doit être entre 1.8 et 2) et de quantifier l'ADN extrait.

La préparation des banques est réalisée grâce au kit TruSeq DNA PCR Free (Illumina®) (Figure 49). Pour chaque patient, 3 µg d'ADN dans 55 µL final de tampon « *Resuspension Buffer* » sont d'abord fragmentés par sonication au Covaris M220 (programme : Peak incident

power 50 W, Duty factor 20%, cycles per burst 200, 65 secondes à 20°C) pour obtenir des fragments de taille d'environ 350 pb. Ces fragments sont ensuite purifiés, puis les extrémités sont réparées et adénylées. Des adaptateurs permettant le séquençage et des index (codes-barres) permettant le multiplexage de l'ADN de plusieurs individus sont ensuite ligués à ces fragments, avant une purification finale. Les librairies sont ensuite vérifiées par Qubit (LifeTechnologies®). La taille des fragments obtenus (comprenant à leurs extrémités les adaptateurs P5, P7 et les séquences index) est validée par Bioanalyzer (Agilent®). La quantification de la librairie est réalisée par qPCR (kit KAPA Library Quantification Kit Illumina Platforms) qui permet de déterminer la concentration en fragments chez chaque patient. Deux banques sont mélangées de façon équimolaire (afin de séquencer la même quantité d'ADN pour chaque patient), dénaturées et diluées à 1,8 pM dans un volume final de 1,3 mL de tampon HT1. Les librairies sont amplifiées de façon clonale sur la *Flow Cell* par *clustering* puis séquencées par synthèse en paired-end 2x101 pb. Ces deux étapes l'amplification et le séquençage sont réalisés sur l'appareil NextSeq500 (Illumina®).

2.3.1. *Analyse bio-informatique*

Le volume important de données de séquençage générées nécessite un traitement bio-informatique spécifique. L'analyse des images de la *Flow Cell* et l'appel des bases (*base calling*) sont réalisés par le logiciel d'analyse en temps réel *Real-time analysis software Illumina RTA2* avec les paramètres définis par défaut. Le démultiplexage, (.BCL), est effectué grâce au logiciel *bcl2fastq* (v2.16), en autorisant au maximum un mésappariement dans la séquence de chacun des index, afin d'attribuer à chaque patient ces propres séquences que l'on appelle *reads*. On obtient des fichiers de format Fastq. Ensuite pour chaque patient, les lectures ainsi obtenues sont alignées sur la séquence du génome humain de référence (version GRCh38/hg38) par le logiciel BWA (algorithme mem, version 0.7.10) (150), qui conduit à l'obtention de fichiers au format .bam (Figure 56). Les alignements sont ensuite analysés par le logiciel BreakDancer (version 1.4.5) (151), ce qui permet de détecter un certain nombre de variations de structure potentielles si la position, l'orientation et la distance séparant les 2 reads de la même paire chez un patient ne sont pas conformes à ce qui est observé dans le génome de référence et le logiciel ERDS version 1.1 (152) qui permet de détecter les CNV. Après l'appel des variants structuraux pour chaque échantillon, grâce au Logiciel Svagga (153), un fichier excel sera produit qui est annoté.

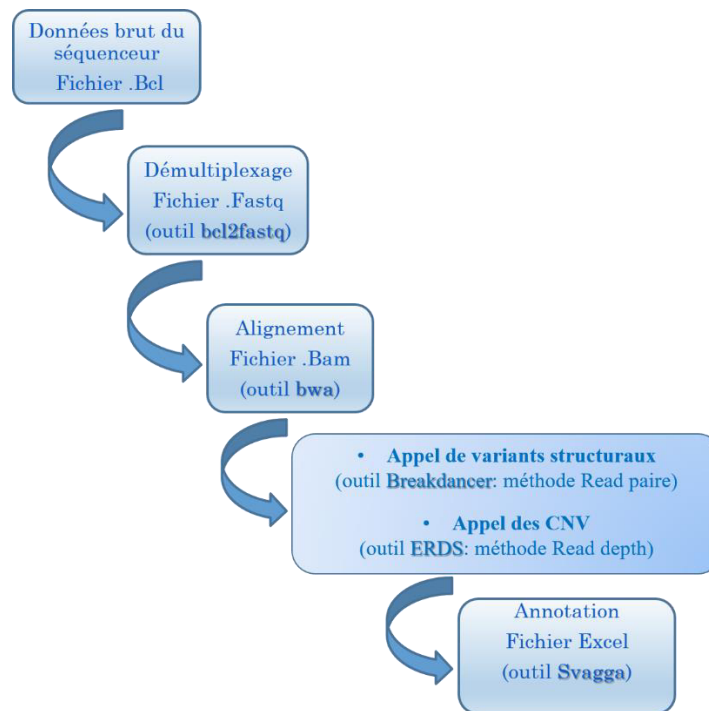


Figure 56 : Pipeline bio-informatique utilisé pour la détection des variants structuraux.

2.3.2. Amplification par PCR de fragments de jonction contenant les points de cassure et séquençage Sanger

Les points de cassure identifiés par WGS sont caractérisés précisément par amplification PCR et séquençage Sanger.

Inventée par K. Mullis en 1985, la PCR est une technique d'amplification de l'ADN. Elle repose sur une succession de réactions de réplifications *in vitro* permettant de générer, à partir des quantités infinitésimales d'extrait d'ADN (ADN matriciel), de multiples copies d'un fragment d'ADN particulier (séquence cible). L'amplification, par PCR, d'une séquence donnée nécessite l'usage d'un couple d'amorces (sens et anti-sens) spécifiques possédants des T_m (melting temperature) voisins (154). Les amorces sont deux oligonucléotides de synthèse d'environ 20 à 25 nucléotides situées de part et d'autre de la séquence à amplifier. La formule empirique pour calculer la température de fusion T_m d'un oligonucléotide est la suivante : $T_m = [4(G+C) + 2(A+T)]$ qui n'est pas valable pour des tailles supérieures à 25 nucléotides.

Une paire d'amorces est choisie de part et d'autre du point de cassure identifié afin d'amplifier le fragment de jonction. A partir des points de cassure identifiés, un modèle du chromosome dérivé est constitué, à partir du génome de référence de la base de données UCSC(155). Les amorces sont choisies à l'aide du logiciel Primer 3.

L'amplification des fragments de jonction est réalisée avec le kit Taq CORE 10 (MP Biomedicals) dans les conditions suivantes : 120 ng d'ADN, 1,75 mM de MgCl₂, 0,2 mM de dNTP, 0,3 μM de chaque amorce et 0,03 U/μL de Taq, dans un volume final de 50 μL. Le programme de PCR suivant est utilisé : dénaturation à 96°C pendant 10 min, puis 35 cycles avec une dénaturation de 45 sec à 96°C, une étape d'hybridation de 30 sec à 60°C (variable en fonction du T_m du couple d'amorces) et une étape d'élongation de 45 sec à 72°C, et une phase d'élongation finale de 10 min à 72°C. Cette réaction d'amplification clonale est effectuée dans un thermocycleur.

L'amplification est réalisée chez le patient et chez un témoin négatif (individu non porteur du RCAE). Un contrôle positif, *ATPIA3*, exons 7-8 est également amplifié chez le patient et chez le témoin négatif afin de vérifier la qualité des ADN et que la manipulation a bien fonctionné. Les produits de PCR sont vérifiés par migration sur puce grâce au LabChip GX (Perkin Elmer). Le produit de PCR est purifié puis séquencé en méthode Sanger sur l'automate ABI3500DX (ThermoFischer®) afin de déterminer la localisation du point de cassure à la base près, par alignement de la séquence du fragment de jonction avec celle du génome de référence à l'aide de l'outil Blat du site de l'UCSC.

2.4. Séquençage d'exome

Un microgramme d'ADN a été fragmenté par sonication sur l'automate Covaris M220 (PerkinElmer) en fragments de 180 à 220 paires de bases (programme : Peak incident power 50 W, Duty factor 20%, cycles per burst 200, 150 secondes à 20°C). Une électrophorèse sur le système Labchip GX Touch (PerkinElmer) est effectuée afin de contrôler la taille de la fragmentation des échantillons. La préparation des librairies est réalisée en utilisant le kit SeqCap EZ MedExome (Roche). Les étapes de la réparation des brins, l'adénylation et l'ajout des adaptateurs sont automatisées sur l'automate Siclone (PerkinElmer). Ensuite une purification est effectuée sur l'automate Zéphyr (PerkinElmer). L'enrichissement des fragments d'intérêt, capture des exons s'est fait grâce au kit Hypercap Target Enrichment (Roche). Son principe consiste à une étape d'hybridation des fragments d'intérêts, suivie par une purification et d'une amplification. Ensuite les librairies sont validées, en utilisant le Bioanalyzer, (Agilent) qui nous permet de vérifier leur taille.

Les librairies sont regroupées de façon équimolaire après dosage par Qubit, en pools de 12 patients afin que tous les patients aient une même profondeur et couverture des exons puis

dénaturées par HT1. Une amplification clonale pour la génération des clusters sur la flow cell suivi d'un séquençage en paired-end pour obtenir des fragments de 150 pb.

Le pipeline bio-informatique est résumé dans la figure 57. Le démultiplexage suivi d'un alignement produisant pour chaque patient un fichier Bam. L'appel de variant a été réalisé à l'aide du programme GATK HaplotypeCaller (156) produisant un fichier VCF génomique pour chaque échantillon qui est annoté grâce aux outils SnpEff-SnpSift.

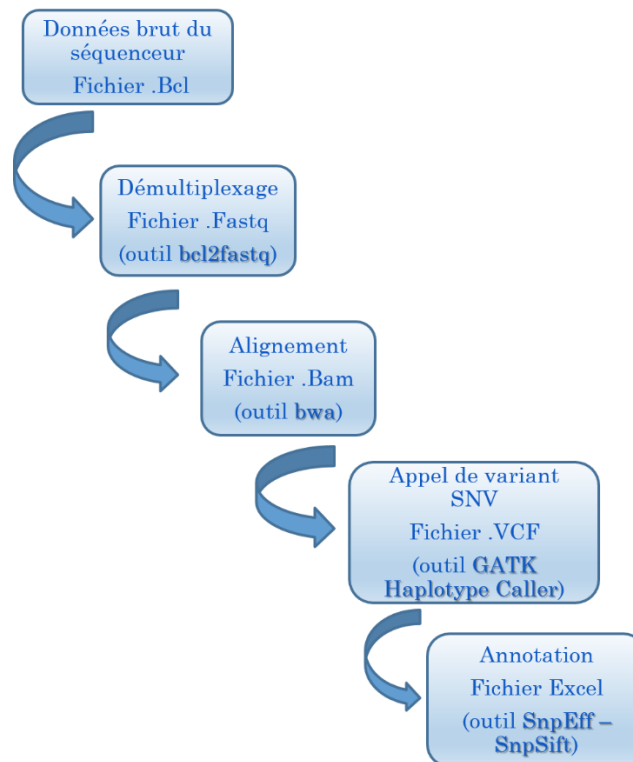


Figure 57 : Pipeline bio-informatique utilisé pour la détection des SNV.

Seuls les variants hautement fiables ont été conservés pour l'analyse. Les variants introniques, les variants présentant un biais de brin (variant présent uniquement sur les lectures du même sens), une fréquence allélique mineure (MAF) >1% dans les bases de données publiques comprenant Exac, dbSNP, Exome Sequencing Project (ESP) et 1000 Genome Project, et celles qui présentent une mosaïque faible (< 10%) ou une faible profondeur (< 10X) ont été filtrées. Les variants faux-sens, non-sens, insertion, délétion ou affectant les sites d'épissage se trouvant dans des régions codantes ont été conservés ainsi que les variants présentant des scores de pathogénicité avec un ou plusieurs logiciels de prédictions *in silico* (SIFT, Polyphen2, Mutation Taster). Notre approche a été la réalisation d'un exome en solo, car malheureusement les parents n'étaient pas disponibles. L'exome en trio (patient et les deux parents) aurait permis un tri de variant facilité.

Dans un premier temps une analyse des variants prenant en compte les modes de transmission autosomique dominant, autosomique récessif homozygote et hétérozygote composite, lié à l'X a été effectuée. Concernant l'étude sur l'infertilité, l'analyse a porté sur les gènes qui étaient liés à l'infertilité masculine monogénique humaine et à l'azoospermie (121).

2.5. Etude d'expression par Reverse transcription (RT) qPCR

Le but de cette technique est d'obtenir de l'ADN complémentaire (ADNc) à partir de l'ARN total. Une reverse transcription est réalisée avec le kit SuperScript II Reverse Transcriptase (Invitrogen) à partir de 0,5 µg d'ARN total extrait des patients et de témoins, dans un volume final de 20 µL. Pour chaque échantillon, deux réactions de RT sont effectuées, annotées RT (+) et RT (-). Le tube RT (-) sert de contrôle de contamination par l'ADN génomique. Pour cela, l'ARN est incubé 5 min à 65°C après ajout de 10 ng/µL d'amorces hexanucléotides (« Random Primers ») et de 0,5 mM de dNTP. Après ajout de tampon 5X First Strand Buffer et 0,01 M de DTT, le mélange est incubé pendant 2 min à 25°C. 10U/µL de l'enzyme SuperScript™ II RT est ensuite ajouté dans les tubes annotés RT (+). Le mélange est de nouveau incubé à 25°C pendant 10 min puis à 42°C pendant 50 min. L'enzyme est ensuite inactivée en chauffant à 70°C pendant 15 min. A la fin de la réaction, 0,01 U/µL de RNase H est ajoutée, suivi d'une incubation de 20 min à 37°C.

2.5.1.1. PCR quantitative (qPCR)

Cette technique est utilisée pour l'étude de l'expression d'un gène ou confirmer l'existence biologique d'un CNV. Cette technique appelée encore « PCR en temps réel » permet de détecter la fluorescence émise à chaque cycle par les produits de PCR. Le profil de cette réaction est composé de trois phases : la première correspond à une phase d'initiation (des premiers cycles jusqu'à une valeur seuil, appelée Ct), une phase exponentielle où à chaque cycle un dédoublement de l'amplicon est visualisé puis la phase de plateau. La Ct ou « cycle threshold » correspond à l'intersection entre la valeur seuil et la courbe d'amplification (Figure 58). Le coefficient d'efficacité de la réaction est calculé grâce à une gamme étalon. Cette gamme est réalisée par une série de dilutions à moitié qui en théorie va donner des courbes décalées d'un cycle de PCR.



Figure 58 : La courbe obtenue d'une amplification par qPCR montrant les trois phases : initiation, exponentielle, et le plateau. Ct représente l'intersection entre le seuil et la courbe d'amplification.

Les amorces ont été choisies à l'aide de Primer 3 et ont été vérifiées pour leur spécificité sur PrimerBlast. La réaction de qPCR est réalisée avec les amorces d'un gène de référence, *ACTB* (exon4-5) et les amorces du gène d'intérêt. Deux gammes d'étalonnage sont réalisées en triplicats, une pour *ACTB* et l'autre pour le gène d'intérêt avec le kit QuantiTect SYBR Green PCR kit (QIAGEN), à partir de dilutions aux 1/10, 1/20, 1/40 et 1/80 de l'ADNc d'un témoin afin d'évaluer l'efficacité de la réaction d'amplification. L'ADNc à tester et des ADNc de contrôles sont dilués au 1/20. A 2 μ L d'ADNc dilués sont ajoutés 1,5 μ M d'amorce sens et d'amorce anti-sens, et 1X de Sybr Green dans un volume final de 20 μ L. La plaque est insérée dans le LightCycler 2000 (Roche) et lancée avec le programme suivant : 1 cycle de dénaturation initiale de 15 min à 95°C ; 45 cycles d'amplification avec une dénaturation de 15 sec à 95°C, une phase d'hybridation de 30 sec à 60°C (dépend du T_m de l'amorce) et une phase d'élongation de 15 sec à 72°C ; 1 cycle de fusion avec une augmentation progressive de 0,1°C/sec jusqu'à 95°C, puis 1 cycle de refroidissement de 30 sec à 40°C. Les résultats obtenus sont normalisés par rapport à l'expression du gène de référence ; et la modification de l'expression est analysée par quantification relative grâce à la méthode du calcul du $2^{-\Delta\Delta Ct}$ (157).

3. Caractérisation moléculaire des RCAE chez 3 patients présentant une infertilité masculine

Dans l'infertilité masculine, la probabilité de trouver une anomalie chromosomique de structure est de 2 à 20 fois plus élevée que dans la population générale (34). Ces anomalies de structure peuvent altérer la spermatogenèse et par la suite les paramètres spermatiques. D'autres hypothèses permettent d'expliquer de l'altération du synapsis chromosomique lors de la gamétogenèse. Au cours de la spermatogenèse, les chromosomes X et Y forment une vésicule sexuelle et seules les régions pseudo-autosomiques s'apparient entre elles. Les régions non appariées subissent une inactivation méiotique. Cependant, lors de la présence d'un réarrangement chromosomique, les régions non appariées subissent un processus d'inactivation de la transcription (47). Deux hypothèses peuvent être responsables de l'arrêt de la division cellulaire. Soit ce processus de l'inactivation de la transcription, lui-même, inactive des gènes nécessaires au bon déroulement de la méiose. Ou bien, ces régions non appariées peuvent interagir avec la vésicule sexuelle, perturbant ainsi l'inactivation des chromosomes sexuels et provoquant l'arrêt précoce de la méiose (48). Par ailleurs, certains patients sont porteurs de RCAE et ne présentent aucune altération spermatique. Plusieurs questions se posent en plus de la perturbation de la méiose : ces RCAE interrompent-ils un gène impliqué dans l'infertilité masculine qui est la cause des anomalies spermatiques ? Est qu'il y a une autre cause génétique en plus des RCAE, responsable du phénotype chez le patient ?

Peu d'études portant sur l'identification des points de cassure des RCAE chez les patients infertiles ont été rapportées. L'étude de Mouka et al. (49) a permis d'identifier un gène candidat, *SYCP3*, comme responsable de l'infertilité après avoir caractériser les points de cassures d'un remaniement complexe entre les chromosomes 7 et 12 chez le patient. Récemment, Schilit et al. ont identifié l'implication du gène *SYCP2* dans l'oligospermie sévère chez des patients présentant soit une variation nucléotidique de type frameshift ou bien une translocation au niveau du chromosome 20q13.3 modifiant l'expression de ce gène.

Dans une première partie, l'objectif est de répondre à ces questions. Dans ce cadre, trois patients présentant une translocation apparemment équilibrée et ayant un trouble de la spermatogenèse ont été caractérisés afin de confirmer ou non l'hypothèse que le RCAE interrompt et/ou modifie l'expression d'un gène par effet de position. Afin d'éliminer une

association fortuite avec une anomalie nucléotidique pathogène, un séquençage d'exome a été effectué.

3.1. Patient 1 :

Ce patient de 32 ans a été référé pour une infertilité après 5 ans de rapports sexuels sans conception. Il présente une azoospermie. Le développement sexuel est normal. Il n'a pas d'antécédents médicaux ni chirurgicaux. Le bilan hormonal, incluant testostérone, cortisol, T3, T4, TSH, estradiol, FSH, LH, prolactine et inhibine-B, est normal. Le caryotype constitutionnel a révélé une translocation réciproque entre les bras longs d'un chromosome 8 et les bras courts d'un chromosome 22 : 46,XY,t(8;22)(q13;p11) (Figure 59). Une ACPA a été effectuée chez ce patient révélant l'absence de remaniement cryptique au niveau des points de cassure et aucun déséquilibre chromosomique n'a été détecté. La recherche d'une microdélétion du chromosome Y était négative.

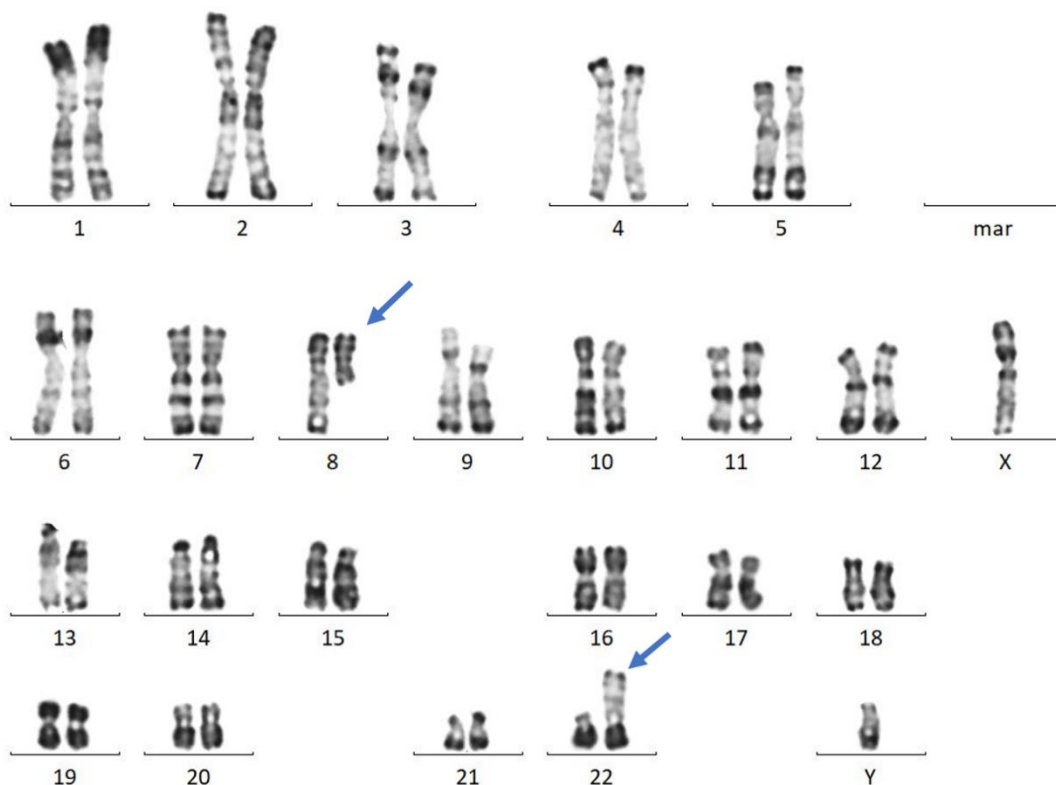


Figure 59 : Caryotype en bande R du patient 1 : 46,XY,t(8;22)(q13;p11).

3.2. Patient 2 :

Ce patient de 46 ans présente une oligo-asthéo-térazoospermie sévère. Le caryotype a montré une translocation réciproque entre les bras courts d'un chromosome 1 et les bras longs d'un chromosome 18: 46,XY,t(1;18)(p21;q12) (Figure 60). Le couple a subi 2 ICSI (IntraCytoplasmic Sperm Injection) avec seulement un transfert embryonnaire sans grossesse clinique.

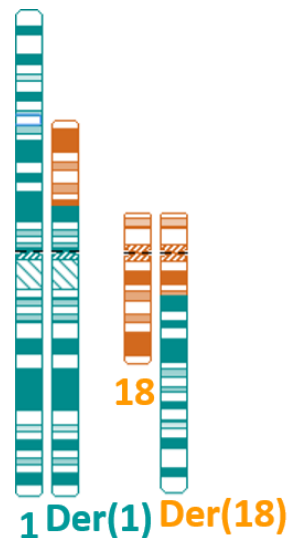


Figure 60 : Idéogramme illustrant la translocation t(1;18)(p21;q12) chez le patient 2.

3.3. Patient 3 :

Ce patient de 26 ans présente une oligo-asthénospermie et une importante térazoospermie. Le caryotype a révélé une translocation réciproque entre les bras longs d'un chromosomes 9 et les bras longs d'un chromosome 21: 46,XY,t(9;21)(q12;q22.3) (Figure 61). Il présente un diabète de type 1 et une hypertension. Les taux d'hormone sériques Testostérone, FSH, LH, prolactine, inhibine-B et hormone anti mullerienne sont normaux. Le couple a subi 2 ICSI, qui ont toutes deux échoué en raison de la fragmentation des embryons avant le stade blastocyste.

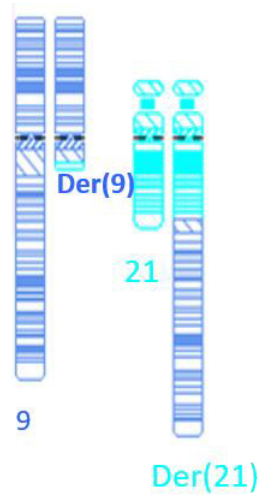


Figure 61 : Idéogramme illustrant la translocation $t(9;21)(q12;q22.3)$ chez le patient 3.

3.3.1. *Résultat*

3.3.1.1. *Patient 1*

Pour le patient 1, la profondeur moyenne de séquençage est de 13.9X. L'analyse des variants de structure proposés par le pipeline utilisant BreakDancer n'a pas permis d'isoler les points de cassure potentiels. Ceci peut s'expliquer par l'implication de la région d'hétérochromatine du bras court du chromosome 22.

Néanmoins, la marche sur le chromosome 8 par hybridation *in situ* en fluorescence a permis de trouver que le point de cassure se situait entre le BAC RP11-152C15 (chr8:69969490_70150856, hg38) et le BAC RP11-11K9 (chr8:71222159_71383240, hg38) (Figure 62).

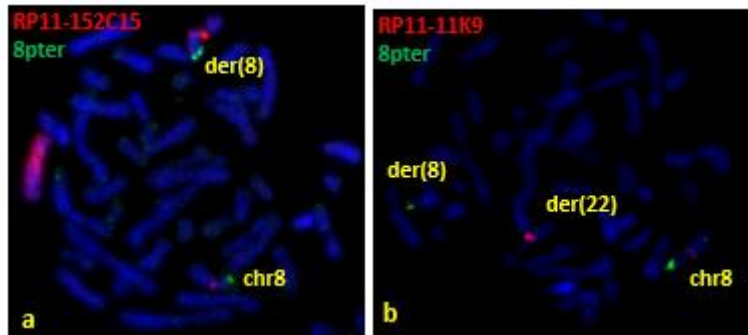


Figure 62 : Images d'hybridation *in situ* en fluorescence pour les sondes : a- RP11-152C15 (BAC) (8q13.3) (rouge) et 8pter (vert) (Cytocell, Cambridge, Royaume-Uni). La sonde RP11-152C15 est située sur le dérivé (8). b- RP11-11K9 BAC (8q13.3) (rouge) et 8pter (vert) (Cytocell, Cambridge, Royaume-Uni). La sonde RP11-11K9 est située sur le dérivé (22).

La délimitation de l'intervalle du point cassure sur le chromosome 8 permet ensuite de balayer manuellement la région d'intérêt sur le logiciel de visualisation IGV. Une zone avec quelques reads discordants est ainsi trouvée (Figure 63). Normalement, les deux reads d'une même paire, sont alignés sur le même chromosome, chacun sur un brin d'orientation différente et sont séparés par une distance d'environ de 400 pb. Lors d'une translocation, chaque *read* d'une même paire est aligné sur un chromosome différent. La visualisation sur Le logiciel IGV représente ces reads discordants par des flèches colorées (Figure 63).

La visualisation de cette zone a permis de situer le point de cassure à la position chr8:70520609_70521528 (hg38) en 8q13.3 tandis que les reads partenaires sont alignés soit sur le bras court du chromosome 21, soit sur une séquence non cartographiée sur le chromosome 22 (chr22_K1270733v1_random).

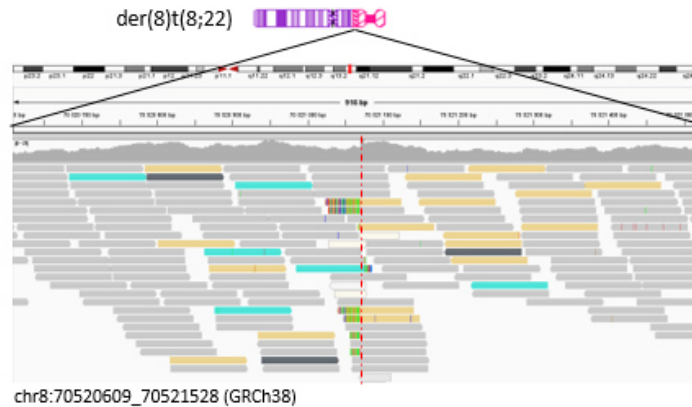
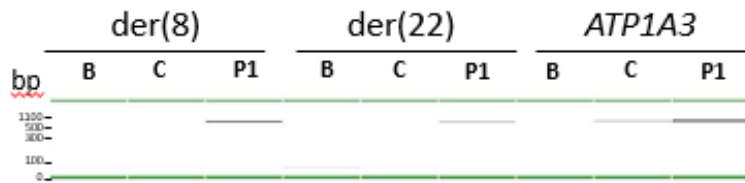


Figure 63 : Visualisation par IGV du point de cassure sur le chromosome 8 chez le patient 1 (points rouges). Les *reads* colorés sont des *reads* dont le *mate* / partenaire correspondant présente une discordance d'alignement sur le génome de référence. Les flèches beige se réfèrent à des paires mal alignées dont le partenaire s'aligne sur le chromosome 21; les flèches bleues se réfèrent à *mate* sur chr22_KI270733v1_random et les flèches gris foncé à *mate* sur chrUn_GL000220v1.

Après amplification par PCR (Figure 64a) et séquençage Sanger des fragments de jonction (Figure 64b), nous avons précisé de façon exacte les positions des points de cassure en (hg38) chr8:70521071_70521076 et chr22_K1270733v1_random:158856_158864. Cette dernière position correspond également à une duplication segmentaire sur le bras court du chromosome 21 (chr21:8241828_8241836). La jonction sur le dérivé 8 présente une structure particulière avec une microhomologie de 5 nucléotides (AAAAT), une insertion de 44 nucléotides d'origine inconnue (CAGAAAAAATCTACTACATAACAACATTCCTCTAGTATGTACTT), et une duplication des séquences autour des points de cassure (chr8:70520702_705215276 et 1587022_K1). La séquence de jonction sur le dérivé 22 montre une microhomologie de 5 nucléotides (TTTTA). Toutes ces caractéristiques suggèrent un mécanisme répliatif (FoSTeS, MM-BIR) expliquant ce réarrangement génomique complexe.

a



b

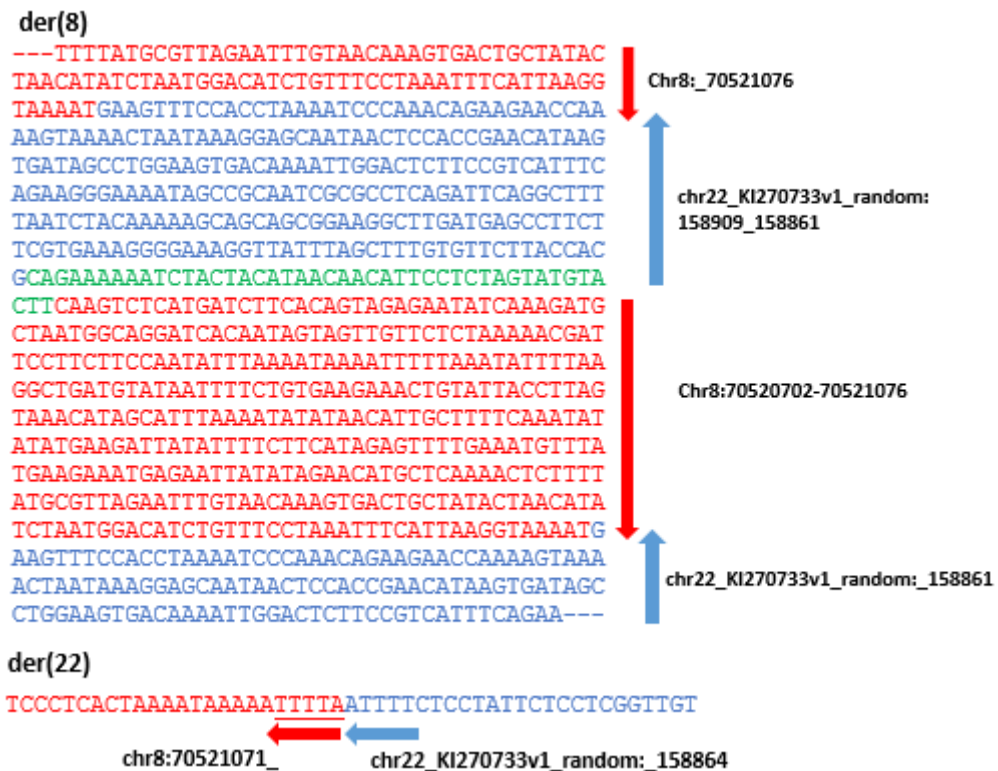


Figure 64 : a- Amplification des fragments de jonction des points de cassures chez le patient 1 par PCR (visualisation Labchip) pour le der (8), le der (22) et une séquence de contrôle (ATP1A3 exons 7-8). B : blanc; C: contrôle sans réarrangement chromosomique, P: patient. b- Séquences des fragments de jonction et coordonnées des points de cassures définies par le WGS et le séquençage de Sanger. En rouge : séquence de référence sur le chromosome 8; en bleu séquence de référence sur le chromosome 22, en vert, séquence d'origine inconnue. Les bases soulignées correspondent à la micro-homologie. Les flèches indiquent l'orientation du brin de référence.

Une technique de FISH supplémentaire a confirmé que les régions NORs (Nucleolus Organizer Regions) du chromosome 22 sont transloquées sur le dérivé 8, et que la région du chromosome 8 sur le dérivé 22 est très proche de la région d'hétérochromatine péricentromérique (Figure 65).

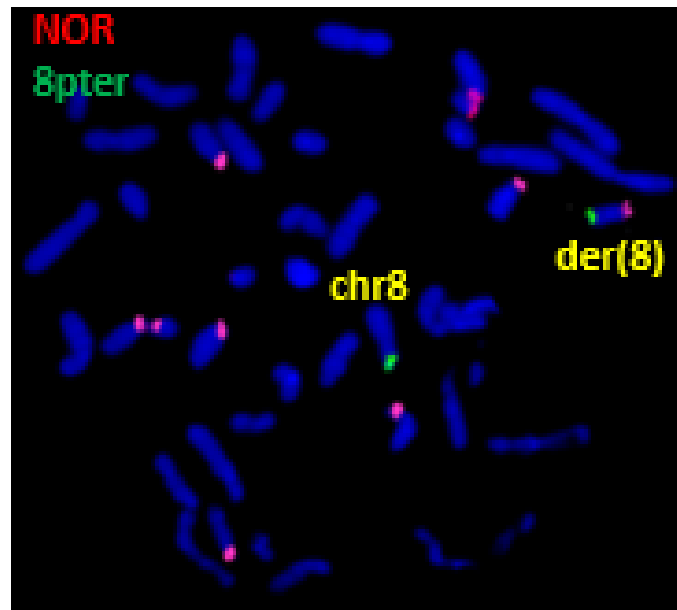


Figure 65 : Images d'hybridation in situ en fluorescence pour les sondes NOR et 8 pter (vert) (Cytocell, Cambridge, Royaume-Uni) montrant la présence de régions NOR sur le dérivé (8).

Le point de cassure au niveau du chromosome 8 n'interrompt aucun gène. L'exploration du domaine TAD (158), Topologically Associated Domain, contenant le point de cassure, nous a permis d'identifier quatre gènes candidat, *NCOA2*, *TRAM1*, *LACTB2* et *XKR9*, tous exprimés dans les testicules (Figure 66). En général, les gènes au sein d'un même TAD interagissent physiquement entre eux plus fréquemment qu'avec les autres TADs. Le repliement du génome en forme de boucle, rapprochent étroitement les enhanceurs de leurs gènes cibles (159).

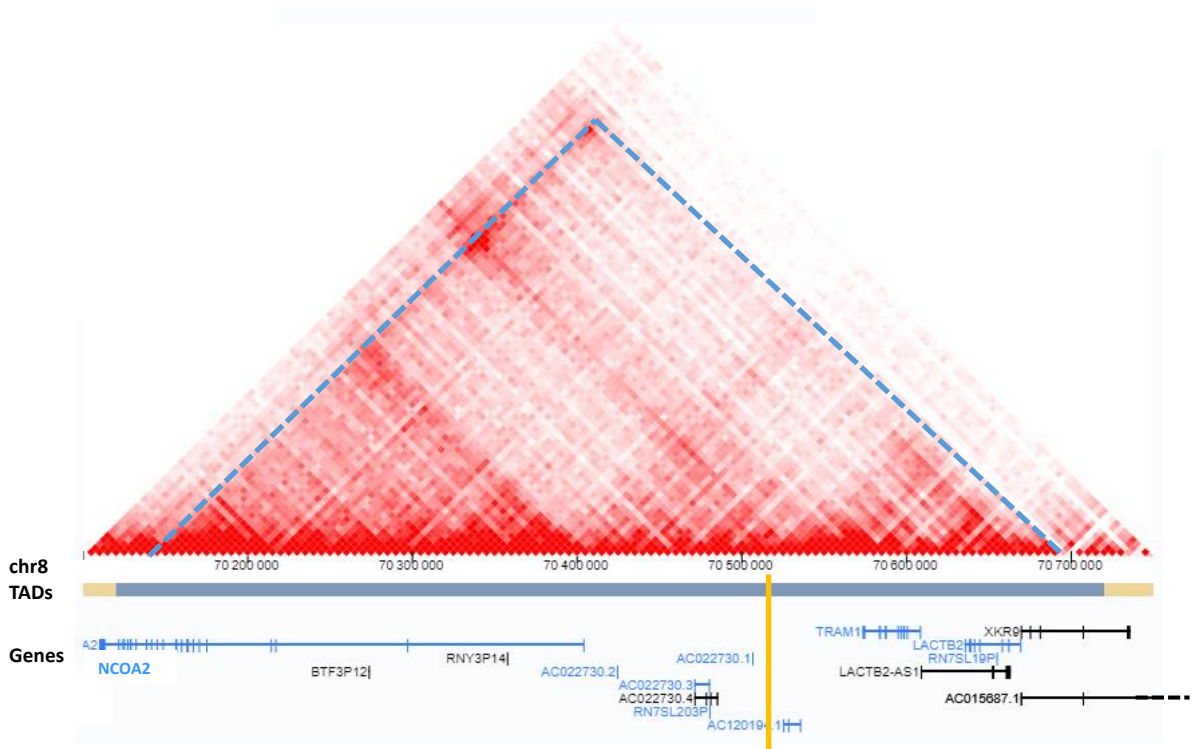
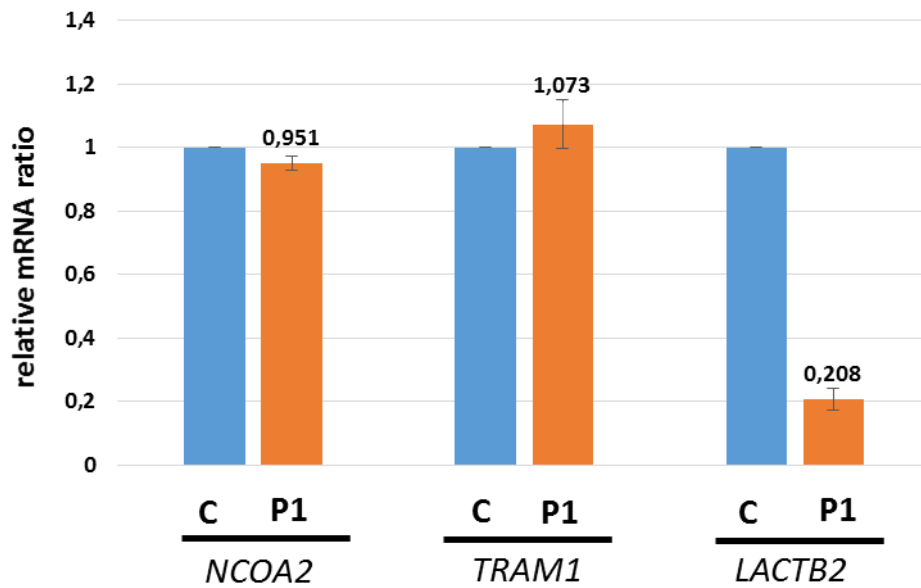


Figure 66 : Exploration de la région de part et d'autre du point de cassure sur le chromosome 8 : 70520702_70521076 sur la carte tridimensionnelle dérivés par la base de données Hi-C (résolution 50 Kb). Les gènes en noir sont localisés sur le brin *forward*, en bleu sur le brin *reverse*, en jaune la localisation du point de cassure (158).

Cette translocation peut affecter l'expression de ces gènes car elle implique les bras courts d'un chromosome acrocentrique, dont l'hétérochromatine peut se propager dans la région euchromatique du chromosome 8.

Les études d'expression ont été réalisées pour les gènes *NCOA2*, *TRAM1* et *LACTB2* (figure 67 a). *XKR9* n'étant pas exprimé dans le sang n'a pas pu être étudié. Elles ont révélé une diminution de près de 80% de l'expression de *LACTB2* par rapport aux témoins fertiles normaux (figure 67 b). Les témoins infertiles ont également montré une diminution de 30% de l'ARNm de *LACTB2* (figure 67 b), ce qui suggère que l'expression de *LACTB2* peut être affectée par des troubles de la spermatogenèse. Cependant, la diminution est plus importante chez le patient 1 que chez les témoins infertiles, ce qui suggère un impact fonctionnel du point de cassure sur le gène *LACTB2* chez le patient 1.

a



b

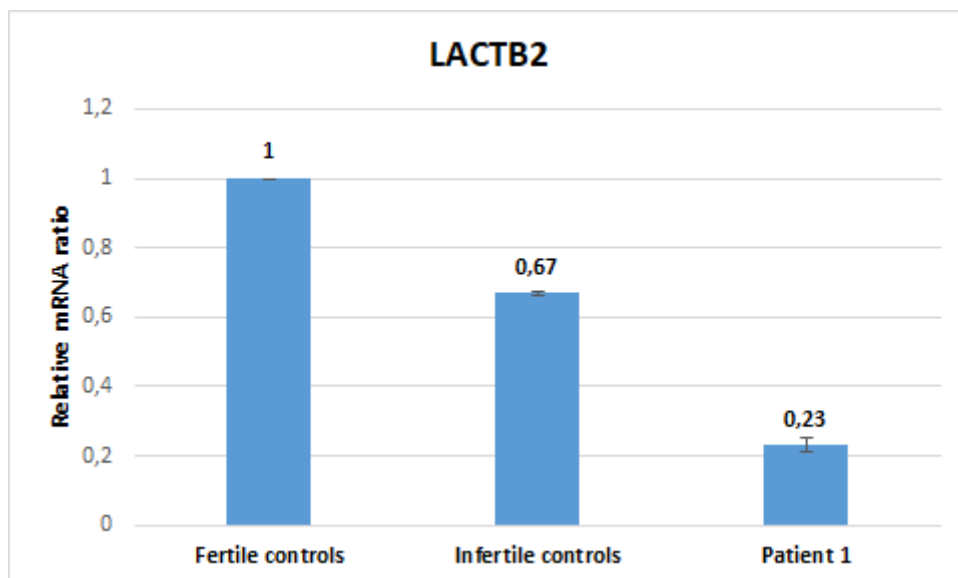


Figure 67 : a- Etude de l'expression des gènes *NCOA2*, *TRAM1* et *LACTB2* dans le tissu sanguin chez le patient 1 par rapport aux témoins fertiles. b- Etude de l'expression du gène *LACTB2* dans le tissu sanguin chez le patient 1 par rapport aux témoins fertiles et infertiles.

Afin d'exclure l'implication d'un SNV pathogène survenu indépendamment du réarrangement chromosomique identifié, un séquençage d'exome a été effectué. En effet la couverture moyenne de 15X du WGS permet de détecter des variant structuraux mais ne permet pas la détection fiable des SNVs.

L'analyse de l'exome chez le patient 1 avec une couverture moyenne de 98% à 30X n'a détecté aucun SNV pathogène dans les gènes impliqués dans l'infertilité.

De plus, nous avons séquencé par séquençage Sanger le gène *LACTB2* dans une cohorte de 50 patients souffrant d'azoospermie non obstructive, ne présentant aucune anomalie chromosomique ni microdélétion du chromosome Y, mais n'avons trouvé aucune variation pathogène.

3.3.1.2. Patient 2

La profondeur moyenne obtenue dans le cadre du WGS pour ce patient est de 15,4X. L'analyse des variants de structure par le pipeline utilisant BreakDancer a permis d'isoler les deux points de cassure en 1p22.2 et en 18q23. La visualisation sur le logiciel IGV de ces deux points a confirmé des défauts d'alignements pour certains reads de cette région (Figure 68).



Figure 68 : Visualisation par IGV des points de cassures sur les chromosomes 1 et 18 chez le patient 2. Les flèches vertes se réfèrent à des paires mal alignées dont le partenaire s’aligne sur le chromosome 18. Les flèches bleues se réfèrent à des paires mal alignées dont le partenaire s’aligne sur le chromosome 1.

L’amplification des deux fragments de jonction a été vérifiée au LabChip (Figure 69a). Le séquençage des fragments de jonction par séquençage Sanger a permis de préciser les positions des cassures en chr1:88626673_88626672 et en chr18:77504599_77504600 (Figure 69b). On note qu’une base T (représentée en vert) a été insérée sur le dérivé 1. Le mécanisme de réparation le plus probable pour cette translocation est la jonction d’extrémités non homologues (Non-Homologous End-Joining ou NHEJ).

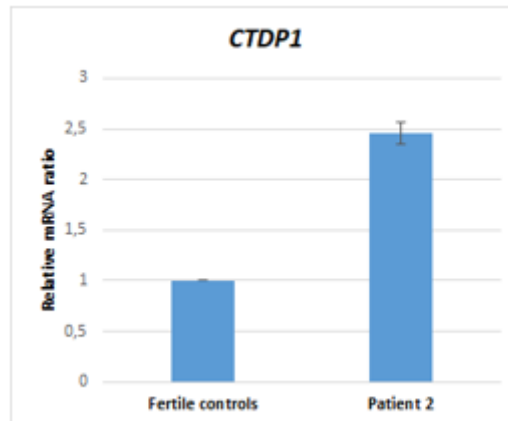


Figure 70 : Etude de l'expression du gène *CTDP1* dans le tissu sanguin chez le patient 3 par rapport aux témoins fertiles.

L'analyse WES réalisée pour le patient 2 avec une couverture moyenne de 97,8% à 30X, n'a détecté aucun SNV pathogène. Il est à noter que le gène *CTDP1* interrompu par la translocation chez ce patient a été séquencé avec une couverture de 100% à 40X et n'a pas montré de variant ponctuel pathogène supplémentaire.

3.3.1.3. Patient 3

La profondeur moyenne obtenue dans le cadre du WGS pour ce patient est de 15,23X. Comme pour le patient 1, l'analyse des variants de structure proposés par le pipeline utilisant BreakDancer n'a pas permis d'isoler les points de cassure potentiels. Ceci peut s'expliquer par l'implication de la région d'hétérochromatine constitutive du chromosome 9 dont les séquences sont mal représentées dans le génome de référence. La marche sur le chromosome 21 par hybridation *in situ* en fluorescence a permis de trouver que le point de cassure se situait entre la sonde RP1-265B9 et la sonde RP11-113F1, soit entre les positions chr21:42795106 et chr21:43635171 (Figure 71).

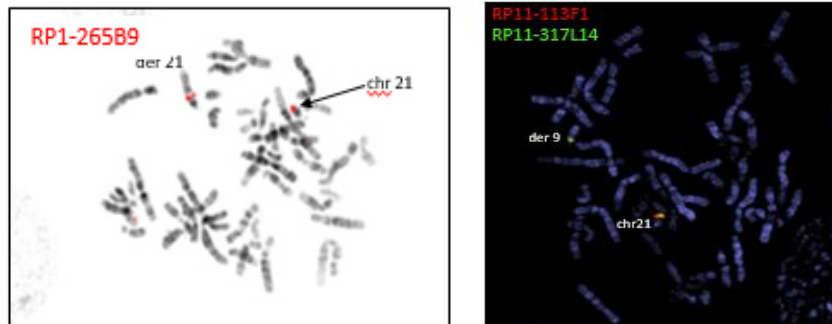


Figure 71 : Images d'hybridation in situ en fluorescence chez le patient 3 pour les sondes : a- RP1-265B9 b- RP11-113F1 (rouge) et RP11-317L14 (vert). Le point de cassure est situé entre RP1-265B9 et RP11-317L14.

La délimitation de cet intervalle permet ensuite de balayer la région manuellement sur le logiciel de visualisation IGV à la recherche du point de cassure. Une zone avec quelques reads discordants est trouvée (Figure 72). La visualisation de cette zone a permis de situer la cassure à la position chr21:41593113-41593114 (Hg38).

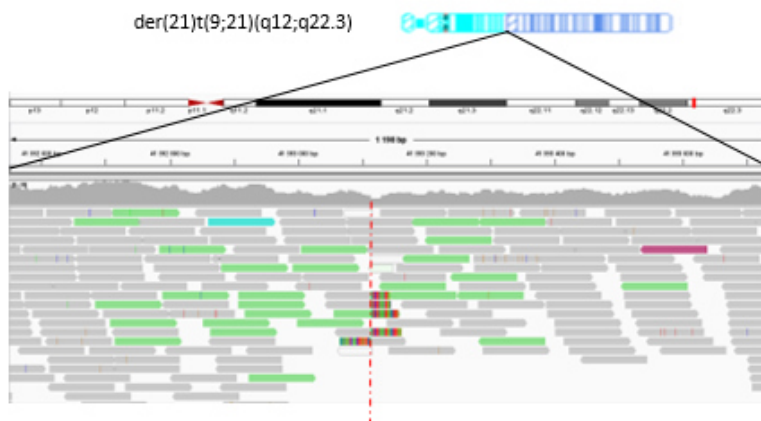


Figure 72 : Visualisation par IGV du point de cassure sur les chromosomes 21 chez le patient 3 (points rouges). Les flèches vertes se réfèrent à des paires mal alignées dont le partenaire s'aligne sur le chromosome 9.

Après amplification par PCR (Figure 73a) et séquençage Sanger des fragments de jonction (Figure 73b), nous avons pu préciser la position du point de cassure au niveau du chromosome 21:41593114_41593115 (Hg38). Cependant, pour le chromosome 9, il n'a pas été possible de préciser la position exacte du point de cassure puisque la séquence correspond à une séquence satellite de type ALR / alpha.

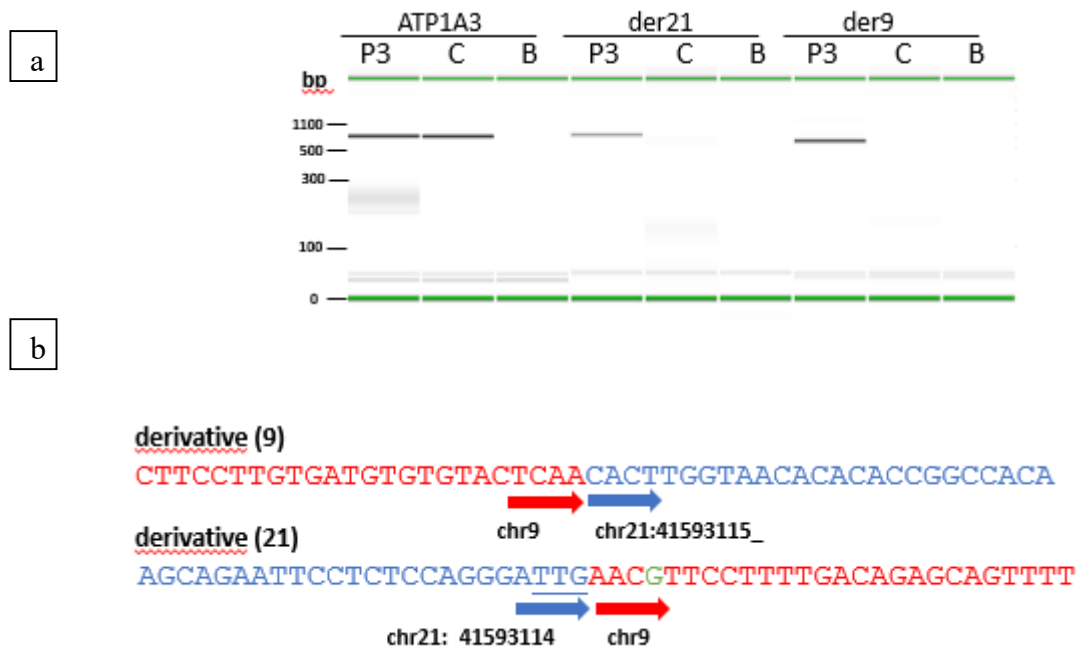


Figure 73 : a- Amplification des fragments de jonction des points de cassures chez le patient 3 par PCR (visualisation Labchip) pour le der (9), le der (21) et une séquence de contrôle (ATP1A3 exons 7-8). B: blanc; C: contrôle sans réarrangement chromosomique, P : patient. b- Séquences des fragments de jonction et coordonnées de points de cassures déterminées par le WGS et le séquençage de Sanger. En rouge : séquence de référence sur le chromosome 9; en bleu séquence de référence sur le chromosome 21, en vert une transversion de C en G. Les bases soulignées correspondent à la micro-homologie. Les flèches indiquent l'orientation du brin de référence.

Le point de cassure sur le chromosome 21 ne perturbe aucun gène. L'exploration de la région voisine des points de cassures dans différentes bases de données ne nous a pas conduit à un gène candidat qui pourrait expliquer le phénotype du patient 3 par effet de position.

L'analyse WES réalisée pour le patient 3 avec une couverture moyenne de 98,7% à 30X, a montré deux variants faux-sens dans le gène *DNAH5* (NM_001369.2) : La première transversion c.13175T>G conduit à la substitution d'une phénylalanine en cystéine, p.(Phe4392Cys). Ce variant est localisé dans le domaine protéique DHC. Le codon impliqué est conservé jusqu'à la drosophile et il existe une grande différence physico-chimique entre Phe et Cys (score de Grantham: 205 [0-215]). Ce variant est rapporté dans les bases de données de sujets contrôles avec une faible fréquence allélique (0.042 % dans gnomAD) et est recensé comme « de signification indéterminée » dans ClinVar. Il a été classé comme délétère par SIFT, PolyPhen2 et MutationTaster. La seconde transversion (*DNAH5*):c.5338A>T conduit à la substitution d'une méthionine en leucine, p.(Met1780Leu). Cette variation implique un acide

aminé conservé jusqu'au Xenope et il est localisé dans le domaine protéique Dynein Heavy Chain 2. Il a été prédit comme toléré par la base de données SIFT, probablement délétère par PolyPhen-2 et délétère par MutationTaster. Il a été signalé deux fois dans la base de données gnomAD. Ces 2 variants ont été classés comme des variants de signification indéterminée.

3.4. Discussion

Dans cette étude, nous avons combiné des techniques de cytogénétique et génétique moléculaire incluant le séquençage à haut débit afin d'identifier le mécanisme moléculaire responsable de l'infertilité chez trois hommes porteurs de RCAE. Pour les 3 patients, nous avons pu définir précisément la localisation et la séquence de chaque point de cassure. Cependant, il convient de noter que cette étape était difficile lorsqu'une région contenant de l'hétérochromatine était impliquée dans le réarrangement des patients.

Nous avons mis en évidence chez le patient 1 un nouveau gène candidat qui pourrait être impliqué dans l'infertilité masculine. En effet, l'expression du gène *LACTB2* est fortement diminuée, de 80%, chez le patient 1. Nous faisons l'hypothèse d'un probable effet de position. Cependant, les contrôles infertiles ont également montré une diminution modérée, de 30%, de l'expression de *LACTB2*, suggérant que sa régulation peut être affectée par des troubles de la spermatogenèse. *LACTB2* code pour une protéine de 288 acides aminés appelée endoribonucléase *LACTB2* ou protéine sécrétoire testiculaire Li 23 qui est fortement exprimée dans les testicules. *LACTB2* est une protéine monomérique soluble située dans la matrice mitochondriale, elle est impliquée dans le renouvellement de l'ARN mitochondrial. Cette protéine est essentielle pour la fonction mitochondriale dans les cellules humaines. Une diminution de l'expression de *LACTB2* dans les cellules cultivées, a entraîné une accumulation modeste de transcrits mitochondriaux correspondants induisant une déformation morphologique cellulaire et la mort cellulaire (160). De plus, le modèle souris mutant de ce gène a montré un développement anormal de ses vésicules séminales (161). Toutes ces données indiquent que cette protéine est essentielle au fonctionnement des mitochondries et à la vitalité cellulaire notamment dans la spermatogenèse. Un examen approfondi de la littérature a montré que *LACTB2* n'a été lié jusqu'à présent à aucune maladie génétique constitutionnelle. Dans ce contexte, ce gène peut être considéré comme un nouveau gène candidat pour l'infertilité masculine, en particulier dans le contexte d'une azoospermie non obstructive. L'identification de patients présentant un défaut dans la séquence et l'expression de *LACTB2* contribuera évidemment à clarifier son rôle dans l'infertilité masculine.

Pour le patient 2, le point de cassure interrompt le gène *CTDPI*. Ce gène, fortement exprimé dans le testicule, est impliqué dans un syndrome récessif associant une cataracte congénitale, une dysmorphie faciale et une neuropathie (CCFDN, MIM # 604168). Les patients CCFDN de sexe masculin peuvent également présenter de petits testicules et un hypogonadisme. L'analyse des données WES chez le patient 2 n'a révélé aucun variant pathogène supplémentaire de *CTDPI*. De plus, les études d'expression de *CTDPI* chez ce patient ont montré un niveau d'ARNm plus élevé par rapport aux témoins fertiles normaux. Ces éléments ne sont pas en faveur de l'implication du gène *CTDPI* pour expliquer l'infertilité chez ce patient.

Chez le patient 3 deux variants faux-sens dans *DNAH5* ont été identifiés : p. (Phe4392Cys) et p. (Met1780Leu). *DNAH5* appartient à la famille des protéines dynéines qui sont des complexes de protéines motrices associés aux microtubules et composés de plusieurs chaînes lourdes, légères et intermédiaires. Les chaînes lourdes de la dynéine (DHC) sont responsables de la production de force et de l'activité ATPase et contiennent un domaine catalytique hautement conservé avec 4 motifs consensus de boucle P impliqués dans la liaison des nucléotides. La structure axonémique centrale du cil mobile et de la queue du sperme a la même disposition microtubulaire ultrastructurale. On s'attend donc à ce que les défauts génétiques des cils mobiles aient également un effet sur la formation de la queue du sperme (120). Les mutations dans *DNAH5* sont responsables de la dyskinésie ciliaire primaire (PCD, MIM # 608644). Ce gène a également été décrit une fois comme étant associé à l'asthénospermie chez les hommes infertiles (162) bien qu'une étude récente suggère que *DNAH5* ne se colocalise pas avec l' α -tubuline le long de l'axonème du sperme (163). La classification des deux variants identifiés reste donc une tâche difficile. Ces deux variants devraient être testés chez les parents pour voir s'ils sont sur un même allèle ou bien sur deux allèles différents. Malheureusement, les parents n'étaient pas disponibles pour l'étude.

Draft Manuscript For Review. Reviewers should submit their review at
<http://mc.manuscriptcentral.com/molehr>

**Deciphering balanced translocations in infertile males by
 Next Generation Sequencing identifies two candidate genes,
 LACTB2 and DNAH5, for spermatogenesis disorders**

Journal:	<i>Molecular Human Reproduction</i>
Manuscript ID	MHR-20-0368
Manuscript Type:	Original Research
Date Submitted by the Author:	16-Oct-2020
Complete List of Authors:	<p>Yammine, Tony; Saint Joseph University, Medical Genetics Unit (UGM), Faculty of Medicine; Institut Neuromyogène, Equipe Métabolisme énergétique et développement neuronal, CNRS UMR 5310, INSERM U1217, Université Lyon 1</p> <p>Reynaud, Nicolas; Hospices Civils de Lyon, Service de Génétique; Hospices Civils de Lyon, Service de Médecine de la Reproduction, Hôpital Femme Mère Enfant</p> <p>Lejeune, Hervé; Hospices Civils de Lyon, Service de Médecine de la Reproduction, Hôpital Femme Mère Enfant; Hospices Civils de Lyon, Laboratoire d'hormonologie et endocrinologie Moléculaire</p> <p>Diguet, Flavie; Hospices Civils de Lyon, Service de Génétique</p> <p>Rollat-Farnier, Pierre-Antoine; Hospices Civils de Lyon, Service de Génétique; Hospices Civils de Lyon, Cellule Bioinformatique</p> <p>Labalme, Audrey; Hospices Civils de Lyon, Service de Génétique</p> <p>Plotton, Ingrid; Hospices Civils de Lyon, Service de Médecine de la Reproduction, Hôpital Femme Mère Enfant; Hospices Civils de Lyon, Laboratoire d'hormonologie et endocrinologie Moléculaire; University Claude Bernard Lyon 1, Unite INSERM 1208</p> <p>Farra, Chantal; Saint Joseph University, Medical Genetics Unit (UGM), Faculty of Medicine; Hotel-Dieu De France, Department of Genetics</p> <p>Sanlaville, Damien; Hospices Civils de Lyon, Service de Génétique; Institut Neuromyogène, Equipe Métabolisme énergétique et développement neuronal, CNRS UMR 5310, INSERM U1217, Université Lyon 1</p> <p>Choueiry El Khoury, Eliane; Saint Joseph University, Medical Genetics Unit (UGM), Faculty of Medicine; Lebanese American University, Department of Human Genetics, Gilbert and Rose-Marie Chagoury School of Medicine</p> <p>Schluth-Bolard, Caroline; Hospices Civils de Lyon, Service de Génétique; Institut Neuromyogène, Equipe Métabolisme énergétique et développement neuronal, CNRS UMR 5310, INSERM U1217, Université Lyon 1</p>
Key Words:	LACTB2, male infertility, translocation, DNAH5, genome sequencing

SCHOLARONE™
Manuscripts

1 **Deciphering balanced translocations in infertile males by Next Generation Sequencing identifies**
2 **two candidate genes, *LACTB2* and *DNAH5*, for spermatogenesis disorders**

3

4

5 T. Yammine^{1,2§}, N. Reynaud^{3,4§}, H. Lejeune^{4,7}, F. Diguët³, PA. Rollat-Farnier^{3,5}, A. Labalme³, I.
6 Plotton^{4,7,8}, C. Farra^{1,6}, D. Sanlaville^{2,3}, E. Choueiry El Khoury^{1,9}, C. Schluth-Bolard^{2,3*}

7 § These authors should be regarded as joint first authors.

8 ¹ Medical Genetics Unit (UGM), Faculty of Medicine, Saint Joseph University, Beirut, Lebanon.

9 ² Institut Neuromyogène, Equipe Métabolisme énergétique et développement neuronal, CNRS UMR
10 5310, INSERM U1217, Université Lyon 1, Lyon, France.

11 ³ Hospices Civils de Lyon, Service de Génétique, Bron, France.

12 ⁴ Service de Médecine de la Reproduction, Hôpital Femme Mère Enfant, Hospices Civils de Lyon, Bron,
13 France.

14 ⁵ Cellule Bioinformatique, Hospices Civils de Lyon, Bron, France.

15 ⁶ Department of Genetics, Hotel Dieu de France Medical Center, Beirut, Lebanon.

16 ⁷ Laboratoire d'hormonologie et endocrinologie Moléculaire, Hospices Civils de Lyon, Bron, France.

17 ⁸ Unite INSERM 1208, Université Lyon 1, Lyon, France.

18 ⁹ Department of Human Genetics, Gilbert and Rose-Marie Chagoury School of Medicine, Lebanese
19 American University, Byblos, Lebanon

20

21 Running title: Next Generation Sequencing in male infertility

22 Keywords: Male Infertility, translocation, Genome sequencing, exome, *LACTB2*, *DNAH5*.

23

24 **Correspondence to Dr Caroline Schluth-Bolard, Service de Génétique, Hospices Civils de Lyon,*
25 *Groupement Hospitalier Est, 59 Bd Pinel, 69677 BRON CEDEX, FRANCE, Tel : 33 4 72 12 96 40 ,*
26 *Fax : 33 4 72 12 97 10, Email : caroline.schluth-bolard@chu-lyon.fr*

27

28 **Abstract**

29 Background: Male infertility affects about 7% of the general male population. Balanced structural
30 chromosomal rearrangements are observed in 0.4 to 1.4% of infertile male and are considered as a
31 well-established cause of infertility. However, underlying pathophysiological mechanisms need still to
32 be clarified.

33 Objectives: We used a combination of standard and high throughput cytogenetic and molecular
34 technologies in order to identify the molecular mechanisms responsible for the infertility in three men
35 carriers of balanced translocation and spermatogenesis defect.

36 Material and Methods: FISH and whole genome paired-end sequencing were used to characterize
37 translocation breakpoints at the molecular level. Exome sequencing was performed in order to exclude
38 any molecular event independent from the chromosomal rearrangement. Expression study of candidate
39 genes was performed on blood RNA.

40 Results: All translocation breakpoints were characterized in the three patients. We identified a
41 position effect on *LACTB2* gene in patient 1, *CTDP1* gene disruption in patient 2 and two single
42 nucleotide variations in *DNAH5* gene in patient 3.

43 Discussion: Molecular characterization of chromosomal rearrangements in infertile male may uncover
44 additional alterations, beside the mechanical effect of the rearrangement on meiosis, such as gene
45 disruption or position effect. Moreover, chromosomal rearrangements may be fortuitously associated
46 with single nucleotide variations that could explain the variability of the impact of chromosomal
47 rearrangement on spermatogenesis.

48 Conclusion: This study confirms the relevance of combining different cytogenetic and molecular
49 techniques to investigate patients with spermatogenesis disorders and pinpoints a potential implication
50 for *LACTB2* and *DNAH5* in male infertility.

51

52

53 **Introduction:**

54 As per to the World Health Organization's definition, human infertility is the failure for a
55 sexually active couple to conceive after one year of unprotected relations (Dohle et al., 2005). Around
56 15% of reproductive age couples experience fertility problems with about half of the cases imputed to
57 male factors (Dohle et al., 2005). The causes of male infertility entail abnormal sperm production or
58 function, problems with sperm release through the ejaculatory ducts, overexposure to certain
59 environmental factors, or damage related to cancer (Coutton et al., 2015; Poongothai et al., 2009). Up
60 to 40% of male infertility remains of unknown origin (Poongothai et al., 2009).

61 In nearly one in 40 men, infertility is associated with genetic factors such as chromosomal,
62 monogenic, multifactorial disorders and defect in mitochondrial DNA (Poongothai et al., 2009). The
63 most recurrent genetic causes of male infertility are chromosomal rearrangements, Y chromosome
64 microdeletions and cystic fibrosis transmembrane conductance regulator (*CTFR*) mutations (Coutton
65 et al., 2015). These altogether are found in only 15-30% of all studied patients. The genetic landscape
66 of infertility is indeed highly complex and uncovering remaining genetic etiologies requires large scale
67 researches (Krausz and Riera-Escamilla, 2018).

68 Balanced Chromosomal Rearrangements (BCR), *i.e.* translocation or inversion, are identified
69 in 0.5-1% of patients with severe oligozoospermia or azoospermia. They may cause subfertility by
70 doubling the risk of miscarriages. Indeed, during meiotic pairing, the normal and derivative
71 chromosomes form a quadrivalent structure leading to the occurrence of unbalanced spermatozoa with
72 percentages ranging from 18.6% to 80.7% (Vozdova et al., 2013). The relative proportions and
73 viability of ensuing embryos depend on the DNA breaks position and recombination sites. The
74 mechanism of how such chromosomal abnormalities may lead to low sperm counts is believed to be
75 mainly related to mechanical issues impacting male meiosis but has not been rigorously investigated
76 (Benet et al., 2005). Moreover, correlations between BCR and spermatic phenotype is still debated
77 (Dul et al., 2012, 2010; Gekas et al., 2001; Vincent et al., 2002). When a structural chromosomal

78 rearrangement is identified in a patient with a spermatogenesis defect, it is usually considered as the
79 cause of the infertility, thus ending subsequent investigations.

80 Recently, Next Generation Sequencing (NGS) technologies have been applied to characterize
81 BCR breakpoints in patients with abnormal phenotypes, in particular intellectual disability and
82 congenital malformations (Schluth-Bolard et al., 2019). Using different DNA preparation protocols
83 and sequencing chemistry, several teams (Nilsson et al., 2017; Ordulu et al., 2016; Redin et al., 2017;
84 Talkowski et al., 2011) demonstrated that NGS procedures are successful in mapping the genetic
85 breakpoints of such rearrangements at the base-pair level (Schluth-Bolard et al., 2013), providing
86 therefore more information about the mechanisms of their occurrence and allowing the identification
87 of candidate disrupted genes. This method was rarely exploited in male infertility. However, some
88 studies showed that beside the mechanical effect of a rearrangement, a chromosomal aberration may
89 present a direct impact on the spermatogenesis through the disturbance of a specific gene. Thereby,
90 Mouka et al. (Mouka et al., 2017) demonstrated the implication of *SYCP3* gene disruption in the
91 phenotype of a patient with azoospermia and a complex chromosomal rearrangement. Moreover,
92 Schilit et al. (Schilit et al., 2020) described *SYCP2* dysregulation by a BCR in a patient with severe
93 oligozoospermia.

94 Based on these findings, we conducted a combined strategy of paired-end whole genome
95 sequencing (WGS) and exome sequencing in three men with BCR and defective spermatogenesis. Our
96 aim was to characterize the breakpoints at the molecular level and to identify the genes affected either
97 directly due to disruption or indirectly through position effect. We added exome sequencing in order to
98 rule out the presence of any additional event unrelated to the chromosomal rearrangement.

99

100 **Patients and methods**

101 **Patients**

102 Three male patients with history of infertility were included in this study. Consanguinity was
103 excluded in all three families.

104 **Patient 1:** A 32-year-old man was referred for infertility after 5 years of sexual intercourse without
105 conception. Semen analysis revealed azoospermia. Sexual development, medical history, physical

106 examination and serum hormone levels (testosterone, cortisol, T3, T4, TSH, estradiol, FSH, LH,
107 prolactin and Inhibin-B) were normal. Standard blood karyotype showed a reciprocal translocation
108 between chromosomes 8 and 22: 46,XY,t(8;22)(q13;p11) (Figure 1a). Chromosomal micro-array
109 analysis did not show any chromosomal imbalances.

110 **Patient 2:** A 46-year-old patient was referred because of severe oligo-astheno-teratozoospermia.
111 Standard blood karyotype showed a reciprocal translocation between chromosomes 1 and 18:
112 46,XY,t(1;18)(p21;q12) (Figure 2a). The couple underwent two ICSI (intra-cytoplasmic sperm
113 injection), with only one embryonic transfer but no pregnancy.

114 **Patient 3:** A 26-year-old patient was included because of a severe oligo-asthenozoospermia and
115 important teratozoospermia, presenting with a reciprocal translocation between chromosomes 9 and
116 21: 46,XY,t(9;21)(q12;q22.3) (Figure 3a). He additionally presented with type I diabetes and
117 hypertension. Serum hormone level were as followed: FSH 3UI/L, LH 5.88 UI/L, PRL 20.3µg/L,
118 testosterone 18nmol/L, inhibine B 5.1nmol.L, AMH 41.2pmol/L. The couple underwent two ICSI, that
119 both failed due to embryos fragmentation before blastocyst stage.

120

121 **Whole genome sequencing and analysis**

122 Genomic libraries were prepared following Illumina TruSeq protocol (Illumina, San Diego,
123 CA, USA) using 3 µg of genomic DNA as previously described (Schluth-Bolard et al., 2019). DNA
124 libraries were then paired-end sequenced, for two samples simultaneously, on Illumina NextSeq500 (2
125 x 101 bp run mode) using high output Flow Cell 300, yielding a mean sequencing depth between 10
126 and 15X, which is sufficient to detect structural variations but not single nucleotide variants (SNV)
127 with an acceptable sensitivity (Meynert et al., 2013). The bioinformatics analysis of sequencing data
128 was performed based on an *in house* pipeline (Schluth-Bolard et al., 2019) including alignment to the
129 Human Reference Genome (Build GRCh38/hg38) with BWA-MEM version 0.7.10, identification of
130 structural variations using BreakDancer version 1.4.5 (Chen et al., 2009) and detection of copy-
131 number variations using ERDS version 1.1 (Zhu et al., 2012). The generated outputs were then
132 annotated using an in-house program, Svagga. Integrative Genomics Viewer (IGV) was used to
133 validate the identified variants and UCSC Genome Browser was used to visualize their genomic

134 context. To assess a putative position effect, genes located in the same topologically associated
135 domain (TAD) as the breakpoint were considered. TADs were defined for GM12878 cells (Rao et al.,
136 2014) owing to 3D Genome Browser (hg38).

137 **PCR verification and Sanger sequencing: Amplification of junction fragments**

138 All identified genomic breakpoints were verified by standard PCR followed by Sanger
139 sequencing, as previously described (Schluth-Bolard et al., 2019) (primer sequences in Supplementary
140 table I). The sequences were aligned to the reference genome (Build GRCh38/hg38) using the UCSC
141 BLAT program. The BLAT tool allowed the identification of the exact location of the breakpoints. It
142 enabled to determine the origin of each part of the junction fragment of each chromosome involved in
143 the translocation, as well as their relative orientation (parallel or antiparallel direction). It helped also
144 to identify the possibility of insertion of one or more bases at the sites of the breakpoints and to deduce
145 a possible repair mechanism.

146

147 **Exome sequencing**

148 After sonication of 1µg of genomic DNA (Covaris M220), library preparation was performed
149 using the SeqCap EZ MedExome kit (Roche Diagnostics, Mannheim, Germany) according to the
150 manufacturers' instructions. Paired-end sequencing was performed on a NextSeq500, (Illumina, San
151 Diego, CA, USA). Genomic alignment against the hg19/GRCh37 assembly and variant calling were,
152 respectively, done with BWA-MEM v.0.7.12 and GATK Haplotype Caller v.3.4 while QC parameters
153 were evaluated using DeCovA (Dimassi et al., 2015). Only highly confident variants were kept for the
154 analysis (total depth 49; no strand bias; mosaicism > 10%). An in house script filter was used to filter
155 the identified variants according to the hypothesized possible inheritance patterns (autosomal
156 recessive, de novo, X-linked inheritance). Intronic variants as well as variants with a Minor Allele
157 Frequency (MAF) > 1% in the ExAC database were excluded in order to only keep rare exonic and
158 splicing variants. Analysis of the remaining variants focused on the genes that were linked to human
159 monogenic male infertility and azoospermia (Supplementary Table II) (Oud et al., 2019). Additionally,
160 the possible impact of the selected variants on the protein structure and function was predicted based

161 on *in silico* methods (Sift, PolyPhen-2 and Mutation Taster). Moreover, for each identified variant the
162 possible testicular expression was checked in GTE_x.

163

164 **Expression studies**

165 RNA was extracted from whole blood using PAXgene kit according to the manufacturer's
166 instruction (Qiagen, Hilden, Germany). Reverse transcription was achieved using 500 ng RNA with
167 the Superscript II RT kit (Invitrogen, Life Technologies, Carlsbad, California, USA). Real-time PCR
168 was performed using 1/20 dilutions of the obtained cDNAs with the QuantiTect SYBR Green PCR kit
169 (Qiagen, Hilden, Germany) on a LightCycler 480 (Roche Applied Sciences, Indianapolis, IN, USA) in
170 triplicate. Primers were designed using Primer 3 and checked for specificity with Primer Blast (primer
171 sequences in supplemental table 1). *ACTB* was used as a reference gene. Samples included patients'
172 cDNA as well as three sex-matched normal fertile controls and two azoosperm patients, additionally to
173 the RT-negative products. Finally, data were analyzed by relative quantification according to the $2^{-\Delta\Delta C_t}$
174 method (Livak and Schmittgen, 2001).

175 **Fluorescent in situ hybridization (FISH)**

176 For a high resolution characterization of chromosomal breakpoints in heterochromatic region,
177 FISH analyses were conducted in patients 1 and 3. BACs were chosen from UCSC databases (RP11-
178 152C15, RP11-11K9, RP1-265B9, RP11-113F1) and the technique was performed as described by
179 Romana *et al.* (Romana et al., 1994). The DNA was fluorescein or rhodamine labeled by nick-
180 translation. Probes were coprecipitated with human Cot-1 DNA (Invitrogen, Carlsbad, CA, USA) and
181 then resuspended in hybridization buffer (50% formamide). After denaturation, overnight
182 hybridization and post-hybridization wash, slides were DAPI counterstained.

183

184 **Results**

185

186 **Identification of translocation breakpoints at the molecular level**

187 Paired-end genome sequencing was first used for the characterization of translocation breakpoint at the
188 molecular level.

189 Patient 1

190 WGS in Patient 1 was achieved with a 13.9X mean depth. WGS data analysis using
191 Breakdancer was not able to find the translocation t(8;22) breakpoints. This may be due to the
192 involvement in the rearrangement of the heterochromatin on the short arm of chromosome 22. We
193 performed thus FISH analyses in order to define an interval with lower and upper bounds. The lower
194 one was the position on the reference genome of the last probe located on the derivative der(8) (probe
195 located upstream of the breakpoint) while the upper bound represented the first probe located on the
196 derivative der(22) (probe located downstream of the breakpoint). This allowed the localization of the
197 breakpoint on chromosome 8 between BAC RP11-152C15 (chr8:69969490_70150856, hg38) and
198 BAC RP11-11K9 (chr8:71222159_71383240, hg38), (Supplementary Figure S1). The data obtained
199 by WGS were then reassessed manually on the basis on IGV visualization to confirm these results
200 (Supplementary Figure S1). We were able to identify the breakpoint at 8q13.3 while the mate reads
201 were mapped either to the short arm of chromosome 21 or to an unplaced sequence on chromosome 22
202 (chr22_K1270733v1_random). An additional FISH technique confirmed that the NORs (Nucleolus
203 Organizer Regions) of chromosome 22 were translocated on the derivative chromosome 8, and that the
204 translocated segment of chromosome 8 long arm was very close to the heterochromatic
205 pericentromeric region on the derivative chromosome 22 (Supplementary Figure S1). Upon PCR
206 amplification (Supplementary Figure S1) and Sanger sequencing of the junction fragments (Figure
207 1b), we specified precisely the positions of the breaks at chr8 (hg38):70521071_70521076 and
208 chr22_K1270733v1_random: 158856_158864. The latter position corresponded to a segmental
209 duplication, also present on the short arm of chromosome 21 (chr21:8241828_8241836). The junction
210 on derivative 8 displayed a particular structure with a 5-nucleotides microhomology (AAAAT), an
211 insertion of 44 nucleotides of unknown origin
212 (CAGAAAAAATCTACTACATAACAACATTCTCTAGTATGTACTT), and a duplication of the
213 sequences around the breakpoints (chr8:70520702_70521076 and
214 chr22_KI270733v1_random:158909_158861). The junction sequence on derivative 22 showed a 5

215 nucleotides microhomology (TTTTA). All these features were suggestive of the occurrence of a
216 replication-based mechanism (FoSTeS, MM-BIR) explaining this complex junction (Zhang et al.,
217 2009).

218 The identified breakpoints did not disrupt any gene. Exploring genes within the disrupted
219 TAD on chromosome 8 allowed us to identify candidate genes expressed in testis, including *NCOA2*,
220 *TRAM1*, *LACTB2* and *XKR9* (Figure 1c).

221

222 Patient 2

223 The mean depth of the WGS data in Patient 2 was evaluated to 15.40X. Data analysis allowed
224 us to isolate two different breakpoints at 1p22.2 and 18q23. The amplification by standard PCR
225 (Supplementary Figure S2) followed by sequencing of the 2 junction fragments confirmed these
226 results and helped us to identify precisely the positions of the breakpoints on
227 chr1:88160989_88160990 and chr18:79744599_79744600 (hg38) (Figure 2b). The molecular
228 signature of the identified breakpoints suggests Non-Homologous End-Joining (NHEJ) as the most
229 likely repair mechanism. The breakpoint on chromosome 1 disrupted the sequence of an anti-sens
230 RNA *PKN2-AS1* while the breakpoint on chromosome 18 disrupted the intron 12/12 in *CTDPI* gene.
231 Analyzing the disrupted TADs on both chromosomes 1 and 18 did not reveal any gene that could
232 explain the disease by position effect (Supplementary Figure 4).

233

234 Patient 3

235 As for Patient 1, the first assessment of the WGS data, with a mean depth of 15.23X, was not
236 able to identify the implicated breakpoints in Patient 3. This may be due to the involvement of the 9q
237 constitutive heterochromatin in the identified rearrangement. FISH analysis was then performed in
238 order to characterize the chromosome 21 breakpoint position. The latter was delimited between the
239 BAC probes RP1-265B9 (chr21:4126518_41423179, hg38) and RP11-113F1
240 (chr21:42214063_42396178, hg38) (Supplementary Fig. S3) in 21q22.3. IGV visualization allowed
241 the confirmation of this finding. PCR amplification (Supplementary Figure S3) followed by Sanger
242 sequencing specified precisely the positions of the breakpoints at chr21:41593114_41593115, hg38

243 (Figure 3b). However, for chromosome 9, it was not possible to point out the exact position of the
244 breakpoint since the sequence corresponded to an ALR/alpha satellite repeats. According to the
245 molecular signature at breakpoints the most likely repair mechanism was NHEJ. The breakpoint on
246 chromosome 21 did not disrupt any gene. Exploring genes within the disrupted TAD on chromosome
247 21 did not lead us to any candidate gene that could explain the disease in Patient 3 by position effect
248 (Supplementary Figure 4).

249

250 Exome sequencing

251 Exome sequencing was performed for the three patients in order to rule out the involvement of
252 any pathogenic SNV in their phenotypes, *i.e.* a supplementary event independent from the identified
253 chromosomal rearrangement.

254 Exome sequencing analysis in Patient 1 yielded a mean coverage of 98% at 30X. No
255 pathogenic SNVs were detected in genes implicated in infertility in this patient.

256 Similarly, Exome sequencing analysis performed with a mean coverage of 97.8% at 30X, for
257 Patient 2, did not detect any pathogenic SNV. Of note the *CTDPI* gene disrupted by the translocation
258 in this patient was sequenced with a coverage of 100% at 40X and did not show any additional
259 pathogenic point variant.

260 Exome sequencing ES data in Patient 3 (mean coverage of 98.7% at 30X) showed two
261 missense variants in the *DNAH5* gene (NM_001369.2): c.13175T>G p.(Phe4392Cys) and c.5338A>T
262 p.(Met1780Leu) (Figure 3c). The first variant, p.(Phe4392Cys) was located in the DHC protein
263 domain. The implicated codon Phe4392 was conserved until *Drosophila* fly (among 11 species from
264 Ensembl Compara database) and there is a large physico-chemical difference between Phenylalanine
265 and Cysteine (Grantham score: 205 [0-215]). This variant was reported in the databases of control
266 subjects with a low allelic frequency (0.042% in gnomAD) and was listed as a variant of unknown
267 significance in ClinVar. It was classified as deleterious by SIFT and disease causing by
268 MutationTaster. The second variant, p.(Met1780Leu) involved a conserved amino acid until *Xenopus*
269 (among 11 species from Ensembl Compara database) and was located in the Dynein Heavy Chain
270 protein domain 2. It was predicted as tolerated by SIFT database, probably damaging by PolyPhen-2

271 and disease causing by MutationTaster. It was reported twice in the gnomAD database. These variants
272 were thus classified as variants of unknown significance.

273

274 **Expression studies**

275 In order to assess the hypothetical functional consequences of the identified chromosomal
276 rearrangements in Patients 1 and 2, we performed quantitative expression studies.

277 In Patient 1, four genes in the vicinity of the breakpoint could possibly have been impacted by
278 position effect. Indeed, the translocation repositioned chromosome 8 euchromatic region near the
279 pericentromeric short arm sequences of chromosome 22; the heterochromatin organization might thus
280 have probably spread into the chromosome 8 region affecting its expression. As *XKR9* is not expressed
281 in blood, its expression could not have been further evaluated. Expression level of *NCOA2* and
282 *TRAMI* was unchanged in patient 1 compared to controls (Supplementary Figure 1) whereas *LACTB2*
283 expression revealed a near 80% decrease in Patient 1 compared to normal fertile controls (Figure 1d).
284 Furthermore, infertile controls showed a 30% decrease in *LACTB2* mRNA (Figure 1d), suggesting
285 that *LACTB2* expression may also be affected by spermatogenesis disorders. However, the decrease
286 was more important in Patient 1 than in infertile controls, suggesting a functional impact of the
287 breakpoint on *LACTB2* gene in Patient 1.

288 In patient 2, *CTDPI* was disrupted by one of the translocation breakpoints. Expression studies
289 for this gene did not reveal any decrease compared to normal fertile controls (Figure 2c).

290

291 **Discussion**

292 We combined, in this study, different standard and high throughput cytogenetic and molecular
293 technologies in order to identify the molecular mechanism responsible for the infertility in three men
294 carriers of BCR.

295 We were able, for all 3 patients, to determine precisely the positions and the sequences for
296 each breakpoint, highlighting the importance of combining different cytogenetic and molecular
297 techniques. It is moreover noteworthy to mention the challenge encountered when sequencing
298 heterochromatin regions involved in the rearrangements.

299 This study draws attention to a new candidate gene in male infertility. Indeed, *LACTB2*
300 expression showed a substantial decrease in Patient 1 probably due to position effect. Furthermore, a
301 moderate decrease in *LACTB2* expression was noted in infertile controls suggesting an impact of
302 spermatogenesis defects on the regulation of this gene. *LACTB2* encodes a 288 amino acids protein
303 called Endoribonuclease LACTB2 or testicular secretory protein Li 23 which is highly expressed in
304 testis. LACTB2 is a soluble monomeric protein located in the mitochondria matrix, and is involved in
305 the turnover of mitochondrial RNA. It is essential for mitochondrial function in human cells (Levy et
306 al., 2016). The knockdown of *LACTB2* in cultured cells resulted in a modest accumulation of
307 corresponding mitochondrial transcripts inducing a cellular morphological deformation and cell death
308 (Levy et al., 2016). In addition, the mutant mouse model for this gene showed an abnormal
309 development of its seminal vesicles (Sollars et al., 2002). All these data indicate that this protein is
310 essential for mitochondrial function and cell vitality especially in spermatogenesis. A thorough review
311 of the literature showed that *LACTB2* has to date not been linked to any constitutional genetic disease.
312 From that perspective, *LACTB2* should be considered as a new candidate gene for male infertility,
313 particularly in cases of non-obstructive azoospermia. To complete this study, we analyzed *LACTB2* in
314 a cohort of 50 patients with non-obstructive azoospermia by Sanger sequencing. We however did not
315 find any pathogenic variant (data not shown). Further analyses on additional cases would contribute to
316 clarify the role of *LACTB2* in male infertility.

317 For Patient 2, the involvement of the disrupted *CTDPI* gene in the clinical manifestations was
318 ruled out. This gene, highly expressed in testis, is associated with an autosomal recessive syndrome
319 including congenital cataracts, facial dysmorphism and neuropathy (CCFDN, MIM#604168). Male
320 patients with CCFDN may also present with small testes and hypogonadism. Exome sequencing did
321 not reveal any additional pathogenic variant on *CTDPI* in Patient 2. Moreover, *CTDPI* expression
322 studies in this patient showed higher mRNA level compared to normal fertile controls. Altogether,
323 these data ruled out any potential role for *CTDPI* in the phenotype of Patient 2.

324 In Patient 3 two missense variants were identified in *DNAH5*: p.(Phe4392Cys) and
325 p.(Met1780Leu). *DNAH5* belongs to the dynein proteins family that are microtubule-associated motor

326 protein complexes composed of several heavy, light and intermediate chains. The dynein heavy chains
327 (DHC) are responsible for force production and ATPase activity and contain a highly conserved
328 catalytic domain with four P-loop consensus motifs involved in nucleotides binding. The core
329 axoneme structure of both the motile cilium and sperm tail have the same ultrastructural microtubular
330 arrangement. It is thus expected that genetic defects in motile cilia would also have an effect on sperm
331 tail formation (Sironen et al., 2019). Mutations in *DNAH5* are responsible for Primary ciliary
332 dyskinesia (PCD, MIM #608644). This gene has also been described once to be associated with
333 asthenospermia in infertile men (Fliegauf et al., 2005) although a recent study suggests that *DNAH5*
334 does not colocalize with α -tubulin along the sperm axoneme (Whitfield et al., 2019). The classification
335 of both identified variants thus remains a challenging task. It would have been interested to determine
336 the inheritance and the allelic configuration of these variants in Patient 3. However, the parents were
337 unfortunately unavailable for further analyses.

338 Although the candidate genes we identified need further investigations, this study highlights
339 the validity of our approach. Molecular characterization of chromosomal rearrangements in infertile
340 males may uncover further alterations of genes involved in spermatogenesis at the breakpoints, like
341 gene disruption or position effect. Moreover, chromosomal rearrangements may be fortuitously
342 associated with single nucleotide pathogenic variations. These alterations would add to the mechanical
343 effects of BCR and may thus explain the variability of spermatogenesis defects in male carriers of
344 BCR. The study of a larger cohort will be helpful to evaluate the role of breakpoints' pathogenicity
345 and other genomic variations in the spermatogenesis phenotypic variability of male carriers of BCR.

346

347 **DECLARATIONS**

348 **Availability of data and material:**

349 The datasets used and analyzed during the current study are available from the corresponding author
350 upon a reasonable request

351 **Acknowledgments**

352 We express our deepest gratitude and sympathy to the patients for their full cooperation throughout the
353 study. We wish to thank Claire Goursaud for her support and advices in exome analysis.

354 **Authors' contributions**

355 CSB, EC and DS conceived and designed the study. TY wrote the manuscript. CSB, DS, EC and CF
356 provided helpful advices and edited the manuscript. TY, NR, FD, PARF, AL and CSB conducted the
357 experiments and performed data analysis and interpretation. TY, NR, HL and IP performed the
358 recruitment and referral of patients. All authors have read and approved the manuscript.

359 **Competing interests/Disclosures**

360 Authors declare no conflict of interest.

361 **Ethics approval and consents:**

362 Approval to conduct the study was obtained from the Ethics Committee of Hospices Civils de Lyon
363 and Saint Joseph University, Beirut, Lebanon. Patients signed an informed consent for participation
364 and sample collection.

365 **References**

- 366 Benet, J., Oliver-Bonet, M., Cifuentes, P., Templado, C., Navarro, J., 2005. Segregation of
367 chromosomes in sperm of reciprocal translocation carriers: a review. *Cytogenet. Genome Res.*
368 111, 281–290.
- 369 Chen, K., Wallis, J.W., McLellan, M.D., Larson, D.E., Kalicki, J.M., Pohl, C.S., McGrath, S.D.,
370 Wendl, M.C., Zhang, Q., Locke, D.P., Shi, X., Fulton, R.S., Ley, T.J., Wilson, R.K., Ding, L.,
371 Mardis, E.R., 2009. BreakDancer: an algorithm for high-resolution mapping of genomic
372 structural variation. *Nat. Methods* 6, 677–681.
- 373 Coutton, C., Escoffier, J., Martinez, G., Arnoult, C., Ray, P.F., 2015. Teratozoospermia: spotlight on
374 the main genetic actors in the human. *Hum. Reprod. Update* 21, 455–485.
- 375 Dimassi, S., Simonet, T., Labalme, A., Boutry-Kryza, N., Campan-Fournier, A., Lamy, R., Bardel, C.,
376 Elsensohn, M.-H., Roucher-Boulez, F., Chatron, N., Putoux, A., de Bellescize, J., Ville, D.,
377 Schaeffer, L., Roy, P., Mougou-Zerelli, S., Saad, A., Calender, A., Sanlaville, D., Lesca, G.,
378 2015. Comparison of two next-generation sequencing kits for diagnosis of epileptic disorders
379 with a user-friendly tool for displaying gene coverage, DeCovA. *Appl. Transl. Genomics* 7,
380 19–25.
- 381 Dohle, G.R., Colpi, G.M., Hargreave, T.B., Papp, G.K., Jungwirth, A., Weidner, W., EAU Working
382 Group on Male Infertility, 2005. EAU guidelines on male infertility. *Eur. Urol.* 48, 703–711.
- 383 Dul, E.C., Groen, H., van Ravenswaaij-Arts, C.M.A., Dijkhuizen, T., van Echten-Arends, J., Land,
384 J.A., 2012. The prevalence of chromosomal abnormalities in subgroups of infertile men. *Hum.*
385 *Reprod. Oxf. Engl.* 27, 36–43.
- 386 Dul, E.C., van Ravenswaaij-Arts, C.M.A., Groen, H., van Echten-Arends, J., Land, J.A., 2010. Who
387 should be screened for chromosomal abnormalities before ICSI treatment? *Hum. Reprod. Oxf.*
388 *Engl.* 25, 2673–2677.
- 389 Fliegau, M., Olbrich, H., Horvath, J., Wildhaber, J.H., Zariwala, M.A., Kennedy, M., Knowles, M.R.,
390 Omran, H., 2005. Mislocalization of DNAH5 and DNAH9 in Respiratory Cells from Patients
391 with Primary Ciliary Dyskinesia. *Am. J. Respir. Crit. Care Med.* 171, 1343–1349.

- 392 Gekas, J., Thepot, F., Turleau, C., Siffroi, J.P., Dadoune, J.P., Briault, S., Rio, M., Bourouillou, G.,
393 Carré-Pigeon, F., Wasels, R., Benzacken, B., Association des Cytogeneticiens de Langue
394 Francaise, 2001. Chromosomal factors of infertility in candidate couples for ICSI: an equal
395 risk of constitutional aberrations in women and men. *Hum. Reprod. Oxf. Engl.* 16, 82–90.
- 396 Krausz, C., Riera-Escamilla, A., 2018. Genetics of male infertility. *Nat. Rev. Urol.* 15, 369–384.
- 397 Levy, S., Allerston, C.K., Liveanu, V., Habib, M.R., Gileadi, O., Schuster, G., 2016. Identification of
398 LACTB2, a metallo- β -lactamase protein, as a human mitochondrial endoribonuclease. *Nucleic
399 Acids Res.* 44, 1813–1832.
- 400 Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time
401 quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods San Diego Calif* 25, 402–
402 408.
- 403 Meynert, A.M., Bicknell, L.S., Hurles, M.E., Jackson, A.P., Taylor, M.S., 2013. Quantifying single
404 nucleotide variant detection sensitivity in exome sequencing. *BMC Bioinformatics* 14, 195.
- 405 Mouka, A., Izard, V., Tachdjian, G., Brisset, S., Yates, F., Mayeur, A., Drévillon, L., Jarray, R.,
406 Leboulch, P., Maouche-Chrétien, L., Tosca, L., 2017. Induced pluripotent stem cell generation
407 from a man carrying a complex chromosomal rearrangement as a genetic model for infertility
408 studies. *Sci. Rep.* 7, 39760.
- 409 Nilsson, D., Pettersson, M., Gustavsson, P., Förster, A., Hofmeister, W., Wincent, J., Zachariadis, V.,
410 Anderlid, B.-M., Nordgren, A., Mäkitie, O., Wirta, V., Käller, M., Vezzi, F., Lupski, J.R.,
411 Nordenskjöld, M., Lundberg, E.S., Carvalho, C.M.B., Lindstrand, A., 2017. Whole-Genome
412 Sequencing of Cytogenetically Balanced Chromosome Translocations Identifies Potentially
413 Pathological Gene Disruptions and Highlights the Importance of Microhomology in the
414 Mechanism of Formation. *Hum. Mutat.* 38, 180–192.
- 415 Ordulu, Z., Kammin, T., Brand, H., Pillalamarri, V., Redin, C.E., Collins, R.L., Blumenthal, I.,
416 Hanscom, C., Pereira, S., Bradley, I., Crandall, B.F., Gerrol, P., Hayden, M.A., Hussain, N.,
417 Kanengisser-Pines, B., Kantarci, S., Levy, B., Macera, M.J., Quintero-Rivera, F., Spiegel, E.,
418 Stevens, B., Ulm, J.E., Warburton, D., Wilkins-Haug, L.E., Yachelevich, N., Gusella, J.F.,
419 Talkowski, M.E., Morton, C.C., 2016. Structural Chromosomal Rearrangements Require

- 420 Nucleotide-Level Resolution: Lessons from Next-Generation Sequencing in Prenatal
421 Diagnosis. *Am. J. Hum. Genet.* 99, 1015–1033.
- 422 Oud, M.S., Volozonoka, L., Smits, R.M., Vissers, L.E.L.M., Ramos, L., Veltman, J.A., 2019. A
423 systematic review and standardized clinical validity assessment of male infertility genes. *Hum.*
424 *Reprod. Oxf. Engl.* 34, 932–941.
- 425 Poongothai, J., Gopenath, T.S., Manonayaki, S., 2009. Genetics of human male infertility. *Singapore*
426 *Med. J.* 50, 336–347.
- 427 Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn,
428 A.L., Machol, I., Omer, A.D., Lander, E.S., Aiden, E.L., 2014. A 3D map of the human
429 genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680.
- 430 Redin, C., Brand, H., Collins, R.L., Kammin, T., Mitchell, E., Hodge, J.C., Hanscom, C., Pillalamarri,
431 V., Seabra, C.M., Abbott, M.-A., Abdul-Rahman, O.A., Aberg, E., Adley, R., Alcaraz-
432 Estrada, S.L., Alkuraya, F.S., An, Y., Anderson, M.-A., Antolik, C., Anyane-Yeboa, K.,
433 Atkin, J.F., Bartell, T., Bernstein, J.A., Beyer, E., Blumenthal, I., Bongers, E.M.H.F., Brilstra,
434 E.H., Brown, C.W., Brüggewirth, H.T., Callewaert, B., Chiang, C., Corning, K., Cox, H.,
435 Cuppen, E., Currall, B.B., Cushing, T., David, D., Deardorff, M.A., Dheedene, A., D’Hooghe,
436 M., de Vries, B.B.A., Earl, D.L., Ferguson, H.L., Fisher, H., FitzPatrick, D.R., Gerrol, P.,
437 Giachino, D., Glessner, J.T., Gliem, T., Grady, M., Graham, B.H., Griffis, C., Gripp, K.W.,
438 Gropman, A.L., Hanson-Kahn, A., Harris, D.J., Hayden, M.A., Hill, R., Hochstenbach, R.,
439 Hoffman, J.D., Hopkin, R.J., Hubshman, M.W., Innes, A.M., Irons, M., Irving, M., Jacobsen,
440 J.C., Janssens, S., Jewett, T., Johnson, J.P., Jongmans, M.C., Kahler, S.G., Koolen, D.A.,
441 Korzelius, J., Kroisel, P.M., Lacassie, Y., Lawless, W., Lemyre, E., Leppig, K., Levin, A.V.,
442 Li, Haibo, Li, Hong, Liao, E.C., Lim, C., Lose, E.J., Lucente, D., Macera, M.J., Manavalan,
443 P., Mandrile, G., Marcelis, C.L., Margolin, L., Mason, T., Masser-Frye, D., McClellan, M.W.,
444 Mendoza, C.J.Z., Menten, B., Middelkamp, S., Mikami, L.R., Moe, E., Mohammed, S.,
445 Mononen, T., Mortenson, M.E., Moya, G., Nieuwint, A.W., Ordulu, Z., Parkash, S., Pauker,
446 S.P., Pereira, S., Perrin, D., Phelan, K., Aguilar, R.E.P., Poddighe, P.J., Pregno, G., Raskin, S.,
447 Reis, L., Rhead, W., Rita, D., Renkens, I., Roelens, F., Ruliera, J., Rump, P., Schilit, S.L.P.,

448 Shaheen, R., Sparkes, R., Spiegel, E., Stevens, B., Stone, M.R., Tagoe, J., Thakuria, J.V., van
449 Bon, B.W., van de Kamp, J., van Der Burgt, I., van Essen, T., van Ravenswaaij-Arts, C.M.,
450 van Roosmalen, M.J., Vergult, S., Volker-Touw, C.M.L., Warburton, D.P., Waterman, M.J.,
451 Wiley, S., Wilson, A., Yereña-de Vega, M. de la C.A., Zori, R.T., Levy, B., Brunner, H.G., de
452 Leeuw, N., Kloosterman, W.P., Thorland, E.C., Morton, C.C., Gusella, J.F., Talkowski, M.E.,
453 2017. The genomic landscape of balanced cytogenetic abnormalities associated with human
454 congenital anomalies. *Nat. Genet.* 49, 36–45.

455 Romana SP, Le Coniat M, Berger R. t(12;21): a new recurrent translocation in acute lymphoblastic
456 leukemia. *Genes, chromosomes & cancer.* 1994; 9:186-91.

457 Schilit, S.L.P., Menon, S., Friedrich, C., Kammin, T., Wilch, E., Hanscom, C., Jiang, S., Kliesch, S.,
458 Talkowski, M.E., Tüttelmann, F., MacQueen, A.J., Morton, C.C., 2020. SYCP2
459 Translocation-Mediated Dysregulation and Frameshift Variants Cause Human Male Infertility.
460 *Am. J. Hum. Genet.* 106, 41–57.

461 Schluth-Bolard, C., Diguët, F., Chatron, N., Rollat-Farnier, P.-A., Bardel, C., Afeñjar, A., Amblard,
462 F., Amiel, J., Blesson, S., Callier, P., Capri, Y., Collignon, P., Cordier, M.-P., Coubes, C.,
463 Demeer, B., Chaussonot, A., Demurger, F., Devillard, F., Doco-Fenzy, M., Dupont, C.,
464 Dupont, J.-M., Dupuis-Girod, S., Faivre, L., Gilbert-Dussardier, B., Guerrot, A.-M., Houlier,
465 M., Isidor, B., Jaillard, S., Joly-Hélas, G., Kremer, V., Lacombe, D., Le Caignec, C., Lebbar,
466 A., Lebrun, M., Lesca, G., Lespinasse, J., Levy, J., Malan, V., Mathieu-Dramard, M., Masson,
467 J., Masurel-Paulet, A., Mignot, C., Missirian, C., Morice-Picard, F., Moutton, S., Nadeau, G.,
468 Pebrel-Richard, C., Odent, S., Paquis-Flucklinger, V., Pasquier, L., Philip, N., Plutino, M.,
469 Pons, L., Portnoï, M.-F., Prieur, F., Puechberty, J., Putoux, A., Rio, M., Rooryck-Thambo, C.,
470 Rossi, M., Sarret, C., Satre, V., Siffroi, J.-P., Till, M., Touraine, R., Toutain, A., Toutain, J.,
471 Valence, S., Verloes, A., Whalen, S., Edery, P., Tabet, A.-C., Sanlaville, D., 2019. Whole
472 genome paired-end sequencing elucidates functional and phenotypic consequences of
473 balanced chromosomal rearrangement in patients with developmental disorders. *J. Med.*
474 *Genet.* 56, 526–535.

- 475 Schluth-Bolard, C., Labalme, A., Cordier, M.-P., Till, M., Nadeau, G., Tevissen, H., Lesca, G.,
476 Boutry-Kryza, N., Rossignol, S., Rocas, D., Dubruc, E., Edery, P., Sanlaville, D., 2013.
477 Breakpoint mapping by next generation sequencing reveals causative gene disruption in
478 patients carrying apparently balanced chromosome rearrangements with intellectual deficiency
479 and/or congenital malformations. *J. Med. Genet.* 50, 144–150.
- 480 Sironen, A., Shoemark, A., Patel, M., Loebinger, M.R., Mitchison, H.M., 2019. Sperm defects in
481 primary ciliary dyskinesia and related causes of male infertility. *Cell. Mol. Life Sci. CMLS.*
- 482 Sollars, V.E., McEntee, B.J., Engiles, J.B., Rothstein, J.L., Buchberg, A.M., 2002. A novel transgenic
483 line of mice exhibiting autosomal recessive male-specific lethality and non-alcoholic fatty
484 liver disease. *Hum. Mol. Genet.* 11, 2777–2786.
- 485 Talkowski, M.E., Ernst, C., Heilbut, A., Chiang, C., Hanscom, C., Lindgren, A., Kirby, A., Liu, S.,
486 Muddukrishna, B., Ohsumi, T.K., Shen, Y., Borowsky, M., Daly, M.J., Morton, C.C., Gusella,
487 J.F., 2011. Next-generation sequencing strategies enable routine detection of balanced
488 chromosome rearrangements for clinical diagnostics and genetic research. *Am. J. Hum. Genet.*
489 88, 469–481.
- 490 Vincent, M.-C., Daudin, M., De, M.P., Massat, G., Mieusset, R., Pontonnier, F., Calvas, P., Bujan, L.,
491 Bourrouillout, G., 2002. Cytogenetic investigations of infertile men with low sperm counts: a
492 25-year experience. *J. Androl.* 23, 18–22; discussion 44-45.
- 493 Vozdova, M., Oracova, E., Kasikova, K., Prinosilova, P., Rybar, R., Horinova, V., Gaillyova, R.,
494 Rubes, J., 2013. Balanced chromosomal translocations in men: relationships among semen
495 parameters, chromatin integrity, sperm meiotic segregation and aneuploidy. *J. Assist. Reprod.*
496 *Genet.* 30, 391–405.
- 497 Whitfield, M., Thomas, L., Bequignon, E., Schmitt, A., Stouvenel, L., Montantin, G., Tissier, S.,
498 Duquesnoy, P., Copin, B., Chantot, S., Dastot, F., Faucon, C., Barbotin, A.L., Loyens, A.,
499 Siffroi, J.-P., Papon, J.-F., Escudier, E., Amselem, S., Mitchell, V., Touré, A., Legendre, M.,
500 2019. Mutations in DNAH17, Encoding a Sperm-Specific Axonemal Outer Dynein Arm
501 Heavy Chain, Cause Isolated Male Infertility Due to Asthenozoospermia. *Am. J. Hum. Genet.*
502 105, 198–212.

503 Zhang, F., Khajavi, M., Connolly, A.M., Towne, C.F., Batish, S.D., Lupski, J.R., 2009. The DNA
504 replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex
505 rearrangements in humans. Nat. Genet. 41, 849–853.

506 Zhu, M., Need, A.C., Han, Y., Ge, D., Maia, J.M., Zhu, Q., Heinzen, E.L., Cirulli, E.T., Pelak, K., He,
507 M., Ruzzo, E.K., Gumbs, C., Singh, A., Feng, S., Shianna, K.V., Goldstein, D.B., 2012. Using
508 ERDS to infer copy-number variants in high-coverage genomes. Am. J. Hum. Genet. 91, 408–
509 421.

510

511

512

513 **Web resources**

514 ClinVar – NCBI : <https://www.ncbi.nlm.nih.gov/clinvar/variation/454741/>

515 Ensembl Compara database : <https://www.ensembl.org/>

516 Human BLAT Search: <http://www.genome.ucsc.edu/cgi-bin/hgBlat>

517 GATK: <https://gatk.broadinstitute.org/hc/en-us>

518 gnomAD. <https://gnomad.broadinstitute.org/>

519 GTEx Portal : <https://gtexportal.org/home/>

520 Integrative Genomics Viewer: <http://software.broadinstitute.org/software/igv/>

521 MutationTaster: <http://www.mutationtaster.org/>

522 PolyPhen-2: prediction of functional effects of human nsSNPs: <http://genetics.bwh.harvard.edu/pph2/>

523 SIFT - Predict effects of nonsynonymous / missense variants: <https://sift.bii.a-star.edu.sg/>

524 UCSC Genome Browser: <https://genome.ucsc.edu/>

525 3D Genome Browser : <http://promoter.bx.psu.edu/hi-c/>

526

527

528

529

530

531 **Figure legends**

532

533 **Figure 1: Patient 1 results a-** Ideogram of the translocation $t(8;22)(q13.3;p11)$. **b-** Sequences of
534 junction fragments and breakpoints coordinates as defined by paired-end genome and Sanger
535 sequencing. In red: reference sequence on chromosome 8; in blue reference sequence on chromosome
536 22, in green, sequence of unknown origin. Underlined bases correspond to micro-homology. Arrows
537 indicate the orientation of the reference strand. On derivative 8, alternate positions for
538 chr22_KI270733v1_random are chr21:8241581_8241833; chr21:8424605_8424857;
539 chr21:8469165_8469417. **c-** Three-dimension (3D)-genome map at the 8q13.3 locus derived from Hi-
540 C data of GM12878 (5 kb resolution) according to 3D-genome browser. TAD: topologically associated
541 domain; genes in black: forward strand; genes in blue: reverse strand. Breakpoint yellow line. **d-**
542 Expression of *LACTB2* in blood tissue for patient 1 compared to fertile and infertile controls.

543

544 **Figure 2: Patient 2 results: a-** Ideogram of the translocation $46,XY,t(1;18)(p21;q12)$. **b-** Sequences of
545 junction fragments and breakpoints coordinates as defined by paired-end genome and Sanger
546 sequencing. In red: reference sequence on chromosome 18; in blue reference sequence on chromosome
547 1; in green insertion of unknown origin. Arrows indicate the orientation of the reference strand. **c-**
548 Expression of *CTDPI* in blood tissue for patient 2 compared to fertile and infertile controls.

549

550 **Figure 3: Patient 3 results a-** Ideogram of the translocation $46,XY,t(9;21)(q12;q22.3)$. **b-** Sequences
551 of junction fragments and breakpoints coordinates as defined by paired-end genome and Sanger
552 sequencing. In red: reference sequence on chromosome 9; in blue reference sequence on chromosome
553 21; in green transversion from C to G. Underlined bases correspond to micro-homology. Arrows indicate
554 the orientation of the reference strand. Exact position on chromosome 9 could not be determined since
555 the sequence correspond to repeated element in centromere and constitutive heterochromatin. **c.**
556 Characteristics of SNVs identified in *DNAH5* gene.

557

558 **Supplementary Data**

559 **Supplementary Figure S1: a-c:** FISH for patient 1: **a** Hybridization of RP11-152C15 bacterial artificial
560 chromosome (BAC) clone (8q13.3) (red) and 8pter (green) (Cytocell, Cambridge, UK). RP11-152C15
561 probe is located on derivative (8). **b** Hybridization of RP11-11K9 BAC clone (8q13.3) (red) and 8pter
562 (green) (Cytocell, Cambridge, UK). The RP11-11K9 is located on derivative (22). **c** Hybridization of
563 Nucleolus Organizer Regions (red) and 8pter (green) (Cytocell, Cambridge, UK) showing presence of
564 NOR regions on derivative (8) . **d** IGV visualization of the chromosome 8 breakpoint in Patient 1 (red
565 dots). Beige arrows refer to mate on chromosome 21; blue arrows refer to mate on
566 chr22_KI270733v1_random and dark grey arrows refer to mate on chrUn_GL000220v1. **e**
567 Amplification of junction fragments in Patient 1 by PCR (Labchip visualization) for derivative (8)
568 der(8), derivative 22 der(22) and a control sequence (*ATPIA3* exons 7-8). B: blank; C: control without
569 chromosomal rearrangement, P: patient. **f** Expression of *NCOA2*, *TRAMI* and *LACTB2* in blood tissue
570 for patient 1 compared to fertile controls.

571
572 **Supplementary Figure S2: a-b** IGV visualization of the breakpoints (red dots) identified in Patient 2
573 on chromosome 1 (a) and 18 (b). **a** Green arrows refer to misaligned pairs with mate on chromosome
574 18. **b** blue arrows refer to misaligned pairs with mate on chromosome 1. *CTDPI* gene is disrupted by
575 the breakpoint. **c** Amplification of junction fragments in Patient 2 by PCR (Labchip visualization) for
576 derivative (1) der(1), derivative 18 der(18) and a control sequence (*ATPIA3* exons 7-8). B: blank; C:
577 control without chromosomal rearrangement, P2: patient.

578
579 **Supplementary Figure S3: a** Hybridization of RP1-265B9 (Red), RP11-113F1 (Red) et RP11-317L14
580 (Green) in Patient 3. Breakpoint on chromosome 21 is localized between RP1-265B9 and RP11-317L14.
581 **b** IGV visualization of the breakpoints (red dots) identified in Patient 3 on chromosome 21. a Green
582 arrows refer to misaligned pairs with mate on chromosome 9. **c** Amplification of junction fragments in
583 Patient 3 by PCR (Labchip visualization) for derivative (21) der(21), derivative 9 der(9) and a control
584 sequence (*ATPIA3* exons 7-8). B: blank; C: control without chromosomal rearrangement, P3: patient.

585

586 **Supplementary Figure S4:** 3D-genome map derived from Hi-C data of GM12878 (5 kb resolution)
587 according to 3D-genome browser. TAD: topologically associated domain; genes in black: forward
588 strand; genes in blue: reverse strand. Breakpoint in yellow. **a-** Patient 2, locus 1p22.2. **b-** Patient 2, locus
589 18q23. **d-** Patient 3, locus 21q22.3.

590

591 **Supplementary Table I:** List of primers used in this study

592

593 **Supplementary Table II:** List of genes involved in male infertility (Oud et al., 2019)

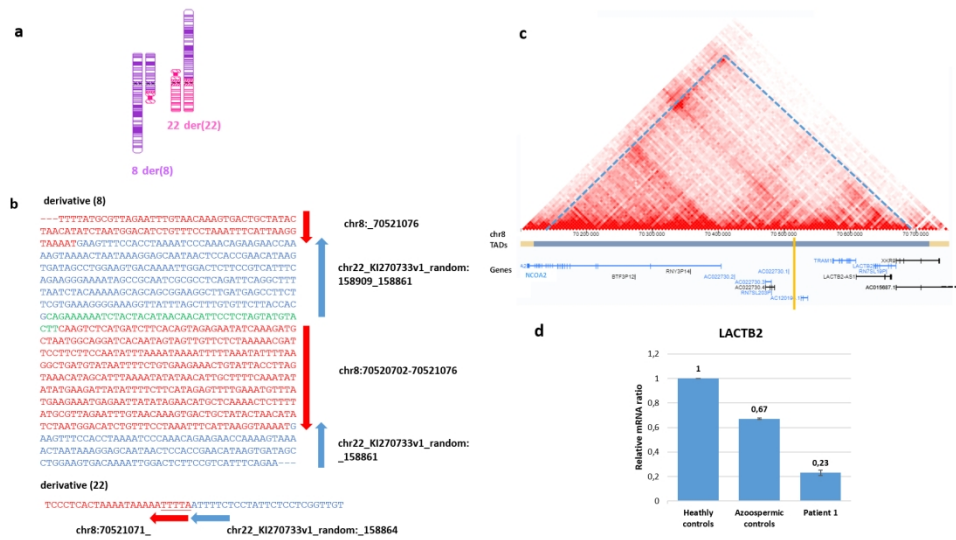


Figure 1: Patient 1 results a- Ideogram of the translocation t(8;22)(q13.3;p11). b- Sequences of junction fragments and breakpoints coordinates as defined by paired-end genome and Sanger sequencing. In red: reference sequence on chromosome 8; in blue reference sequence on chromosome 22, in green, sequence of unknown origin. Underlined bases correspond to micro-homology. Arrows indicate the orientation of the reference strand. On derivative 8, alternate positions for chr22_KI270733v1_random are chr21:8241581_8241833; chr21:8424605_8424857; chr21:8469165_8469417. c- Three-dimension (3D)-genome map at the 8q13.3 locus derived from Hi-C data of GM12878 (5 kb resolution) according to 3D-genome browser. TAD: topologically associated domain; genes in black: forward strand; genes in blue: reverse strand. Breakpoint yellow line. d- Expression of LACTB2 in blood tissue for patient 1 compared to fertile and infertile controls.

346x196mm (150 x 150 DPI)

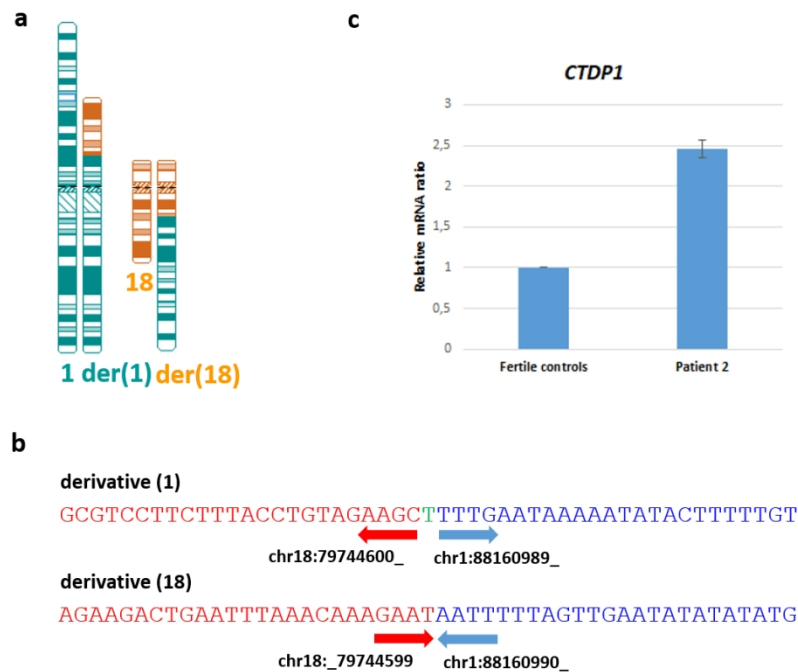


Figure 2: Patient 2 results: a- Ideogram of the translocation 46,XY,t(1;18)(p21;q12). b- Sequences of junction fragments and breakpoints coordinates as defined by paired-end genome and Sanger sequencing. In red: reference sequence on chromosome 18; in blue reference sequence on chromosome 1; in green insertion of unknown origin. Arrows indicate the orientation of the reference strand. c- Expression of CTDP1 in blood tissue for patient 2 compared to fertile and infertile controls.

225x161mm (150 x 150 DPI)

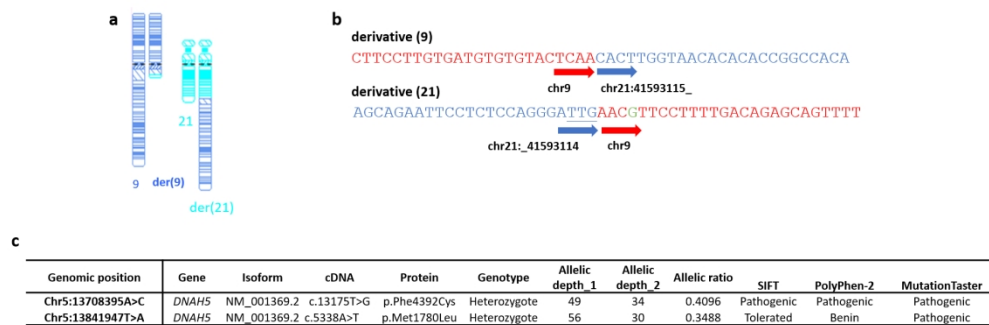


Figure 3: Patient 3 results a- Ideogram of the translocation 46,XY,t(9;21)(q12;q22.3). b- Sequences of junction fragments and breakpoints coordinates as defined by paired-end genome and Sanger sequencing. In red: reference sequence on chromosome 9; in blue reference sequence on chromosome 21; in green transversion from C to G. Underlined bases correspond to micro-homology. Arrows indicate the orientation of the reference strand. Exact position on chromosome 9 could not be determined since the sequence correspond to repeated element in centromere and constitutive heterochromatin. c. Characteristics of SNVs identified in DNAH5 gene.

334x119mm (150 x 150 DPI)

Supplementary material

Content:

Supplementary Figure S1: Breakpoint characterization and expression study in patient 1.

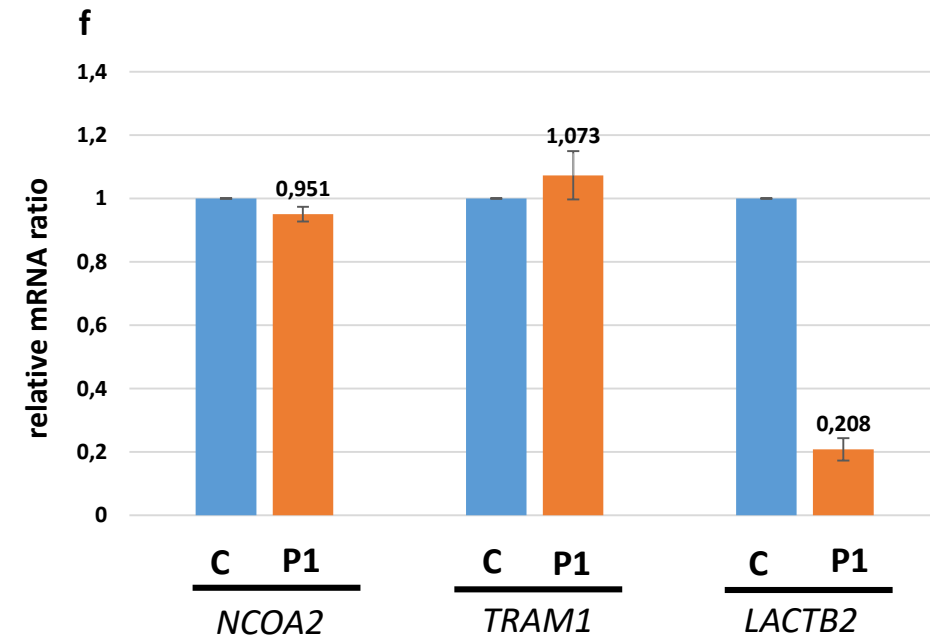
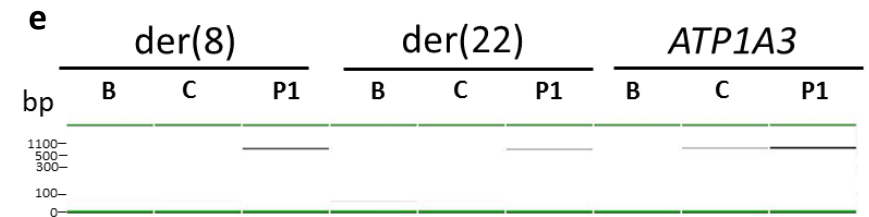
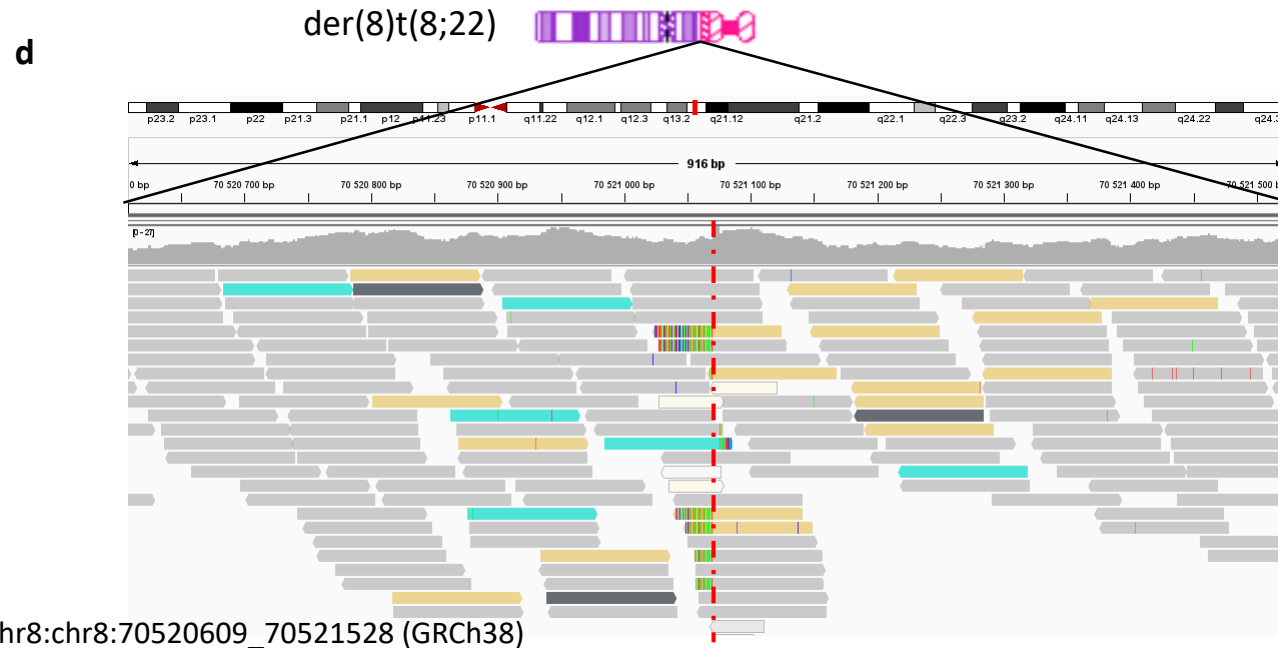
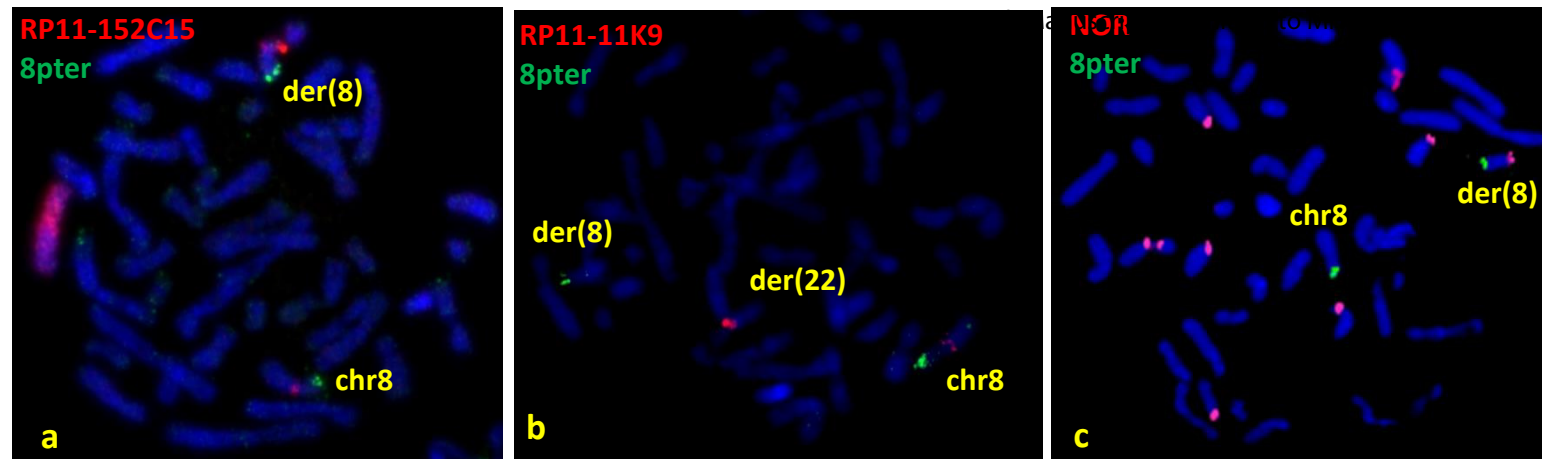
Supplementary Figure S2: Breakpoint characterization in patient 2.

Supplementary Figure S3: Breakpoint characterization in patient 3.

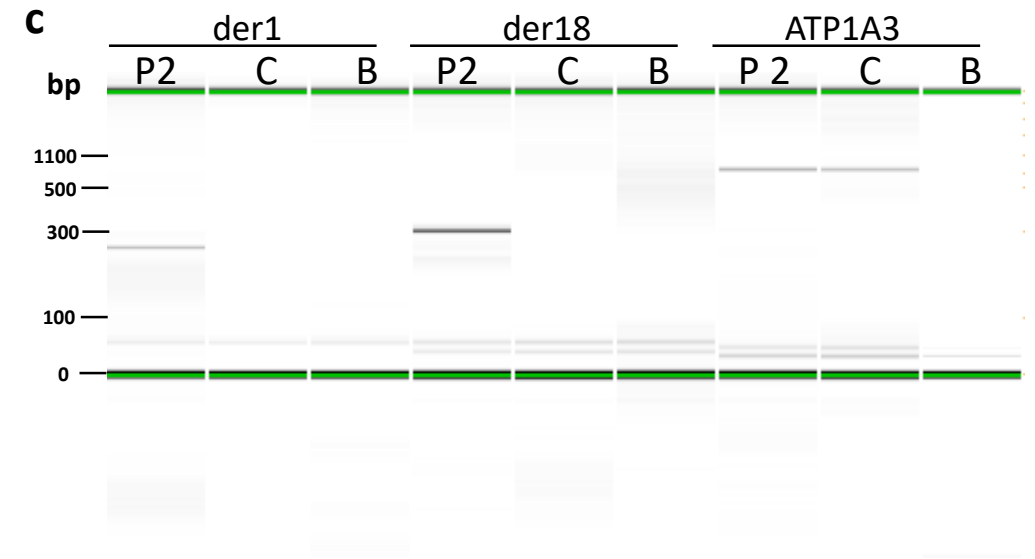
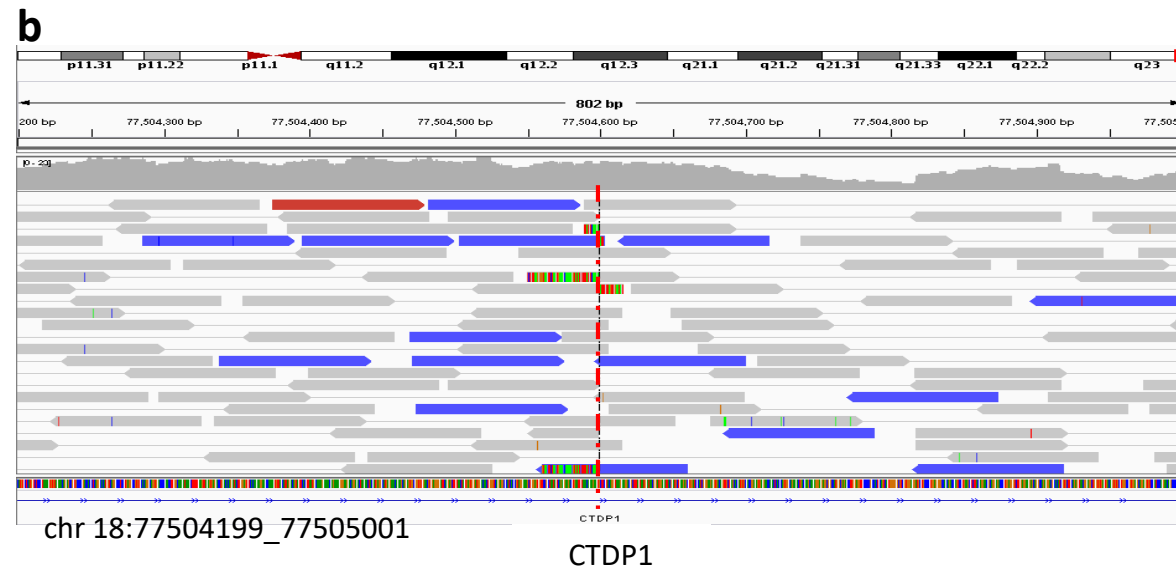
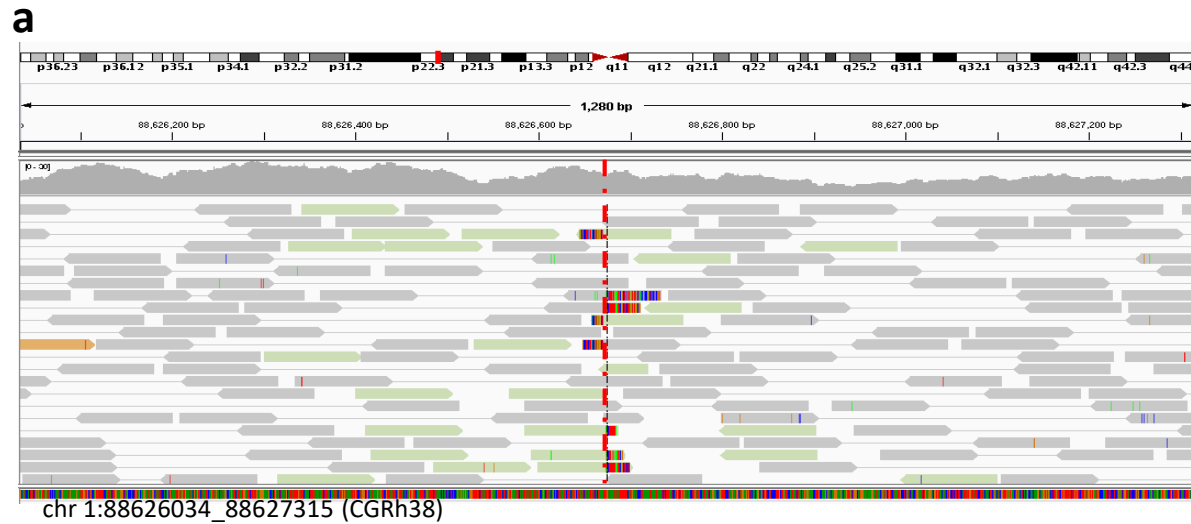
Supplementary Figure S4: 3D-genome map of breakpoints.

Supplementary Table I: List of primers used in this study

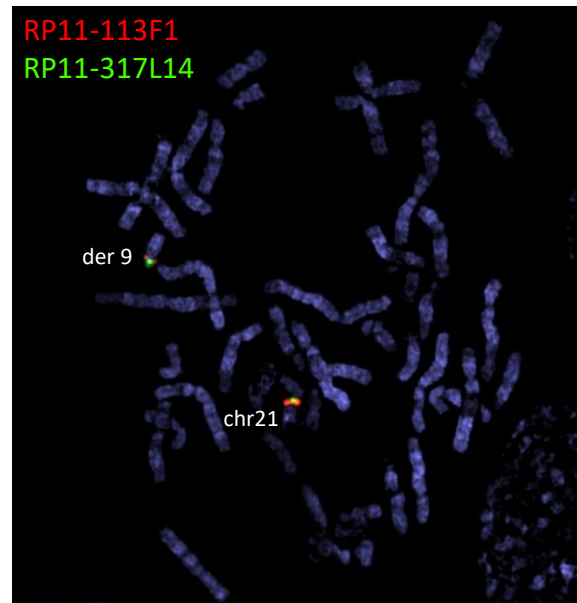
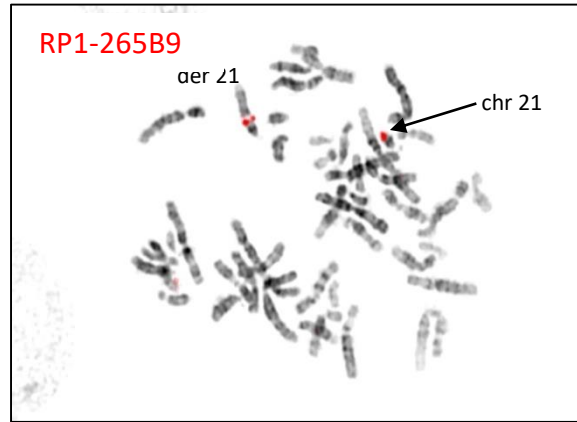
Supplementary Table II: List of genes involved in male infertility (Oud et al., 2019)



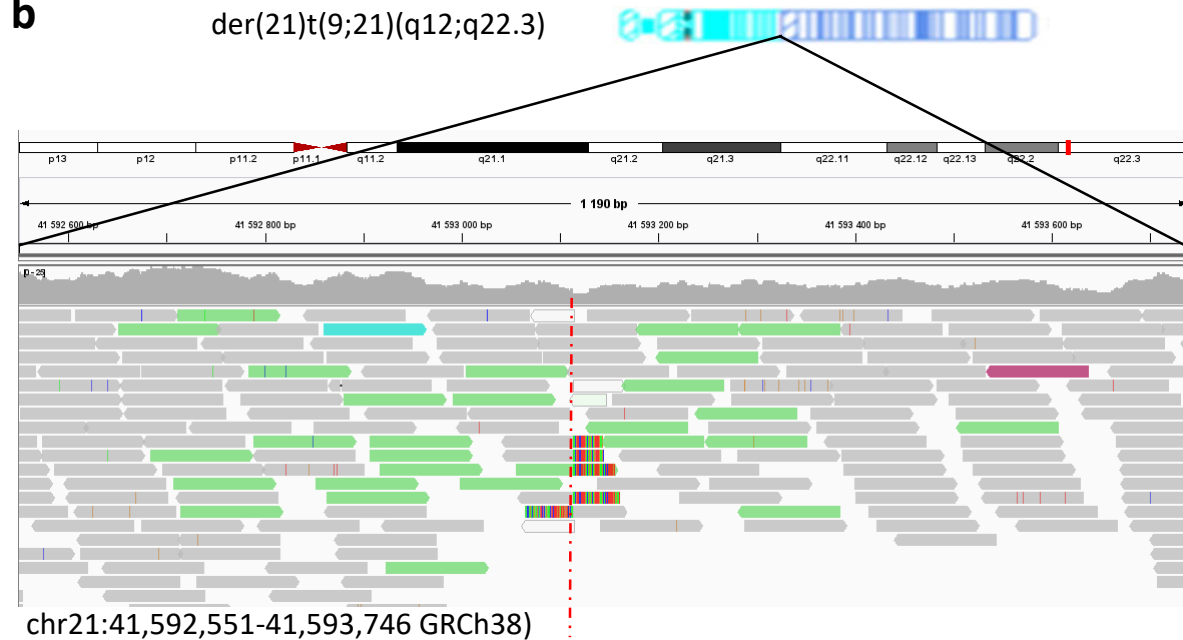
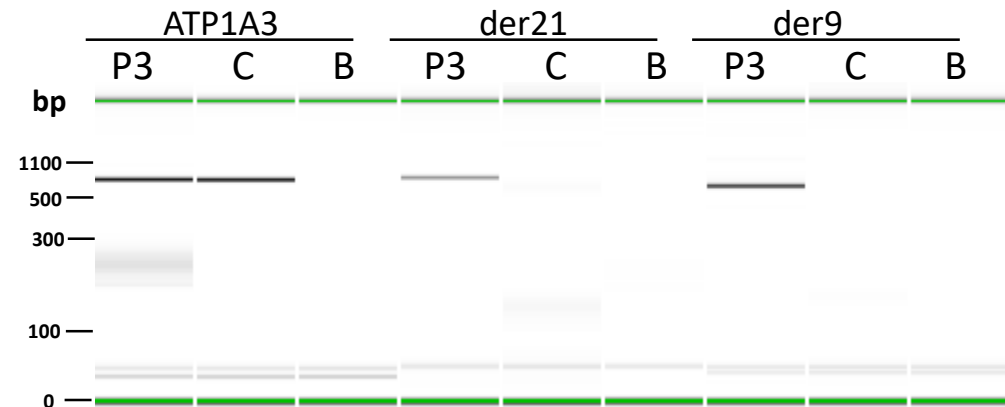
Supplementary Figure S1: Breakpoint characterization and expression study in patient 1. **a-c** FISH for patient 1: **a** Hybridization of RP11-152C15 bacterial artificial chromosome (BAC) clone (8q13.3) (red) and 8pter (green) (CytoCell, Cambridge, UK). RP11-152C15 probe is located on derivative (8). **b** Hybridization of RP11-11K9 BAC clone (8q13.3) (red) and 8pter (green) (CytoCell, Cambridge, UK). The RP11-11K9 showed is located on derivative (22). **c** Hybridization of Nucleolus Organizer Regions (red) and 8pter (green) (CytoCell, Cambridge, UK) showing presence of NOR regions on derivative (8). **d** IGV visualization of the chromosome 8 breakpoint in Patient 1 (red dots). Beige arrows refer to mate on chromosome 21; blue arrows refer to mate on chr22_K1270733v1_random and dark grey arrows refer to mate on chrUn_GL000220v1. **e** Amplification of junction fragments in Patient 1 by PCR (Labchip visualization) for derivative (8) der(8), derivative 22 der(22) and a control sequence (*ATP1A3* exons 7-8). B: blank; C: control without chromosomal rearrangement, P: patient. **f** Expression of *NCOA2*, *TRAM1* and *LACTB2* in blood tissue for patient 1 compared to fertile controls.



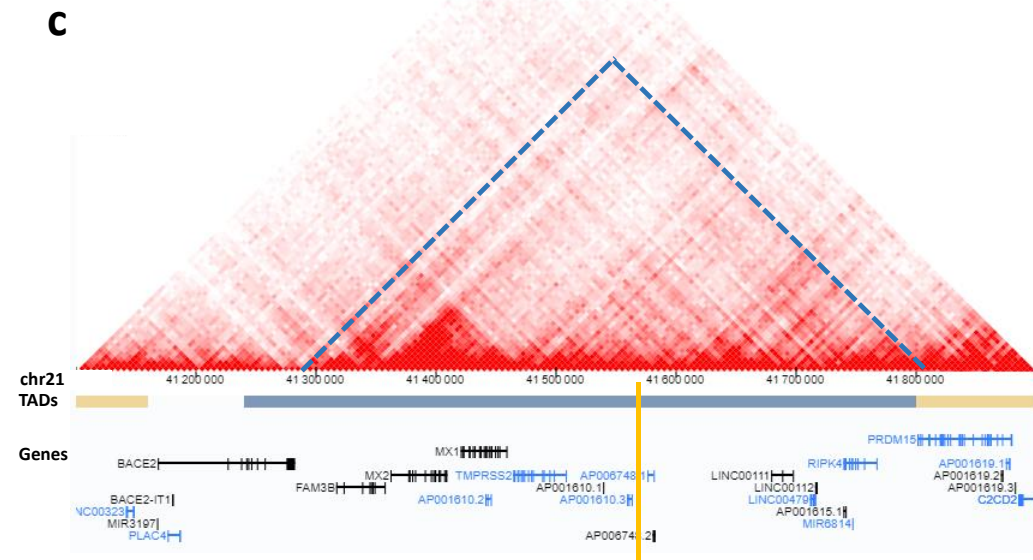
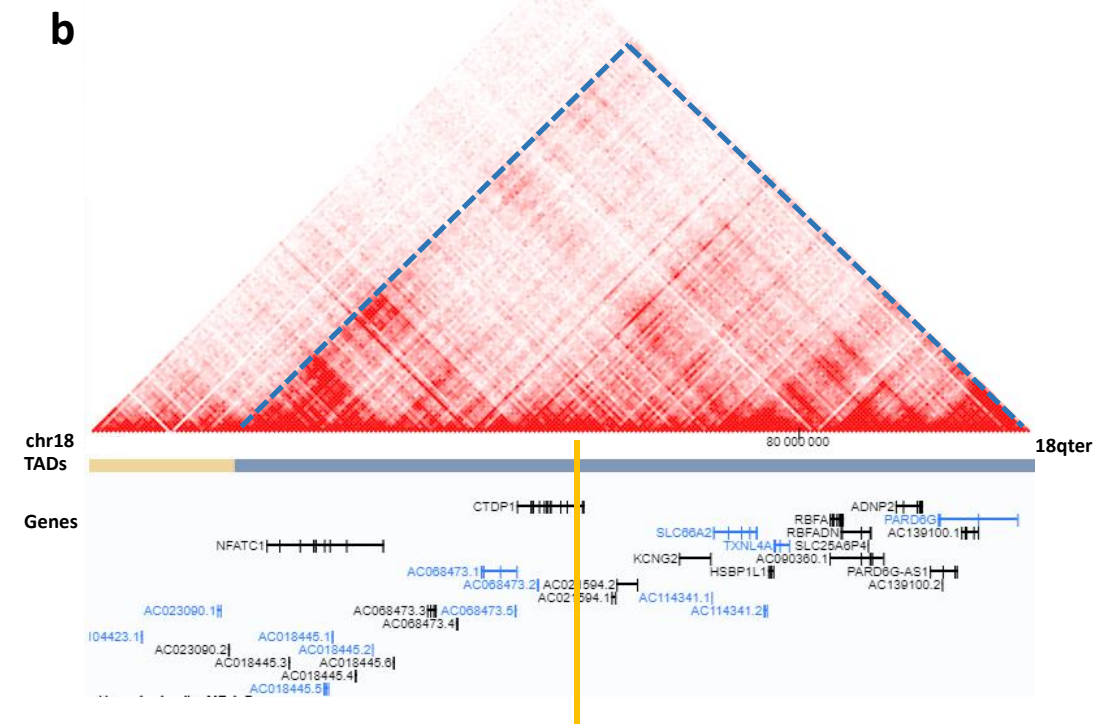
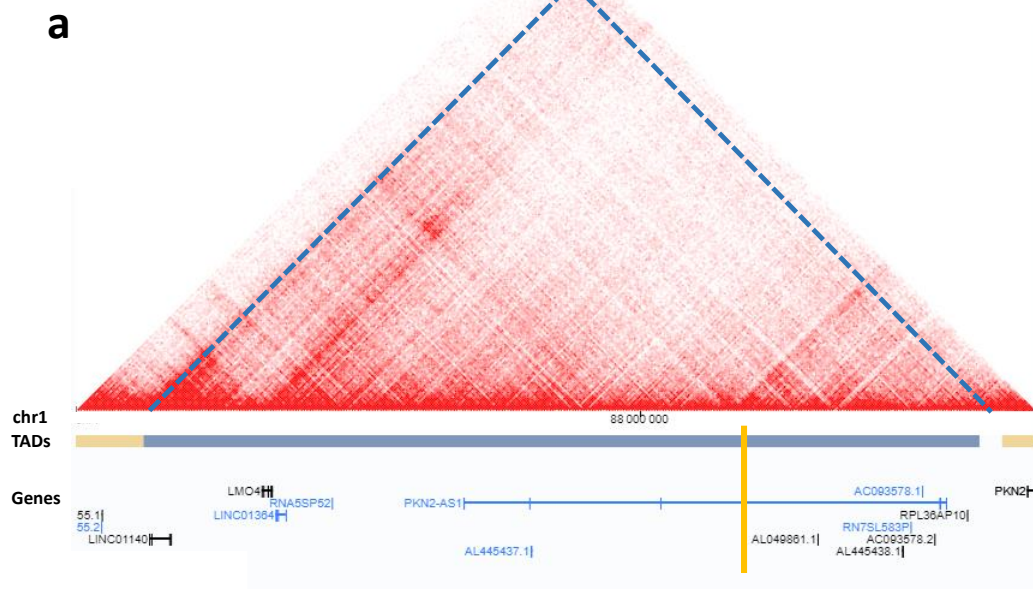
Supplementary Figure S2: Breakpoint characterization in patient 2. a-b IGV visualization of the breakpoints (red dots) identified in Patient 2 on chromosome 1 (a) and 18 (b). a Green arrows refer to misaligned pairs with mate on chromosome 18. b blue arrows refer to misaligned pairs with mate on chromosome 1. *CTDP1* gene is disrupted by the breakpoint. c Amplification of junction fragments in Patient 2 by PCR (Labchip visualization) for derivative (1) der(1), derivative 18 der(18) and a control sequence (*ATP1A3* exons 7-8). B: blank; C: control without chromosomal rearrangement, P2: patient 2.

a**b**

der(21)t(9;21)(q12;q22.3)

**c**

Supplementary Figure S3: Breakpoint characterization in patient 3. **a** Hybridization of RP1-265B9 (Red), RP11-113F1 (Red) et RP11-317L14 (Green) in Patient 3. Breakpoint on chromosome 21 is localized between RP1-265B9 and RP11-317L14. **b** IGV visualization of the breakpoints (red dots) identified in Patient 3 on chromosome 21. a Green arrows refer to misaligned pairs with mate on chromosome 9. **c** Amplification of junction fragments in Patient 3 by PCR (Labchip visualization) for derivative (21) der(21), derivative 9 der(9) and a control sequence (*ATP1A3* exons 7-8). B: blank; C: control without chromosomal rearrangement, P3: patient 3.



Supplementary Figure S4: 3D-genome map of breakpoints. The maps are derived from Hi-C data of GM12878 (5 kb resolution) according to 3D-genome browser. TAD: topologically associated domain; genes in black: forward strand; genes in blue: reverse strand. Breakpoint in yellow. a- Patient 2, locus 1p22.2. b- Patient 2, locus 18q23. d- Patient 3, locus 21q22.3.

Junction sequencing		
	Primer title	Sequence
Patient 1	ATP1A3 (exon 7-8)-F	5'-CTTCTAGCTGTGATCTCCAG-3'
	ATP1A3 (exon 7-8)-R	5'-ACACGGGTCCAGGCCTCTA-3'
	der-8-F	5'-GGATTCAAGACTTAGGGAGGTTTC-3'
	der-8-R	5'-TGATAGCTTCTTTGCCGTGGT-3'
Patient 2	der-22-F	5'-AGAATGCCTACCTCTGGACT-3'
	der-22-R	5'-GGAGACGGACCCTATTGACA-3'
	der-1-F	5'-GCAACTGTCACTCTTCCCGA-3'
	der-1-R	5'-TGTAGCACTTTCCTGAGCCA-3'
Patient 3	der-18-F	5'-AACGTGAGGGGATTCAGAGC-3'
	der-18-R	5'-AGAGGAGAAGCGGGTGTATTAT-3'
	der-21-F	5'-TCTCTGAATGCTGACCTGCC-3'
	der-21-R	5'-TCCAAATGTCCACTTCCAGGT-3'
	der-9-F	5'-TGCGATGTGTCTTCAACT-3'
	der-9-R	5'-TCGCAATGACGGGAACAGAA-3'
Expression study		
	Primer title	Sequence
Patient 1	ACTB-F	5'-ATTGGCAATGAGCGGTTTC-3'
	ACTB-R	5'-CGTGGATGCCACAGGACT-3'
	LACT2B-F	5'-TGATTAAGACTGAGGGAGCCAC-3'
	LACT2B-R	5'-ACCGTTGTTCCCTCCCTAG-3'
	NCOA2-F	5'-GCTGCCAACATAGATGAAGTGC-3'
	NCOA2-R	5'-CCTTCCAGGTTCACTACAAAGA-3'
	TRAM1-F	5'-TCCTCGTCAGCTTGTCTACA-3'
Patient 2	TRAM1-R	5'-ACAGGCGGGAAATGTGGAAA-3'
	CTDP1-F	5'-AACACGGGGAAAGCTCATCAG-3'
	CTDP1-R	5'-CTTACCAAACATCTGCGGC-3'

Supplementary table I: Primers sequences

gene name	Gene locus	Disease category	OMIM Number	Inheritance
ABCA1	9q31.1	Leydig cell dysfunction	600046	AR
ACE	17q23.3	Oligozoospermia and azoospermia	106180	Not reported
ACR	22q13.33	Unexplained male infertility	102480	AR
ACTL7A	9q31.3	Azoospermia and oligozoospermia	604303	Not reported
ACTL7B	9q31.3	Azoospermia and oligozoospermia	604304	Not reported
ADAM20	14q24.2	Fertilization failure	603712	AD
ADCY10	1q24.2	Asthenozoospermia	605205	AR
ADGRG2	Xp22.13	Abnormal development of vas deferens	300572	XL
AGFG1	2q36.3	Globozoospermia	600862	AR
AK7	14q32.2	Spermiogenesis defect	615364	AR
AKAP2	9q31.3	Kallmann syndrome; hypogonadotropic hypogonadism, delayed puberty, cryptorchidism, small testes	604582	AD
AKAP3	12p13.32	Multiple morphological abnormalities of the sperm flagella	604689	AR
AKAP4	Xp11.22	Spermiogenesis defect	300185	XL
AKR1C2	10p15.1	Hypospadias	600450	Not reported
AKR1C3	10p15.1	Hypospadias	603966	Not reported
AKR1C4	10p15.1	Hypospadias	600451	Not reported
AKR1D1	7q33	Hypospadias	604741	AR
ALKBH5	17p11.2	Male infertility	613303	Not reported
AMELY	Yp11.2	Oligozoospermia	410000	YL
AMH	19p13.3	Abnormal development of reproductive organs	600957	AR
AMHR2	12q13.13	Abnormal development of reproductive organs	600956	AR
AMN1	12p11.21	Hypogonadotropic hypogonadism and Kallmann syndrome	-	AR
ANOS1	Xp22.31	Abnormal hypothalamus development and function	300836	XL
APOA1	11q23.3	Pre-meiotic arrest	107680	AD
AR	Xq12	Abnormal development of reproductive organs	313700	XL
ART3	4q21.1	Non-obstructive azoospermia	603086	AD
ARX	Xp21.3	5 α -reductase deficiency	300382	XL
ATF3	1q32.3	Hypospadias	603148	Not reported
AURKC	19q13.43	Spermiogenesis defect	603495	AR
AXL	19q13.2	Kallmann syndrome; Anosmia in combination with Hypogonadotropic Hypogonadism. Additional features may include cryptorchidism, unilateral absence of vas deferens, and/or micropenis	109135	AD
B4GALT1	9p21.1	Fertilization failure	137060	Not reported
B4GAT1	11q13.2	Hypogonadotropic hypogonadism	605517	Not reported
BCL2L2	14q11.2	Non-obstructive azoospermia	601931	AR
BEX3	Xq22.2	Non-obstructive azoospermia	300361	Not reported
BMP4	14q22.2	Abnormal development of urethra	112262	AD
BMP7	20q13.31	Abnormal development of urethra	112267	AD
BMPR1B	4q22.3	Azoospermia	603248	AR
BNC2	9p22.3-p22.2	Abnormal development of urethra	608669	AD

BOLL	2q33.1	Azoospermia and oligozoospermia	606165	AR
BPY2	Yq11.223	Azoospermia and oligozoospermia	400013	YL
BRAF	7q34	Abnormal development of reproductive organs	164757	AD
BRDT	1p22.1	Spermiogenesis defect/Non-obstructive azoospermia	602144	AR
BRWD1	21q22.2	Non-obstructive azoospermia	617824	AR
BSCL2	11q12.3	Spermiogenesis defect	606158	AR
C11orf70	11q22.1	Spermiogenesis defect	618058	AR
CACNA2D1	7q21.11	Azoospermia	114204	AD
CALCA	11p15.2	Cryptorchidism	114130	Not reported
CALCB	11p15.2	Cryptorchidism	114160	Not reported
CALCRL	2q32.1	Cryptorchidism	114190	AR
CAMK4	5q22.1	Azoospermia	114080	AD
CAPZA3	12p12.3	Male infertility	608722	AR
CASR	3q13.33-q21.1	Hypogonadotropic hypogonadism	601199	Not reported
CATSPER1	11q13.1	Spermiogenesis defect	606389	AR
CATSPER2	15q15.3	Asthenozoospermia without deafness/Asthenozoospermia with deafness	607249	AR
CATSPER3	5q31.1	Asthenozoospermia	609120	AR
CATSPER4	1p36.11	Asthenozoospermia	609121	AR
CATSPERE	1q44	Fertilization defect	617510	AR
CBX2	17q25.3	46,XX Disorder of sex development	613080	AR
CCDC103	17q21.31	Spermiogenesis defect	614677	AR
CCDC141	2q31.2	Abnormal hypothalamus development and function	616031	AR
CCDC155	19q13.33	Meiotic arrest	618125	AR
CCDC39	3q26.33	Spermiogenesis defect/Asthenozoospermia	613798	AR
CCDC40	17q25.3	Spermiogenesis defect/Asthenozoospermia	613799	AR
CCKAR	4p15.2	Hypogonadotropic hypogonadism	118444	AR
CCKBR	11p15.4	Hypogonadotropic hypogonadism and Kallmann syndrome, undescended testes	118445	AD
CCNA1	13q13.3	Azoospermia and oligozoospermia	604036	Not reported
CD46	1q32.2	Idiopathic male infertility	120920	AR
CDC14A	1p21.2	Pre-meiotic arrest/Meiotic arrest/Spermiogenesis defect	603504	AR
CDC20	1p34.2	Non-obstructive azoospermia	603618	AD
CDK2	12q13.2	Non-obstructive azoospermia	116953	AD
CDY1	Yq11.23	Non-obstructive azoospermia	400016	YL
CEP135	4q12	Spermiogenesis defect	611423	AR
CEP290	12q21.32	Spermiogenesis defect	610142	AR
CFAP43	10q25.1	Spermiogenesis defect	617558	AR
CFAP44	3q13.2	Spermiogenesis defect	617559	AR
CFAP69	7q21.13	Spermiogenesis defect	617949	AR
CFTR	7q31.2	Low fertilization and pregnancy rate	602421	AR
CHD7	8q12.2	Abnormal hypothalamus development and function	608892	AD
CLCA4	1p22.3	Azoospermia or oligoasthenozoospermia	616857	AD
CNTN2	1q32.1	Hypogonadotropic hypogonadism	190197	AD
COL1A1	17q21.33	Sertoli cell only syndrome	120150	Not reported
CRCP	7q11.21	Cryptorchidism	606121	Not reported

CREM	10p11.21	Azoospermia	123812	AD
CRISP2	6p12.3	Asthenozoospermia and teratozoospermia	187430	AR
CS	12q13.3	Azoospermia and oligozoospermia	118950	AD
CSNK2A2	16q21	Globozoospermia	115442	Not reported
CSNK2B	6p21.33	Globozoospermia	115441	Not reported
CTCFL	20q13.31	OAT	607022	AR
CUL4B	Xq24	Sertoli cell only syndrome	300304	XL
CXCR4	2q22.1	Hypogonadotropic hypogonadism and Kallmann syndrome	162643	Not reported
CYP11A1	15q24.1	Adrenal gland dysfunction	118485	AR
CYP11B1	8q24.3	Adrenal gland dysfunction	610613	AR
CYP17A1	10q24.32	Adrenal gland dysfunction	609300	AR
CYP19A1	15q21.2	Adrenal gland dysfunction	107910	AD
CYP21A2	6p21.33	Adrenal gland dysfunction	613815	AR
DAZ1	Yq11.223	Azoospermia	400003	YL
DAZL	3p24.3	Azoospermia and oligozoospermia	601486	AR
DCAF17	2q31.1	Hypogonadotropic hypogonadism	612515	AR
DCC	18q21.2	Abnormal hypothalamus development and function	120470	AD
DDX25	11q24.2	Azoospermia and oligozoospermia	607663	AR
DDX3Y	Yq11.221	Azoospermia	400010	YL
DDX53	Xp22.11	Non-obstructive azoospermia	-	XL
DEFB126	20p13	Male infertility	-	AR
DHH	12q13.12	Abnormal development of reproductive organs	233420	AR
DKK1	10q21.1	Hypospadias	605189	Not reported
DLX5	7q21.3	Hypogonadotropic hypogonadism	600028	Not reported
DMC1	22q13.1	Meiotic arrest	602721	AR
DMRT1	9p24.3	Sertoli cell-only syndrome/Pre-meiotic arrest/Meiotic arrest	602424	AD
DNAAF2	14q21.3	Spermiogenesis defect	612517	AR
DNAAF4	15q21.3	Spermiogenesis defect	608706	AR
DNAAF5	7p22.3	Spermiogenesis defect	614864	AR
DNAH1	3p21.1	Spermiogenesis defect	603332	AR
DNAH11	7p15.3	Asthenozoospermia	603339	AR
DNAH5	5p15.2	Asthenozoospermia	603335	AR
DNAH6	2p11.2	Azoospermia	603336	AR
DNAH9	17p12	Spermiogenesis defect	603330	AR
DNAI1	9p13.3	Spermiogenesis defect	604366	AR
DNAI2	17q25.1	Spermiogenesis defect	605483	AR
DNAJB13	11q13.4	Spermiogenesis defect	610263	AR
DNMT1	19p13.2	Sertoli cell-only syndrome/pre-meiotic arrest/meiotic arrest	126375	AD
DNMT3A	2p23.3	Oligozoospermia	602769	AR
DNMT3B	20q11.21	Sertoli cell-only syndrome/pre-meiotic arrest/meiotic arrest	602900	AD
DNMT3L	21q22.3	Azoospermia and oligozoospermia	606588	AR
DPP6	7q36.2	Azoospermia	126141	AD
DPY19L2	12q14.2	Spermiogenesis defect	613893	AR
DUSP6	12q21.33	Kallmann syndrome; Hypogonadotropic hypogonadism	602748	AD

DYNLT3	Xp11.4	Azoospermia	300302	XL
E2F1	20q11.22	Leydig cell dysfunction/Sertoli cell-only syndrome/Pre-meiotic arrest/Meiotic arrest/Spermiogenesis defect	189971	AD
EBF2	8p21.2	Kallmann syndrome; hypogonadotropic hypogonadism	609934	Not reported
EDNRB	13q22.3	Hypogonadotropic hypogonadism	131244	Not reported
EFNA5	5q21.3	Hypogonadotropic hypogonadism	601535	AD
EGF	4q25	Hypogonadotropic hypogonadism	126150	Not reported
EGFR	7p11.2	Hypogonadotropic hypogonadism	131550	Not reported
EGR4	2p13.2	Oligozoospermia and azoospermia	128992	AR
EIF5A2	3q26.2	Oligozoospermia and azoospermia	605782	Not reported
EMX2	10q26.11	Kallmann syndrome; hypogonadotropic hypogonadism	607637	AD
EPHA5	4q13.1-q13.2	Hypogonadotropic hypogonadism	600004	Not reported
EPPIN	20q13.12	Male infertility	609031	Not reported
ERBB4	2q34	Abnormal hypothalamus development and function	600543	AD
ESR1	6q25.1-q25.2	Oligozoospermia and azoospermia	133430	Not reported
ESR2	14q23.2-q23.3	Oligozoospermia and azoospermia	601663	Not reported
ETV5	3q27.2	Non-obstructive azoospermia	608067	Not reported
EXT1	8q24.11	Oligozoospermia	608177	AD
FABP9	8q21.13	Teratozoospermia	-	AR
FANCA	16q24.3	Sertoli cell-only syndrome	607139	AR
FANCM	14q21.2	Sertoli cell-only syndrome/pre-meiotic arrest	609644	AR
FATE1	Xq28	Sertoli cell-only syndrome/Pre-meiotic arrest/Meiotic arrest/Spermiogenesis	300450	XL
FEZ1	11q24.2	Hypogonadotropic hypogonadism	604825	Not reported
FEZF1	7q31.32	Abnormal hypothalamus development and function	613301	AR
FGF10	5p12	Hypospadias	602115	Not reported
FGF17	8p21.3	Abnormal hypothalamus development and function	603725	AD
FGF18	5q35.1	Kallmann syndrome	603726	Not reported
FGF8	10q24.32	Abnormal hypothalamus development and function	600483	AD
FGF9	13q12.11	46,XX Disorders of sex development	600921	AD
FGFR1	8p11.23	Abnormal hypothalamus development and function	136350	AD
FGFR2	10q26.13	Isolated hypospadias	176943	AD
FHL5	6q16.1	Azoospermia and oligozoospermia	605126	AR
FKBP4	12p13.33	Hypospadias	600611	Not reported
FKBP6	7q11.23	Azoospermia	604839	AR
FKBPL	6p21.32	Azoospermia	617076	AD
FLNA	Xq28	Abnormal development of reproductive organs	300017	XL
FLRT3	20p12.1	Kallmann syndrome	604808	Not reported
FOXL2	3q22.3	Oligozoospermia in combination with Blepharophimosis, corneal vascularization, deafness, and acroosteolysis	605597	AD
FSHB	11p14.1	Pituitary gland dysfunction	136530	AR
FSHR	2p16.3	Pre-meiotic arrest	136435	AR
FSIP2	2q32.1	Spermiogenesis defect	615796	AR
FTHL17	Xp21.2	Azoospermia	300308	XL
FTO	16q12.2	Oligozoospermia	610966	AR
FUS	16p11.2	Non-obstructive azoospermia	137070	AD
GALNTL5	7q36.1	Spermiogenesis defect	615133	AD

GAP43	3q13.31	Kallmann syndrome	162060	Not reported
GAPDHS	19q13.12	Multiple morphological abnormalities of the sperm flagella	609169	AR
GAS8	16q24.3	Spermiogenesis defect	605178	AR
GATA4	8p23.1	Abnormal development of reproductive organs	600576	AD
GGN	19q13.2	Oligozoospermia	609966	AR
GH1	17q23.3	Pituitary gland dysfunction	139250	AD
GHR	5p13.1-p12	Hypogonadotropic hypogonadism	600946	Not reported
GJA1	6q22.31	Oligozoospermia in combination with Blepharophimosis, corneal vascularization, deafness, and acroosteolysis	121014	Not reported
GJB2	13q12.11	Oligozoospermia in combination with Blepharophimosis, corneal vascularization, deafness, and acroosteolysis	220290	AD
GJB6	13q12.11	Oligozoospermia in combination with Blepharophimosis, corneal vascularization, deafness, and acroosteolysis	604418	Not reported
GLI3	7p14.1	Kallmann syndrome	165240	AD
GNRH1	8p21.2	Abnormal hypothalamus development and function	152760	AR
GNRHR	4q13.2	Pituitary gland dysfunction	138850	AR
GOPC	6q22.1	Globozoospermia	606845	AR
GPX4	19p13.3	Oligoasthenozoospermia	138322	AD
GTF2H3	12q24.31	Pre-meiotic arrest/Meiotic arrest/Spermiogenesis defect	601750	AR
H19	11p15.5	Sertoli cell only syndrome	103280	Not reported
H1FNT	12q13.11	Azoospermia and oligozoospermia	618565	AR
H2AFX	11q23.3	Spermatogenic failure	601772	Not reported
HASPIN	17p13.2	Male infertility	609240	Not reported
HAUS7	Xq28	Meiotic arrest	300540	XL
HESX1	3p14.3	Abnormal hypothalamus development and function	601802	AD
HFE	6p22.2	Hypogonadism due to iron excess	613609	AR
HGF	7q21.11	Hypogonadotropic hypogonadism	142409	Not reported
HMGB2	4q34.1	Testicular failure	163906	Not reported
HOOK1	1p32.1	Acephalic sperm	607820	AR
HORMAD1	1q21.3	Meiotic arrest	609824	AR
HOXA10	7p15.2	Cryptorchidism	142957	AR/AD
HOXA11	7p15.2	Cryptorchidism	142958	Not reported
HOXA13	7p15.2	Cryptorchidism	142959	AD
HOXA4	7p15.2	Cryptorchidism	142953	AD
HOXA9	7p15.2	Cryptorchidism	14295	Not reported
HOXB13	17q21.32	Cryptorchidism	604607	AR
HOXB6	17q21.32	Hypospadias	142961	AD
HOXB9	17q21.32	Cryptorchidism	142964	Not reported
HOXC10	12q13.13	Cryptorchidism	605560	AD
HOXC11	12q13.13	Cryptorchidism	605559	AD
HOXC12	12q13.13	Cryptorchidism	142975	AD
HOXC13	12q13.13	Cryptorchidism	142976	AD
HOXC9	12q13.13	Cryptorchidism	142971	AD
HOXD10	2q31.1	Cryptorchidism	142984	AD
HOXD11	2q31.1	Cryptorchidism	142986	AD
HOXD12	2q31.1	Cryptorchidism	142988	AD
HOXD13	2q31.1	Cryptorchidism	142989	AD

HOXD9	2q31.1	Cryptorchidism	142982	AD
HRAS	11p15.5	Cryptorchidism	190020	AD
HS6ST1	2q14.3	Abnormal hypothalamus development and function	604846	AD
HSD17B3	9q22.32	Abnormal development of reproductive organs	605573	AR
HSD3B2	1p12	Adrenal gland dysfunction	613890	AR
HSF2	6q22.31	Meiotic arrest/Spermiogenesis defect	140581	AD
HSFY1	Yq11.222	Azoospermia	400029	YL
HSP90AA1	14q32.31	Varicocele	140571	Not reported
HYDIN	16q22.2	Spermiogenesis defect	610812	AR
IGF1	12q23.2	Hypogonadotropic hypogonadism	147440	Not reported
IGSF10	3q25.1	Delayed puberty	617351	AD
IL17RD	3p14.3	Abnormal hypothalamus development and function	606807	AR
INSL3	19p13.11	Abnormal development of reproductive organs	146738	AD
INSL6	9p24.1	Asthenozoospermia/cryptorchidism	606414	AR
IZUMO1	19q13.33	Fertilization failure	609278	AR
JAG1	20p12.2	Abnormal hypothalamus development and function	601920	AD
KATNAL1	13q12.3	Non-obstructive azoospermia	614764	AR
KATNB1	16q21	Oligoasthenozoospermia	602703	AR
KCNQ1	11p15.5-p15.4	Sertoli cell only syndrome	607542	Not reported
KDM3A	2p11.2	Spermiogenesis defect	611512	AD
KISS1	1q32.1	Central Precocious Puberty	603286	AR
KISS1R	19p13.3	Abnormal hypothalamus development and function	604161	AR
KIT	4q12	Azoospermia	164920	AD
KLB	4p14	Hypogonadotropic hypogonadism	611135	AR
KLF6	10p15.2	Hypospadias	602053	AD
KLHL10	17q21.2	Meiotic arrest/Spermiogenesis defect	608778	AD
KLK12	19q13.41	Male infertility	605539	Not reported
KLK14	19q13.41	Male infertility	606135	AR
KLK15	19q13.33	Male infertility	610601	Not reported
KLK3	19q13.33	Male infertility	176820	Not reported
KLK4	19q13.41	Male infertility	603767	AR
KLK6	19q13.3	Male infertility	602652	Not reported
KLK7	19q13.33	Male infertility	604438	Not reported
KLK8	19q13	Male infertility	605644	AR
KRAS	12p12.1	Cryptorchidism	190070	AD
LEP	7q32.1	Hypogonadotropic hypogonadism	164160	AR
LEPR	1p31.3	Hypogonadotropic hypogonadism	601007	Not reported
LHB	19q13.3	Pituitary gland dysfunction	152780	AR
LHCGR	2p16.3	Leydig cell dysfunction	152790	AR
LHX3	9q34.3	Hypogonadotropic hypogonadism	600577	AR
LHX4	1q25.2	Hypogonadotropic hypogonadism	602146	AD
LHX9	1q31.3	46,XY Disorders of Sex Development	606066	AD
LIF	22q12.2	Hypogonadotropic hypogonadism	159540	Not reported
LIMK2	22q12.2	Testicular failure	601988	AR

LIPE	19q13.2	Testicular failure	151750	AR
LRR6	8q24.22	Spermiogenesis defect	614930	AR
LRWD1	7q22.1	Sertoli cell only syndrome	615167	Not reported
LSM6	4q31.22	Azoospermia	607286	AD
MAGEB4	Xp21.2	Sertoli cell defect/Pre-meiotic defect/Meiotic defect	300153	XL
MAMLD1	Xq28	Abnormal development of reproductive organs	300120	XL
MAP2K1	15q22.31	Cryptorchidism	176872	AD
MAP2K2	19p13.3	Abnormal development of reproductive organs	601263	AD
MAP3K1	5q11.2	Abnormal development of reproductive organs	613762	AD
MASTL	10p12.1	Hypogonadotropic Hypogonadism	608221	AD
MC4R	18q21.32	Abnormal hypothalamus development and function	155541	AR
MCM8	20p12.3	Azoospermia	608187	AR
MEI1	22q13.2	Meiotic arrest	608797	AR
MEIG1	10p13	Azoospermia and oligozoospermia	614174	AR
MEIOB	16p13.3	Meiotic arrest	617670	AR
MET	7q31	Hypogonadotropic Hypogonadism	164860	AD
MIER3	5q11.2	46,XY Disorders of Sex Development	-	Not reported
MLH3	14q24.3	Meiotic arrest	604395	AR
MNS1	15q21.3	Spermiogenesis defect	610766	AR
MSH4	1p31.1	Maturation arrest	602105	AR
MT-ATP6	mitochondria	Varicocele/OAT	516060	NA
MT-ATP8	mitochondria	Varicocele/OAT	516070	NA
MT-CO1	mitochondria	Asthenozoospermia	516030	NA
MT-CO2	mitochondria	Asthenozoospermia	516040	NA
MT-CO3	mitochondria	Varicocele/OAT	516050	NA
MT-CYB	mitochondria	Asthenozoospermia	516020	NA
MTHFR	1p36.22	Azoospermia	236250	AD
MT-ND1	mitochondria	Male infertility	516000	NA
MT-ND2	mitochondria	OAT	516001	NA
MT-ND3	mitochondria	Varicocele/OAT	516002	NA
MT-ND4	mitochondria	Asthenozoospermia	516003	NA
MT-ND4L	mitochondria	Varicocele	516004	NA
MT-ND5	mitochondria	OAT	516005	NA
MTOR	1p36.22	Abnormal hypothalamus development and function	601231	AD
MT-RNR1	mitochondria	Asthenozoospermia and teratozoospermia	561000	NA
MT-RNR2	mitochondria	Asthenozoospermia and teratozoospermia	561010	NA
MT-TG	mitochondria	Varicocele	590035	NA
MT-TH	mitochondria	Asthenozoospermia	590040	NA
MT-TK	mitochondria	OAT	590060	NA
MT-TL1	mitochondria	Asthenozoospermia	590050	NA
MT-TR	mitochondria	Varicocele	590005	NA
MT-TS1	mitochondria	OAT	590080	NA
NANOS1	10q26.11	OAT	608226	AD
NANOS2	19q13.32	Sertoli cell-only syndrome	608228	AR

NANOS3	19p13.13	Azoospermia and oligozoospermia	608229	Not reported
NECTIN2	19q13.32	Teratozoospermia	600798	AD
NECTIN3	3q13.13	Teratozoospermia	607147	AD
NLRP14	11p15.4	Azoospermia and severe oligozoospermia	609665	AR
NLRP3	1q44	Sertoli cell-only syndrome/Pre-meiotic arrest/Meiotic arrest/Spermiogenesis defect	606416	AD
NNT	5p12	Adrenal gland dysfunction	607878	AR
NOS1	12q24.22	Abnormal hypothalamus development and function	163731	AD
NOTCH1	9q34.3	Abnormal hypothalamus development and function	190198	AD
NOTCH2	1p12	Persistent mullerian duct syndrome	618026	AD
NPAS2	2q11.2	Non-obstructive azoospermia	603347	AR
NPFFR1	10q22.1	Gonadotrophin-releasing hormone-dependent pubertal disorders	607448	AR
NPVF	7p15.3	Gonadotrophin-releasing hormone-dependent pubertal disorders	616984	AR
NPY1R	4q32.2	Hypogonadotropic hypogonadism	162641	Not reported
NROB1	Xp21.2	Adrenal gland dysfunction/46,XX Disorders of Sex Development	300473	XL
NRSA1	9q33.3	Abnormal development of reproductive organs	612965	AD
NRAS	1p13.2	Abnormal development of reproductive organs	164790	AD
NRP1	10p11.22	Hypogonadotropic hypogonadism	602069	Not reported
NRP2	2q33.3	Hypogonadotropic hypogonadism	602070	Not reported
NSMF	9q34.3	Abnormal hypothalamus development and function	608137	AD
NSUN7	4p14	Asthenozoospermia	617185	Not reported
NTN1	17p13.1	Kallmann syndrome; with or without anosmia	601614	AD
NXF2	Xq22.1	Non-obstructive azoospermia	300315	XL
OAZ3	1q21.3	Azoospermia	605138	AR
ODF1	8q22.3	Male infertility	182878	AR
OTX2	14q22.3	Hypogonadotropic hypogonadism	600037	AD
OXCT2	1p34.2	Male infertility	610289	Not reported
PACRG	6q26	Male infertility	608427	AR
PANK2	20p13	Abnormal development of vas deferens	606157	AR
PAPOLB	7p22.1	Male infertility	607436	Not reported
PAX6	11p13	Hypogonadotropic hypogonadism	607108	Not reported
PCDH10	4q28.3	Azoospermia	608286	AD
PCSK1	5q15	Kallmann syndrome	162150	AR
PDE3A	12p12.2	Kallmann syndrome	123805	AD
PDHA2	4q22.3	Sertoli cell-only syndrome/Pre-meiotic arrest/Meiotic arrest/Spermiogenesis defect	179061	AR
PGAM4	Xq21.1	Non-obstructive azoospermia	300567	Not reported
PICK1	22q13.1	Globozoospermia	605926	AR
PIH1D3	Xq22.3	Spermiogenesis defect	300933	XL
PIN1	19p13	Hypogonadotropic hypogonadism	601052	AD
PIWIL1	12q24.33	Azoospermia	605571	AD
PKD1	16p13.3	Spermiogenesis defect	601313	AD
PLA2G6	22q13.1	Asthenozoospermia	603604	AR
PLCZ1	12p12.3	Fertilization defect	608075	AR
PLEKHA5	12p12.3	Abnormal hypothalamus development and function	607770	AD
PLK4	4q28.1	Sertoli cell-only syndrome	605031	AD

PLXNA1	3q21.3	Abnormal hypothalamus development and function	601055	AD
PLXNB1	3p21.31	Hypogonadotropic hypogonadism	601053	Not reported
PMFBP1	16q22.2	Spermiogenesis defect	618085	AR
POLG	15q26.1	Male infertility	174763	AR
POLR3B	12q23.3	Abnormal hypothalamus development and function	614366	AR
POR	7q11.23	Testicular regression syndrome	124015	AR
POU1F1	3p11.2	Hypogonadotropic hypogonadism	173110	Not reported
POU3F2	6q16.1	Hypogonadotropic hypogonadism	600494	AD
PPP1CC	12q24.11	Testicular failure	176914	AD
PRDM9	5p14.2	Non-obstructive azoospermia	609760	AR
PRKAR1A	17q24.2	Carney complex with oligoteratozoospermia or azoospermia	188830	AD
PRM1	16p13.13	Oligozoospermia or azoospermia	182880	Not reported
PRM2	16p13.13	Oligozoospermia or azoospermia	182890	Not reported
PRM3	16p13.13	Oligozoospermia or azoospermia	-	AR
PROK2	3p13	Abnormal hypothalamus development and function	607002	AR
PROKR2	20p12.3	Abnormal hypothalamus development and function	607123	AR
PSMC3IP	17q21.2	Meiotic arrest	608665	AR
PTN	7q33	Hypospadias	162095	AD
PTPN11	12q24.13	Cryptorchidism	176876	AD
PUM2	2p24.1	Spermatogenic failure	607205	AD
PYGO2	1q21.3	Oligozoospermia or azoospermia	606903	AD
RAD21L1	20p13	Sertoli cell only and meiotic arrest	-	Not reported
RAF1	3p25.2	Cryptorchidism	164760	AD
RBMX	Xq26.3	Non-obstructive azoospermia	300199	XL
RBMXL2	11p15	Meiotic arrest/Spermiogenesis defect	605444	AD
RBMX1A1	Yq11.223	Azoospermia	400006	YL
RD3	1q32.3	Hypogonadotropic hypogonadism and Kallmann syndrome	180040	AR
REC8	14q12	Meiotic arrest	608193	AD
RELN	7q22.1	Abnormal hypothalamus development and function	600514	AD
RHOXF1	Xq24	Severe oligozoospermia	300446	XL
RHOXF2	Xq24	Severe oligozoospermia	300447	XL
RHOXF2B	Xq24	Severe oligozoospermia	300973	XL
RNF17	13q12.12	Non-obstructive azoospermia	605793	Not reported
RNF220	1p34.1	Spermiogenesis defect	616136	AR
ROBO3	11q24.2	Hypogonadotropic Hypogonadism	608630	Not reported
ROCK1	18q11.1	46,XX Disorders of Sex Development	601702	AD
RSPH1	21q22.3	Spermiogenesis defect	609314	AR
RSPH3	6q25.3	Spermiogenesis defect	615876	AR
RSPH9	6p21.1	Spermiogenesis defect	612648	AR
RSPO1	1p34.3	Abnormal development of reproductive organs	609595	AR
RXFP2	13q13.1	Cryptorchidism	606655	AR
SBF1	22q13.33	Testicular failure	603560	AD
SDHA	5p15.33	Azoospermia and oligozoospermia	600857	AR
SDHB	1p36.13	Azoospermia and oligozoospermia	185470	AR

SDHC	1q23.3	Azoospermia and oligozoospermia	602413	AR
SDHD	11q23.1	Azoospermia and oligozoospermia	602690	AR
SECISBP2	9q22.2	Spermiogenesis defect	607693	AR
SEMA3A	7q21.11	Abnormal hypothalamus development and function	603961	AD
SEMA3E	7q21.11	Kallmann syndrome	608166	AD
SEMA4D	9q22.2	Hypogonadotropic Hypogonadism	601866	Not reported
SEMA7A	15q24.1	Hypogonadotropic Hypogonadism	607961	AD
SEMG1	20q13.12	Asthenozoospermia	182140	AR
SEMG2	20q13.12	Asthenozoospermia	182141	AR
SEPT12	16p13.3	Non-obstructive azoospermia	611562	AD
SERPINA5	14q32.13	Azoospermia and teratozoospermia	601841	AR
SIN3A	15q24.2	Azoospermia	607776	AD
SLC10A7	4q31.22	Hypogonadotropic hypogonadism	611459	AD
SLC26A3	7q22.3-q31.1	Hydrocele/spermatocoele	126650	AR
SLC26A8	6p21.31	Asthenozoospermia	608480	AR
SLC34A2	4p15.2	Testicular Microlithiasis	604217	AR
SLC9C1	3q13.2	Asthenozoospermia	612738	AR
SLIT2	4p15.31	Abnormal hypothalamus development and function	603746	AD
SOHLH1	9q34.3	Non-obstructive azoospermia	610224	AR
SOS1	2p22.1	Abnormal development of reproductive organs	182530	AD
SOX10	22q13.1	Abnormal hypothalamus development and function	602229	AD
SOX2	3q26.33	Abnormal hypothalamus development and function	184429	AD
SOX3	Xq27.1	Abnormal development of reproductive organs	313430	XL
SOX8	16p13.3	Abnormal development of reproductive organs	605923	AD
SOX9	17q24.3	Abnormal development of reproductive organs	608160	AD
SPACA1	6q15	Globozoospermia	612739	AR
SPACA3	17q11.2	Infertility	612749	AR
SPAG16	2q34	Asthenozoospermia	612173	AR
SPAG17	1p12	Spermiogenesis defect	616554	AR
SPATA16	3q26.31	Spermiogenesis defect	609856	AR
SPATA17	1q41	Non-obstructive azoospermia	611032	AR
SPATA33	16q24.3	Non-obstructive azoospermia	615409	AR
SPINK2	4q12	Spermiogenesis defect	605753	AR
SPO11	20q13.31	Meiotic arrest	605114	AR
SPRY2	13q31.1	46,XX Disorders of sex development	602466	AD
SPRY4	5q31.3	Hypogonadotropic hypogonadism	607984	AD
SRA1	5q31.3	Abnormal hypothalamus development and function	603819	AR
SRD5A2	2p23.1	Abnormal development of reproductive organs	607306	AR
SRY	Yp11.2	Abnormal development of reproductive organs	400045	YL
STK36	2q35	Spermiogenesis defect	607652	AR
STRA8	7q33	Non-obstructive azoospermia	609987	AR
STRC	15q15.3	Asthenozoospermia with deafness	606440	AR
STS	Xp22.31	Hypogonadotropic hypogonadism	300747	XL
STX2	12q24.33	Pre-meiotic arrest/Meiotic arrest	132350	AR

STYX	14q22.1	Testicular failure	615814	AD
SUN5	20q11.21	Spermiogenesis defect	613942	AR
SYCE1	10q26.3	Meiotic arrest	611486	AR
SYCP3	12q23.2	Meiotic arrest	604759	AD
TAC3	12q13.3	Abnormal hypothalamus development and function	162330	AR
TACR3	4q24	Abnormal hypothalamus development and function	162332	AR
TAF4B	18q11.1	Sertoli cell-only syndrome/pre-meiotic arrest	601689	AR
TAF7L	Xq22.1	Azoospermia	300314	XL
TBPL1	6q23.2	Testicular failure	605521	AD
TDRD6	6p12.3	Spermiogenesis defect	611200	AR
TDRD7	9q22.33	Spermiogenesis defect	611258	AR
TDRD9	14q32.33	Spermiogenesis defect	617963	AR
TEKT2	1p34.3	Asthenozoospermia	608953	AR
TEX11	Xp11	Meiotic arrest	300311	XL
TEX13A	Xq22.3	Azoospermia	300312	XL
TEX13B	Xq22.3	Azoospermia	300313	XL
TEX14	17q22	Meiotic arrest	605792	AR
TEX15	8p12	Meiotic arrest	605795	AR
TLE4	9q21.31	Hypogonadotropic hypogonadism	605132	Not reported
TNP1	2q35	Oligozoospermia or varicocele	190231	AR
TNP2	16p13.13	Oligozoospermia	190232	AR
TP63	3q28	Syndromic teratozoospermia	603273	AD
TRAPPC9	8q24.3	Hypogonadotropic hypogonadism and Kallmann syndrome	611966	AR
TRIM37	17q22	Leydig cell dysfunction/Sertoli cell-only syndrome/pre-meiotic arrest/Meiotic arrest/Spermiogenesis defect	605073	AR
TSC22D3	Xq22.3	Sertoli cell only syndrome	300506	XL
TSPAN11	12p11.21	Hypogonadotropic hypogonadism and Kallmann syndrome	-	AD
TSPY1	Yp11.2	Male infertility	480100	YL
TSPYL1	6q22.1	Azoospermia/Disorder of Sex Development	604714	AD
TSPYL5	8q22.1	Azoospermia	614721	AD
TSSK1B	5q22.2	Male infertility	610709	Not reported
TSSK2	22q11.21	Male infertility	610710	Not reported
TSSK4	14q12	Azoospermia and oligozoospermia	610711	AD
TTL5	14q24.3	Spermiogenesis defect	612268	AR
TYRO3	15q15.1	Hypogonadotropic hypogonadism	600341	AR
UBE2B	5q31.1	Spermiogenesis defect	179095	AR
UBR2	6p21.1	Meiotic arrest/Spermiogenesis defect	609134	AD
USP1	1p31.3	Non-obstructive azoospermia	603478	AD
USP26	Xq26.2	Pre-meiotic arrest/Meiotic arrest/Spermiogenesis defect	300309	XL
USP9Y	Yq11.221	Azoospermia	400005	YL
UTP14C	13q14.3	Azoospermia and oligozoospermia	608969	AR
VAMP7	Xq28 and Yq12	Abnormal development of reproductive organs	300053	Copy gain
WDR11	10q26.12	Abnormal hypothalamus development and function	606417	AD
WDR66	12q24.31	Spermiogenesis defect	618146	AR
WNK3	Xp11.22	Azoospermia	300358	XL

WNT4	1p36.12	46,XX Disorders of sex development	603490	AD
WT1	11p13	Abnormal development of reproductive organs	607102	AD
WTAP	6q25.3	Hypospadias	605442	AD
XRCC2	7q36.1	Meiotic arrest	600375	AR
YBX2	17p13.1	Azoospermia and oligozoospermia	611447	AD
ZMYND15	17p13.2	Spermiogenesis defect	614312	AR
ZNF214	11p15.4	Azoospermia and severe oligozoospermia	605015	Not reported
ZNF215	11p15.4	Azoospermia and severe oligozoospermia	605016	Not reported
ZPBP	7p12.2	Spermiogenesis defect	608498	AR

4. Caractérisation moléculaire de remaniements chromosomiques chez des patients atteints de déficience intellectuelle et/ou de malformations congénitales

Environ un tiers des anomalies congénitales sont secondaires à une anomalie chromosomique. Nous étudierons, dans cette partie, neuf patients atteints de déficience intellectuelle et/ou malformations congénitales (DI/MC). Cinq patients sont porteurs d'une anomalie chromosomique déséquilibrée identifiée au caryotype et quatre autres sont porteurs d'un remaniement chromosomique apparemment équilibré (RCAE) sur le caryotype. Ces remaniements ont été caractérisés par ACPA et par WGS afin de préciser le mécanisme de survenue de ces remaniements et de déterminer les gènes impliqués. Nous souhaitons ainsi mieux caractériser les corrélations génotype - phénotype et tenter d'identifier de nouveaux gènes candidats. Nous présenterons les résultats pour les anomalies déséquilibrées d'une part et pour les RCAE d'autre part.

4.1. Caractérisation moléculaire de remaniements chromosomiques déséquilibrés chez des patients atteints de déficience intellectuelle et/ou de malformations congénitales

Le tableau III résume les cinq patients présentant un déséquilibre chromosomique au niveau du caryotype.

4.1.1. *Patient 1*

Le 1^{er} patient est un garçon suivi depuis l'âge de neuf mois pour des problèmes d'arthrite inflammatoire. Il n'y a pas d'antécédent familial de déficience intellectuelle ou d'autre maladie neurologique. L'histoire prénatale était sans particularité et à la naissance, le patient pesait 3,7 kg, mesurait 53 cm et avait un périmètre crânien de 37 cm. A partir de l'âge de 9 mois, le patient a présenté des épisodes récurrents associant une forte fièvre, une arthrite et des éruptions cutanées transitoires, entraînant le diagnostic de maladie de Still à l'âge de 2 ans. À l'âge de 5 ans, il a subi une intervention chirurgicale pour testicule droit non descendu. L'examen clinique à l'âge de 12 ans a révélé un retard de croissance staturale, une microcéphalie, une dysmorphie faciale avec des oreilles bas implantées et décollées, des fentes palpébrales obliques vers le bas, un ptosis gauche, une bouche tombante, des lèvres fines, une microrétrognathie et des lentigines (Figure 74). L'examen du thorax a mis en évidence des mamelons largement espacés et un discret pectus excavatum.



Figure 74 : Patient à l'âge de 12 ans.

On peut noter la dysmorphie faciale : oreilles bas implantées et décollées, lèvres fines, microrétrognathie et lentigines

L'examen cardiaque a retrouvé une légère régurgitation mitrale. Au niveau des membres, de grandes articulations interphalangiennes proximales et des lésions inflammatoires liées à l'arthrite entraînant des limitations et des déformations secondaires ont été observées.

L'évaluation de son développement psychomoteur a révélé un retard important avec un score global de QI de 40 (test Matrix).

Le caryotype a mis en évidence un remaniement de structure déséquilibré avec présence de matériel additionnel au niveau des bras longs du chromosome 12 (Figure 75). Les caryotypes parentaux étaient normaux.

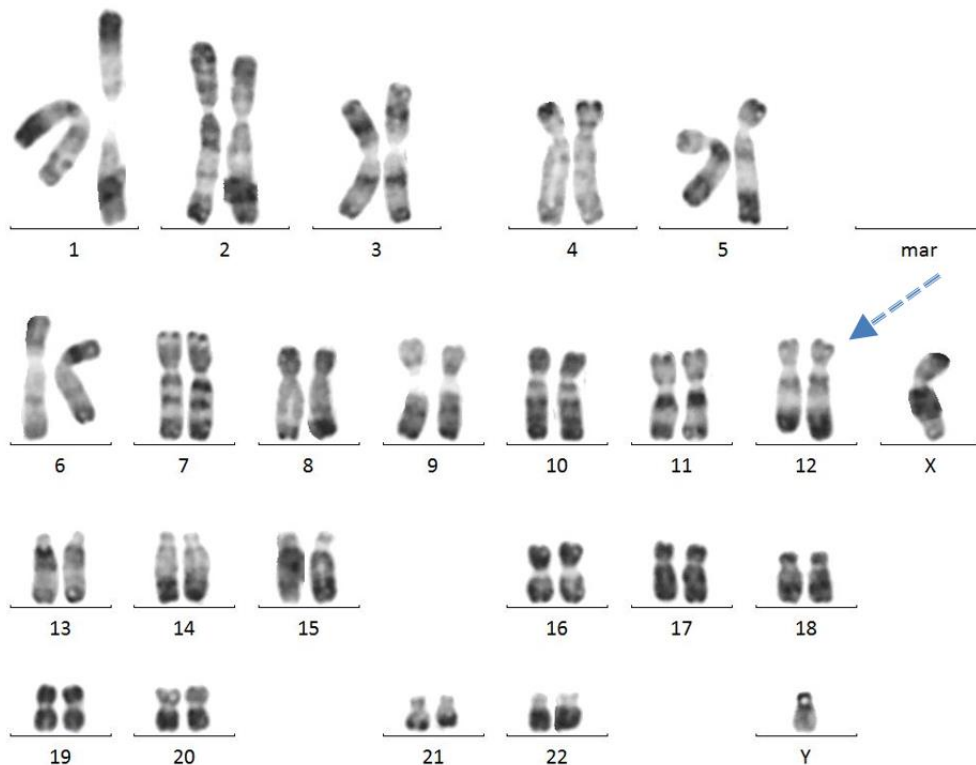


Figure 75 : Caryogramme (en bande R, 300-400 bandes) du patient 1 :

Présence de matériel additionnel sur les bras longs d'un chromosome 12 (flèche).

L'ACPA a montré un gain d'une copie de 11,7 Mb au niveau des bras longs du chromosome 12 dans la région 12q23.1q24.13 en faveur d'une duplication et dont les bornes sont les suivantes : chr12:101500368 et chr12:113290066 Mb (Hg19) (Figure 76).

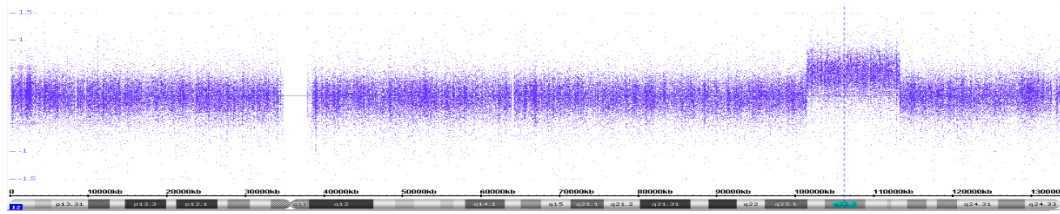


Figure 76 : Profil de SNP-array (Affymetrix.) Mise en évidence d'un gain d'une copie en chr12:101500368_113290066 Mb (Hg19) dans la région 12q23.1q24.13.

L'analyse de profondeur des données de WGS a révélé une augmentation de la profondeur de lecture au niveau de la région du chromosome 12 chr12:101499266_113290369 (Figure 77). L'étude des paires de reads a montré qu'il s'agissait d'une duplication en tandem et a permis de préciser à la base près les bornes de la duplication : chr12:101499268_113290370 (hg19) avec une micro-homologie de deux bases au point de jonction (Figure 78). Les points de cassure de la duplication interrompent deux gènes : *ANO4* en 12q23.1 et *RPH3A* en 12q24.13.

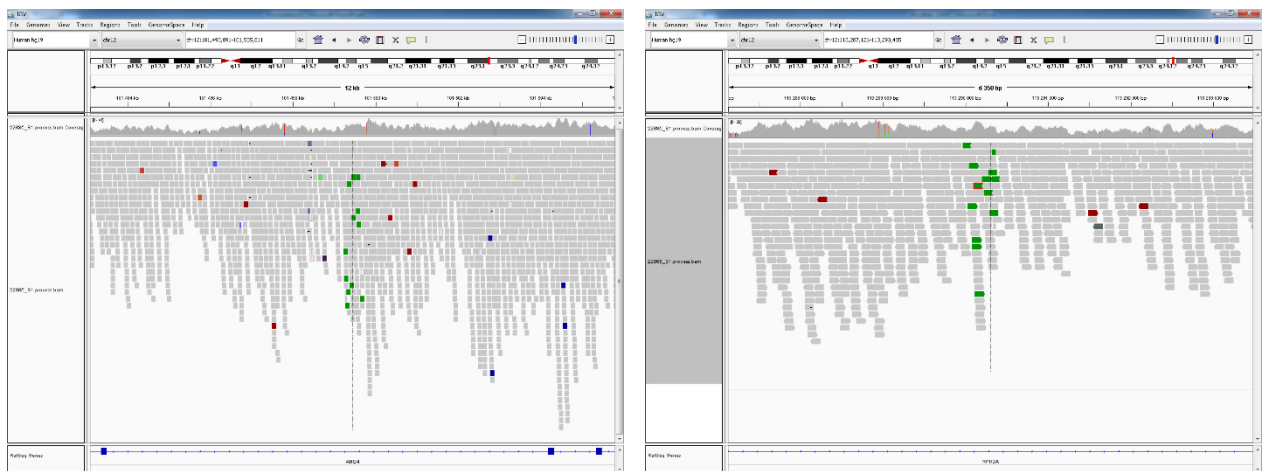


Figure 77 : Visualisation par IGV montrant une augmentation de la profondeur de lecture sur le bras long du chromosome 12 en chr12:101499266_113290369 Mb (hg19). Les reads verts correspondent aux paires mal alignées et leur orientation indique une duplication en tandem.

Chr12+ AGAGTCTCCCAGTTTTTATAGCTCTTCTTAGTCTAAATTCCAAATTCCT
 der (12) AGAGTCTCCCAGTTTTTATAGCTCAGATGGTTGGTCAAATTTGGCCACA
 Chr12+ CACAGATGGTTAAAAAAGAAAATCAGATGGTTGGTCAAATTTGGCCACA

chr12:113290370::101499268

Figure 78 : Séquence de jonction de la duplication chez le patient 1 d’après l’analyse des split-read. La séquence rouge correspond à la région chr12:113290345_113290395 et la séquence bleue correspond à la séquence chr12 :101499243_101499293 (hg19). La séquence rouge et bleu représente la jonction sur le dérivé. A noter une micro-homologie TC au niveau de la jonction.

Le point de cassure proximal est localisé dans un transposon hAT-Charlie (MER58C). Trois hypothèses peuvent expliquer l’origine de cette duplication : i) une translocation entre deux chromosomes homologues lors de la gamétogenèse chez un parent ; ii) un échange inégal entre chromatides sœurs du même chromosome ; iii) une insertion d’un segment d’une chromatide soit dans le chromosome homologue soit dans la chromatide sœur.

Comme les points de cassures interrompent deux gènes, nous avons vérifié l’hypothèse d’un gène de fusion pouvant aboutir à un transcrit *ANO4-RPH3A* par des études *in silico*. L’utilisation du portail ExPaSy a permis d’exclure la formation d’un transcrit de fusion et d’une protéine chimérique.

A ce jour, les gènes *ANO4* et *RPH3A* n’ont pas été rapportés en pathologie constitutionnelle. Seule une étude d’association a mis en évidence un lien entre le locus *NAA25*, proche de *RPH3A* sur le chromosome 12, et la polyarthrite rhumatoïde (164). Le gène *RPH3A* est fortement exprimé dans le cerveau et faiblement exprimé dans le sang d’après la base de données GTEX. L’étude d’expression du gène *RPH3A* chez le patient et chez des témoins normaux au niveau des lymphocytes sanguin n’a pas révélé d’expression significative de ce gène.

La région dupliquée comporte 119 gènes dont 86 sont des gènes OMIM (Tableau IV).

Tableau IV : Liste des gènes impliqués dans la duplication chez le patient 1.

<i>ANO4</i>	<i>HSP90B1</i>	<i>LOC100287944</i>	<i>DAO</i>	<i>ANAPC7</i>	<i>MAPKAPK5</i>
<i>SLC5A8</i>	<i>MIR3652</i>	<i>RFX4</i>	<i>SVOP</i>	<i>ARPC3</i>	<i>ADAM1A</i>
<i>UTP20</i>	<i>C12orf73</i>	<i>LOC100505978</i>	<i>USP30</i>	<i>GPN3</i>	<i>TMEM116</i>
<i>ARL1</i>	<i>TDG</i>	<i>RIC8B</i>	<i>USP30-AS1</i>	<i>FAM216A</i>	<i>ERP29</i>
<i>SPIC</i>	<i>GLT8D2</i>	<i>TMEM263</i>	<i>ALKBH2</i>	<i>VPS29</i>	<i>NAA25</i>
<i>MYBPC1</i>	<i>HCFC2</i>	<i>MTERF2</i>	<i>UNG</i>	<i>RAD9B</i>	<i>MIR3657</i>
<i>CHPT1</i>	<i>NFYB</i>	<i>CRY1</i>	<i>ACACB</i>	<i>PPTC7</i>	<i>TRAFD1</i>
<i>SYCP3</i>	<i>MIR7641-2</i>	<i>BTBD11</i>	<i>FOXN4</i>	<i>TCTN1</i>	<i>HECTD4</i>
<i>GNPTAB</i>	<i>TXNRD1</i>	<i>PWP1</i>	<i>LINC01486</i>	<i>HVCN1</i>	<i>MIR6861</i>
<i>DRAM1</i>	<i>EID3</i>	<i>PRDM4</i>	<i>MYO1H</i>	<i>PPP1CC</i>	<i>RPL6</i>
<i>WASHC3</i>	<i>CHST11</i>	<i>LOC101929162</i>	<i>KCTD10</i>	<i>CCDC63</i>	<i>PTPN11</i>
<i>NUP37</i>	<i>MIR3922</i>	<i>ASCL4</i>	<i>UBE3B</i>	<i>MYL2</i>	<i>RPH3A</i>
<i>PARBP</i>	<i>SLC41A2</i>	<i>LOC728739</i>	<i>MMAB</i>	<i>LINC01405</i>	<i>MIR1302-1</i>
<i>PMCH</i>	<i>C12orf45</i>	<i>WSCD2</i>	<i>MVK</i>	<i>LOC105369980</i>	
<i>IGF1</i>	<i>ALDH1L2</i>	<i>CMKLR1</i>	<i>FAM222A</i>	<i>CUX2</i>	
<i>LINC00485</i>	<i>LOC414300</i>	<i>LINC01498</i>	<i>FAM222A-AS1</i>	<i>MIR6760</i>	
<i>PAH</i>	<i>WASHC4</i>	<i>FICD</i>	<i>TRPV4</i>	<i>FAM109A</i>	
<i>ASCL1</i>	<i>APPL2</i>	<i>SART3</i>	<i>MIR4497</i>	<i>SH2B3</i>	
<i>LOC101929058</i>	<i>KCCAT198</i>	<i>ISCU</i>	<i>GLTP</i>	<i>ATXN2</i>	
<i>C12orf42</i>	<i>C12orf75</i>	<i>TMEM119</i>	<i>TCHP</i>	<i>ATXN2-AS</i>	
<i>LOC105369945</i>	<i>CASC18</i>	<i>SELPLG</i>	<i>GIT2</i>	<i>BRAP</i>	
<i>LINC02401</i>	<i>NUAK1</i>	<i>MIR4496</i>	<i>ANKRD13A</i>	<i>ACAD10</i>	
<i>STAB2</i>	<i>CKAP4</i>	<i>CORO1C</i>	<i>C12orf76</i>	<i>ALDH2</i>	
<i>NT5DC3</i>	<i>TCP11L2</i>	<i>SSH1</i>	<i>IFT81</i>	<i>MIR6761</i>	
<i>TTC41P</i>	<i>POLR3B</i>	<i>MIR619</i>	<i>ATP2A2</i>	<i>MAPKAPK5-AS1</i>	

Parmi ces gènes se trouve *PTPN11* (#MIM 176876), impliqué dans les syndromes de Noonan (NS) (# MIM 163950) et LEOPARD (LS) (# MIM 151100) de transmission autosomique dominante. Le syndrome de LEOPARD ou syndrome de Noonan à lentigines multiples est un syndrome à forte pénétrance et à expressivité variable. LEOPARD est un acronyme pour les principales caractéristiques de ce syndrome, défini par des lentigines multiples, des anomalies de la conduction cardiaque, un hypertélorisme, une sténose pulmonaire, des anomalies génitales, un retard de croissance et une surdité neurosensorielle (165). Le LS partage de nombreuses caractéristiques avec le NS en dehors de la présence des lentigines et d'une surdité. Des études moléculaires ont montré que ces deux syndromes sont des troubles alléliques causés par différentes mutations faux-sens dans *PTPN11*, un gène codant pour la protéine tyrosine phosphatase SHP-2. Des études moléculaires et biochimiques ont montré que les mutations

sont responsables d'un gain de fonction de la protéine (166). Les duplications de la région 12q23q24 ayant comme conséquence l'augmentation de dosage du gène *PTPN11* ont la même conséquence fonctionnelle que le gain de mutations observées dans LS et NS (167). A ce jour, onze patients porteurs de duplications isolées de la région 12q23q24 ont été décrits, dont deux cas présentant un Leopard/Noonan typique (167,168). Une variabilité clinique est cependant observée du fait de la taille variable des duplications. Par ailleurs, des études fonctionnelles ont rapportées le rôle de la protéine SHP-2 dans la physiopathologie de la polyarthrite rhumatoïde (169). Ainsi la duplication 12q23.1q24.13 identifiée chez le patient permet d'expliquer son phénotype et met en évidence un rôle majeur de *PTPN11*.

4.1.2. *Patient 2*

La seconde patiente est issue d'une union entre cousins germains. Deux antécédents de fausses couches ont été notés chez le couple. La grossesse n'a pas été suivie sur le plan médical et aucune échographie anténatale n'a été réalisée. L'accouchement s'est déroulé normalement. A la naissance, la patiente a présenté une microcéphalie, des problèmes respiratoires, une malformation cardiaque et des réflexes ralentis. Elle est décédée à l'âge de 6 mois.

Le caryotype a révélé un dérivé 18 d'une translocation réciproque maternelle entre les bras longs d'un chromosome 16 et les bras courts d'un chromosome 18 : 46,XX,der(18)t(16;18)(q21;p11)mat. Ce dérivé chromosomique entraîne chez la patiente une monosomie partielle du bras court du chromosome 18 et une trisomie partielle du bras long du chromosome 16 (Figure 79).

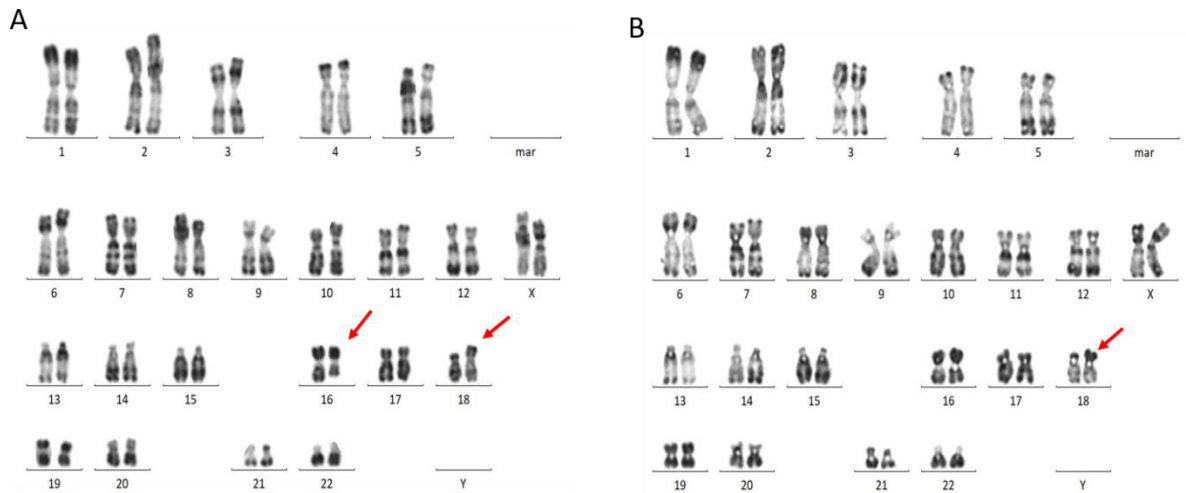


Figure 79 : A- Caryogramme (en bande R) de la mère du patient 2 présentant une translocation réciproque entre les chromosomes 16 et 18. B- Caryogramme (en bande R) de la patiente 2 présentant un dérivé 18 de la translocation maternelle.

L'ACPA a précisé la taille des fragments en gain et en perte chez la patiente. Elle a détecté un gain d'une copie au niveau du chromosome 16 de 18,9 Mb localisé en chr16:71209913_90155062 en 16q22.2q24.3, et une perte d'une copie au niveau du bras court du chromosome 18 de 8,9 Mb située en chr18:136226_9091008 en 18p11.22p11.32 (Figure 80). Ces CNVs comprennent de très nombreux gènes et expliquent le phénotype chez la patiente.

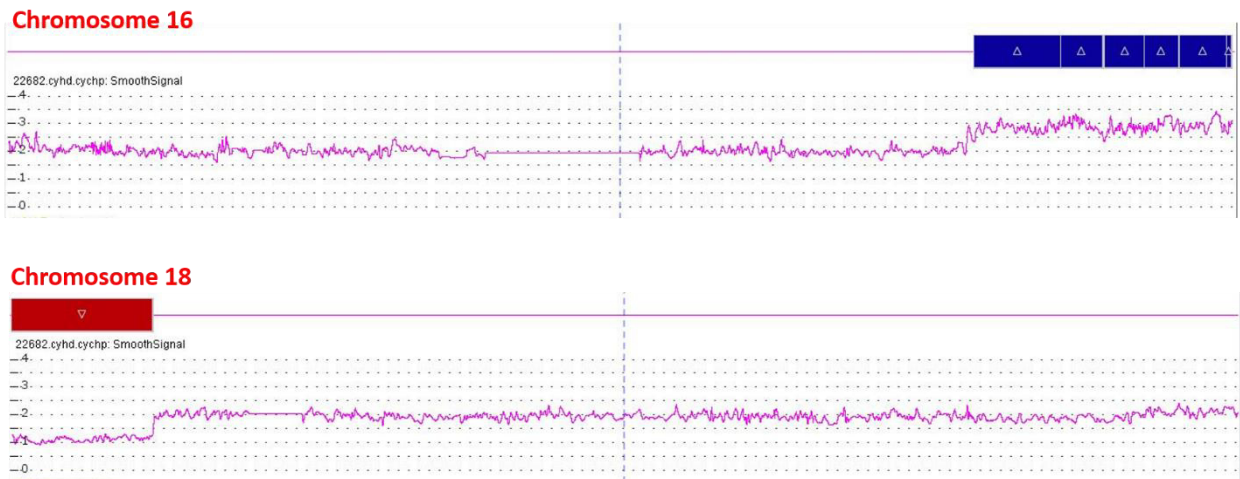


Figure 80 : Profil d'ACPA montrant le gain terminal d'une copie sur le chromosome 16 en 16q22.2q24.3 et la perte terminale d'une copie sur le chromosome 18 en 18p11.22p11.32 chez la patiente 2 (SNP-array, affymetrix).

Le WGS a détecté une diminution de profondeur de reads au niveau du chromosome 18 du télomère 18pter jusqu'à la position chr18:9092437 (Hg19) en 18p11.22 et une augmentation de profondeur des reads sur le chromosome 16 à partir de la position chr16:70592367 en 16q22.1 jusqu'au télomère 16qter (Figure 81). L'étude des paires de reads et des splits-read confirme la séquence de jonction et la position des points de cassure pour le dérivé 18. Une étude en WGS chez la mère aurait été nécessaire pour déterminer avec certitude la position des points de cassure sur le dérivé 16.

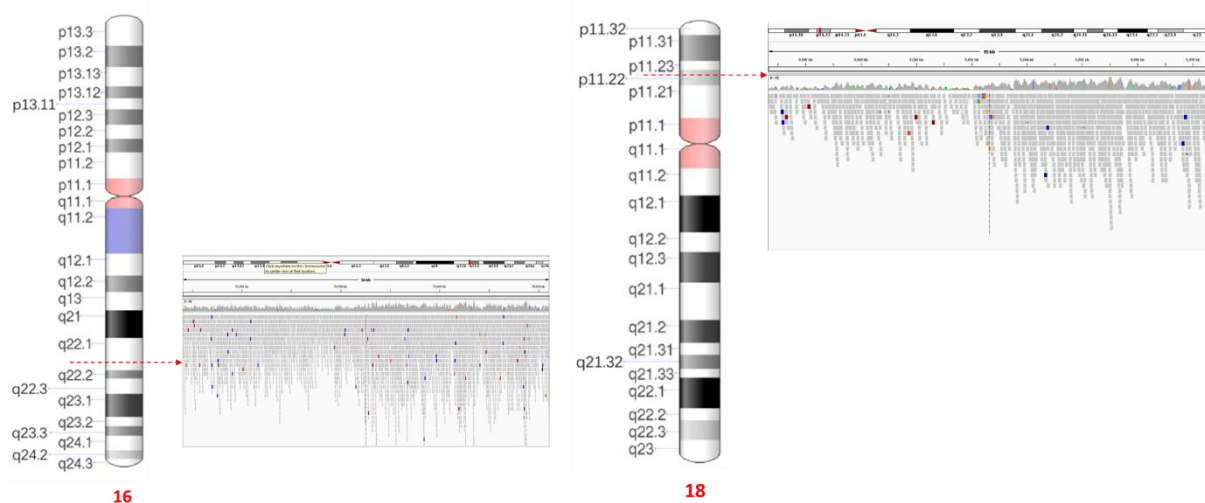


Figure 81 : Image illustrant les points de cassures sur le chromosome 16 et le chromosome 18 détectés par WGS chez la patiente 2. Pour chaque chromosome : idéogramme du chromosome avec point de cassure indiqué par la flèche rouge (à gauche) et image IGV à droite.

Le point de cassure sur le chromosome 16 est localisé dans un transposon hAT-Charlie (MER33) et le point de cassure sur le chromosome 18 est localisé dans une région répétitive du type SINE de la famille AluSg. Il existe une micro-homologie de deux paires de bases au niveau de la jonction (Figure 82). Les informations recueillies ne permettent pas de faire une hypothèse sur le mécanisme de survenue de la translocation.

der(18): chr16:70592367::chr18:9092437

```
chr16-      CAGCAATTTTATATTAACTACATGTTAGAATGATTTTCAGATATACAGA
der (18)    CAGCAATTTTATATTAACTACATTTTTGTATTTTAGTAGAGACGGGG
chr18+     TGTGTGCCACCACACCTGGCTAATTTTTGTATTTTAGTAGAGACGGGG
```

Figure 82 : Séquence de jonction du dérivé 18 d’après l’analyse des split-read chez la patiente 2. La séquence rouge (haut) correspond à la séquence de référence du chromosome 16 et la séquence bleue (bas) correspond à la séquence de référence du chromosome 18. Au milieu, la séquence du fragment de jonction sur le dérivé 18. A noter une micro-homologie AT au niveau de la jonction.

4.1.3. *Patient 3*

Le patient est issu d’un mariage entre cousins germains. Il a présenté à la naissance une macrocéphalie, une hypotonie, une malposition des oreilles, un hypertélorisme, des doigts chevauchants, des talons proéminents, une cryptorchidie et un micropenis. Le patient est décédé deux jours après sa naissance. Le caryotype a révélé un chromosome Y surnuméraire et une délétion au niveau des bras longs d’un chromosome 4 : 47,XYY,del(4)(q?) (Figure 83). Les caryotypes parentaux étaient normaux.

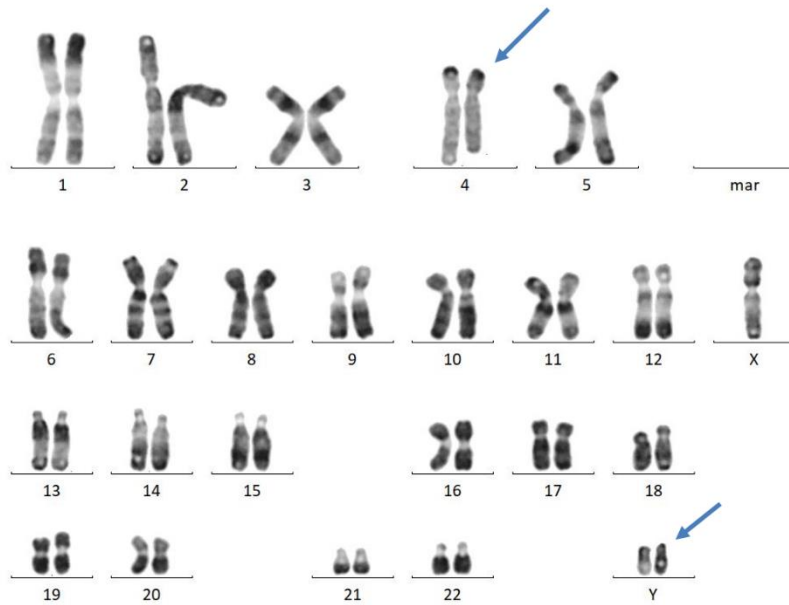


Figure 83 : Caryogramme (en bande R) du patient 3 présentant un chromosome Y surnuméraire et une délétion au niveau des bras longs d'un chromosome 4.

L'ACPA a révélé une perte d'une copie au niveau du bras long du chromosome 4 de 20.4 Mb localisée en 4q21.1q22.3 et la présence de deux copies du chromosome Y : arr [hg19] (X)x1,(Y)x2,4q21.1q22.3(77309280_97798507)x1 (Figure 84-85).

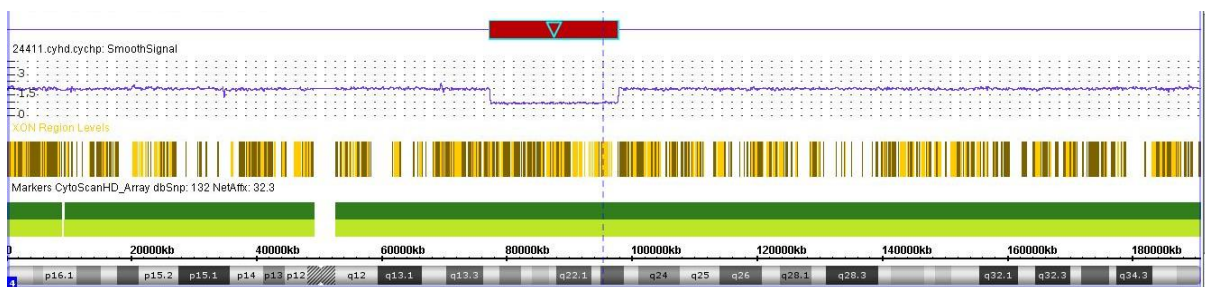


Figure 84 : Profil d'ACPA montrant une délétion interstitielle au niveau du chromosome 4 chez le patient 3 (SNP-array, Affymetrix).



Figure 85 : Profil d'ACPA montrant le gain d'une copie complète du chromosome Y. A noter que les régions PAR surnuméraires sont visualisées comme gain sur le chromosome X. (SNP-array, Affymetrix).

L'analyse des données de WGS a révélé une diminution de la profondeur des reads sur le chromosome 4 au niveau de la région chr4:77310419_97795300 (hg19) confirmant la délétion interstitielle vue sur le caryotype et en ACPA. De plus elle a révélé une augmentation de la profondeur des reads sur tout le chromosome Y (50X), avec profondeur double par rapport à celle du chromosome X (27X) et à peu près identique à celle des autosomes (51X).

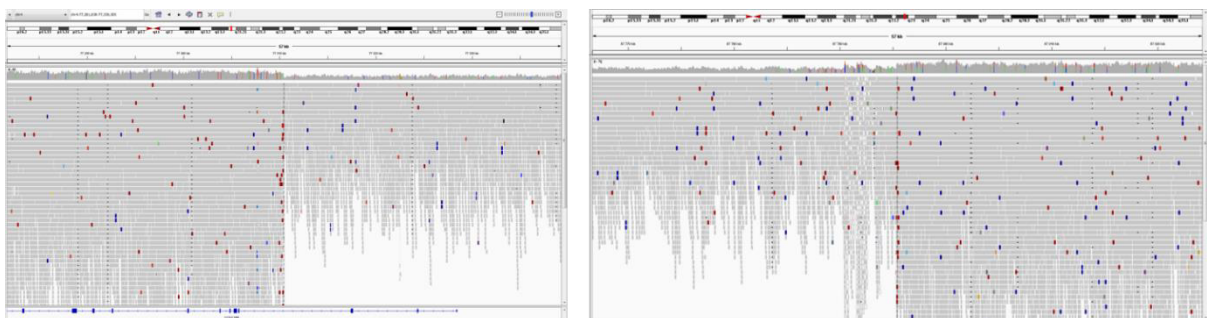


Figure 86 : Visualisation IGV des données WGS montrant une diminution de la profondeur de lecture sur le bras long du chromosome 4 en chr4:77310419_97795300Mb (hg19) .

chr4:77310419::97795300

Chr4+ TCTATCAGATAAAATTTAACAAAGAGATTAAAAATAATTAGAAAGAATCAAGCAGAAAT
Der(4) TCTATCAGATAAAATTTAACAAAGAGATTAACAGTGAATTTCAAAGCTTGCATGAGTGA
Chr4+ CCCAGCACCTTACATTAAATTCCTATTAACAGTGAATTTCAAAGCTTGCATGAGTGA

Figure 87 : Séquence de jonction de la délétion chez le patient 3 d'après l'analyse des split-read. La séquence rouge (haut) correspond à la séquence de référence du chromosome 4 proximale et la séquence bleue (bas) correspond à la séquence de référence du chromosome 4 distale. Au milieu, la séquence du fragment de jonction correspondant à la délétion. A noter une micro-homologie ATTAA au niveau de la jonction.

Le point de cassure au niveau de la région 4q21.1 est localisé dans un élément répété de type LINE (L1MA2) et interrompt le gène *CCDC158* (Coiled-Coil Domain Containing 158). Ce gène n'est pas impliqué en pathologie humaine à ce jour. Le deuxième point de cassure n'interrompt aucun gène. La présence d'une micro-homologie de 5 paires de bases au niveau de la jonction est évocateur d'un mécanisme répliatif de type MMBIR/FoSTeS (Figure 87). Le chromosome Y surnuméraire est dû à une non-disjonction méiotique chez le père.

La délétion comprend 72 gènes OMIM, dont 16 gènes responsables de pathologies autosomiques dominantes. Le tableau V représente la liste des gènes dominants inclus dans cette délétion.

Tableau V : Liste des gènes dominants impliqués dans la délétion au niveau du bras long du chromosome 4 chez le patient 3.

Maladie	Gène	Phenotype MIM number
Multiple system atrophy	<i>COQ2</i>	146500
Deafness, autosomal dominant 39, with dentinogenesis	<i>DSPP</i>	605594
Dentin dysplasia, type II	<i>DSPP</i>	125420
Dentinogenesis imperfecta, Shields type II	<i>DSPP</i>	125490
Dentinogenesis imperfecta, Shields type III	<i>DSPP</i>	125500
Polycystic kidney disease 2	<i>PKD2</i>	613095
Dementia, Lewy body	<i>SNCA</i>	127750
Parkinson disease 1	<i>SNCA</i>	168601
Parkinson disease 4	<i>SNCA</i>	605543
Muscular dystrophy, limb-girdle, autosomal dominant 3	<i>HNRNPDL</i>	609115
Adermatoglyphia	<i>SMARCAD1</i>	136000
Basan syndrome	<i>SMARCAD1</i>	129200
Brachydactyly, type A1, D	<i>BMPR1B</i>	616849
Brachydactyly, type A2	<i>BMPR1B</i>	112600
?Microcephaly 18, primary, autosomal dominant	<i>WDFY3</i>	617520
Uric acid concentration, serum, QTL1]	<i>ABCG2</i>	138900

4.1.4. *Patient 4 (Fœtus 1)*

Le patient 4 est un fœtus qui présentait un hygroma kystique de 3,37 mm à la 15^{ème} semaine d'aménorrhée (Patient 13 de l'étude de Nicolas CHATRON et al. (43)). Le caryotype sur examen direct des villosités chorales a montré une formule déséquilibrée avec deux dérivés de translocation robertsonienne entre les chromosomes 13 et 15 entraînant une trisomie 13 et une trisomie 15 : 46,XX,rob(13;15)(q10;q10),+der(13;15)(q10;q10) (Figure 88A). Le caryotype sur la culture des villosités chorales a montré une formule déséquilibrée avec une translocation robertsonienne entre les chromosomes 13 et 15 et un chromosome 13 surnuméraire: 46,XX,rob(13;15)(q10;q10),+13 (Figure 88B). Le caryotype sur liquide amniotique a révélé une formule à 45 chromosomes avec un dérivé de translocation robertsonienne entre les chromosomes 13 et 15 présentant des anomalies de structure supplémentaires (Figure 88C). La technique de FISH réalisée sur les métaphases de la culture du liquide amniotique a confirmé la présence d'anomalies structurales sur le dérivé 13 (Figure 88D). La sonde localisée au niveau de la région 13q14.3 se trouve près du télomère du bras long sur le dérivé 13.

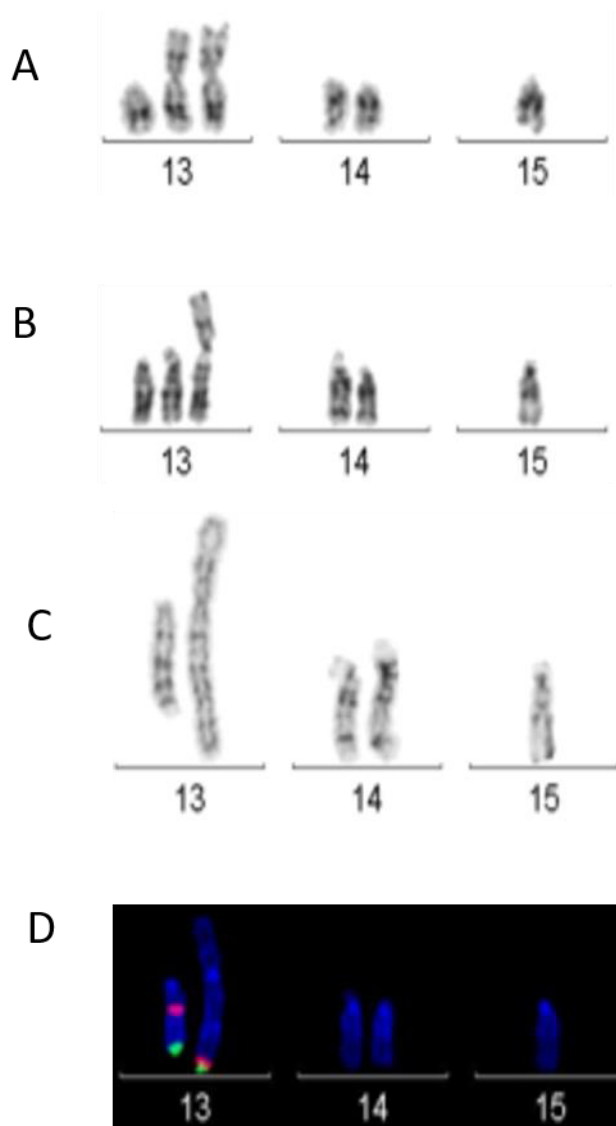


Figure 88 : A- Caryogramme (bande G) partielle du fœtus 1 de l'examen direct des villosités choriales : rob(13;15)(q10;q10),+rob(13;15)(q10;q10). B- Caryogramme (bande G) partielle de la culture longue des villosités choriales : rob(13;15)(q10;q10),+13. C- Caryogramme (bande G) partielle de la culture du liquide amniotique : rob(13;15) avec anomalie de structure au niveau du dérivé du chromosome 13. D- FISH métaphasiques sur la culture de liquide amniotique en utilisant les sondes 13q14.3 (Rouge) et 13qter (Vert) indiquant la présence d'anomalies structurales au niveau du der(13).

La CGH-array (Agilent) sur l'ADN issu du liquide amniotique a révélé la présence de 7 gains et de 4 pertes sur le chromosome 13 avec un déséquilibre total de 37,95 Mb (Figure 89).

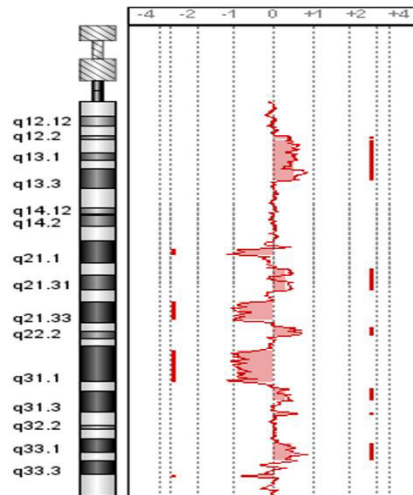


Figure 89 : Profil d'ACPA montrant les gains et les pertes au niveau du chromosome 13 détectés par la puce Agilent.

Le WGS à partir de l'ADN extrait du sang fœtal, a révélé 57 points de cassures (Figure 90), dont 15 points de cassures avec des gains et 8 points de cassures avec des pertes. Ces points de cassures sont représentés dans le tableau VI. La totalité des régions en gain représente 35,9 Mb et les régions en perte représentent 15,24 Mb. L'étude des split reads a montré que la majorité de ces points de cassures présente une microhomologie au niveau des points de cassures suggérant un mécanisme répliatif de type MMBIR /FoSTeS, permettant ainsi de définir un chromoanasythesis chez ce fœtus. Les régions en perte et en gain contiennent de nombreux gènes. Les points de cassures interrompent 8 gènes : *STARD13*, *NBEA*, *SMAD9*, *FAM155A*, *FGF14*, *RBM26*, *GPC5* et *GPC6*. Deux de ces gènes sont impliqués dans des maladies dominantes : *SMAD9* dans l'hypertension pulmonaire primaire et *FGF14* dans l'ataxie spinocérébelleuse. *GPC6* est un gène de transmission récessive impliqué dans l'omodysplasie. Les autres gènes jusqu'à présent ne sont pas rapportés en pathologie.

Tableau VI : Liste des points de cassure au niveau du chromosome 13 chez le patient 4. La séquence en Noire désigne la séquence du partenaire A, en rouge celle de B et en vert une séquence insérée.

Patient 4 (foetus 1)	Génome	Chromosome (read1)	Position	Chromosome (read2)	Position	Séquence selon les Split read	Mécanisme
Patient 4	Hg19	13	3424549	13	34266161		NA
Patient 4	Hg19	13	34266160	13	34245489	CATATGGTAGTTAAAAGACCATATGTCTTCCAATATGGAAGCCTACCATG CACTAATGTGCAATCATT/AAAGGTAGTCCGGACCCAAACAGTGAGCAGC A	NHEJ
Patient 4	Hg19	13	37442385			TTCCAATATGGAAGCCTACCATGCACTAATGTGCAATCAT/TAAAGGTAGT CCGGACCCAAACAGTGAGCAGCAGCAAGATTTATTGTCTAAAGTGAAAG A	MMBIR/FoSTeS
Patient 4	Hg19	13	38766265			AAATTGCCAGTGGAGCTTGCTAAAAATGCAGATTCCTGGCCCCATCCTTA GCAATTCGGATTCAATAGGTCAG/CAATTCGGATTCAATAGGTCACTGGCC	MMBIR/FoSTeS
Patient 4	Hg19	13	38609517	13	54657062	AACTAATCACAAGTCCTTTAGAATGGAGGTGGTGAATGGTGGTTGATTAA GCACTGGCCTTGAGG/CCAGTGACCTATTGAATCGGAATTGCTGACCTATT	MMBIR/FoSTeS
Patient 4	Hg19	13	54657058			GCATTGATAGGCACACATTAATAACCTGTGTATtTTATACACAggtataaaaTA CACACTTTTTTTTACATGTCTAGGCTCCAAAGACATCCTGG/ACACCTCTG AGAATATACTCGCACCTGCAATCCTAGCACTCTGGGAGGCCGAG	MMBIR/FoSTeS
Patient 4	Hg19	13	38907916	13	55020286	TAAAAAAAGTGTGTATTTTATACCTGTGTATAAAAATACACAGGTATTTA ATGTGTGCCTAT/CAATGCACACATTAAAAAAATGGCTACATAGCAATCGT	NHEJ - MMBIR/FoSTeS
Patient 4	Hg19	13	55020286	13	38907916	GAAAAAAGGTAAGTTCATCATTCCATATCAGTTATGATAAATGTTATGTT/ TTAAATCTGCCTTCATTAACCACTAAAGCAGCAATTGCTTCTCTTTCTTC	NHEJ - MMBIR/FoSTeS
Patient 4	Hg19	13	57053883	13	57056442	AATTATTATGAAGAAAGAGAAGCAATTGCTGCTTTAGTGGTTAATGAAGG CAG/ATTTAAACATAACATTTATCATAACTGATATGGAATGATGAACCTA	MMBIR/FoSTeS
Patient 4	Hg19	13	57056442	13	57053883	CCCTAGAGGTTGAATGCACTGGGTGTTAATTCTCAATAAATGCAT/TTTTA AAATAAAAAATTGCGGCTTCATATTTTCATTGTGAATATCCCATCACTTAT	MMBIR/FoSTeS
Patient 4	Hg19	4	1.9065E+10	13	59924924	AATGAAAATATGAAGCCGCAATTTTTATTTTAaaAATGCATTTATTGAGAA TTAACACCCAGTGCATTCAAC/CTCTAGGGAGGTTGGGGAAATTTTCATGA AAAATGTCCTCACA	MMBIR/FoSTeS
Patient 4	Hg19	13	59924924	4	1.91E+08	AACTCCTGGACTCAAGCTATCCTCCATCCTTGGCCTCCCAATTACTAAGC TTACAGGC/TTAGACTCCCTTCTAAGCCTCTTCATCTGAAGAGACTGACCC TCAGACTCTTCATCTATAAAGTTAATAAGATAAATAC	MMBIR/FoSTeS
Patient 4	Hg19	13	38804210			CCCAAATTACTAAGCTTACAGGCTTAGACTCCCTTCTAAGCCTCTTCATCT GAAGAGAC/TGACCCTCAGACTCTTCATCTATAAAGTTAATAAGATAAATAC	

Patient 4	Hg19	13	38804183	13	81159501	CAAAGCCTAGGCAAGGCTGTACTATTACAGCCTTTAAATCTGTACATAAA GTACAGCCTTTATGTACTTT/ATTGACATAAAGCGGGGAATAAAAGACT	MMBIR/FoSTeS
Patient 4	Hg19	13	81159501	13	38804183	CTGGAGGCCAACATCCTTAGCAAATAATGCAGAAACAGAAAACCAAT ACCAGGTGTTCTCACTTAC/TGTACAAAAAATACTGTCATCCCTGTAAGTG GTCTCATTTTAAATATAAAATATTAATAATTTTGTGATGGGTGCATTCAACA AAATGTCCTTCCTGTGTATCCCTGC	MMBIR/FoSTeS
Patient 4	Hg19	13	63006551	13	8159526		
Patient 4	Hg19	13	81159526			AATACTGTCATCCCTGTAAGTGGTCTCATTTTAAATATAAAATATTAATAAT TTGTTGATGGGTGCATT/AAACAAAATGTCTTCCTGTGTGTATCCCTGC	MMBIR/FoSTeS
Patient 4	Hg19	13	24453512	13	58859577		
Patient 4	Hg19	13	58859577	13	24453512	AGAAAATACTTGTCTATACCAGGAATGCTCAAAACAACCTTACTGGCAAG GTAAGT/TAAAAAttatATTAGATGATTTTCTAATAATAAATTCCTTTGGAA ATTCAGTGTCTTAATATTTTAAATC	MMBIR/FoSTeS
Patient 4	Hg19	13	24453536	13	79929560	AGAAAATACTTGTCTATACCAGGAATGCTCAAAACAACCTTACTGG/CAA GGTAACTGTAAAAAttatATTAGATGATTTTCTAATAATAAATTCCTTTGGAA ATTCAGTGTCTTAATATTTTAAATC	MMBIR/FoSTeS
Patient 4	Hg19	13	79929560	13	24453536	AGAAAATACTTGTCTATACCAGGAATGCTCAAAACAACCTTACTGGCAAG GTAAGTGTAAAAAttatATTAGATGATTTTCTAATA/TAAATTCCTTTGGAA ATTCAGTGTCTTAATATTTTAAATC	NHEJ
Patient 4	Hg19	13	58859576	13	69950856	GATTTAAAATATTAAGAACAGTGAATTTCCAAAGGAA/ATTTATATTAGA AAATCATCTAATAATAATTATTACAGTTACCTTGCCAGTAAAGTTGTTTTGA	NHEJ
Patient 4	Hg19	13	69950853	13	58859579	CCATGGTGGTTTGCTGTACCCATCAACCTATCATCTACATTA/GGTGAGTC AAAACAAGAAATAAAACAAATGTCCATCAAATGGTTATTGATAAACACA T	NHEJ - MMBIR/FoSTeS
Patient 4	Hg19	13	59348651	13	70728648	CAATAACCATTTGATGGACATTTGTTTTATTCTTGTGTTTACTC/ACCTAA TGATAGATGATAGGTTGATGGGTACAGCAAACCACCATGGCACGTGTATA	NHEJ - MMBIR/FoSTeS
Patient 4	Hg19	13	70728648	13	59348651	ATACTTTATAGCATTTATACTGAAATATAAAATATTT/ATATTTTATAAATA TTTATAAATATTTATAAAATATTTAAAAATATTTATAAATATTTATAAA ATATTTATATtATATTTATATAGTATAGTATACTATACTGATGCACCATC TTTTGGTGTGGTACTGAGGCGGGAATTA	MMBIR/FoSTeS
Patient 4	Hg19	13	67854880	13	70731432		
Patient 4	Hg19	13	70731432	13	67854880	CTGATAGGGGTGCATTTG/TTGAATACCACATTCTGGATGTTTTATTTTCT CATATTAGACATGTTTTTAAAAATCAAGATTATGTATATTTTCAAATA	MMBIR/FoSTeS

Patient 4	Hg19	13	74724020	13	75055605	TTAAAATAGAATGTTGTGTCTCAAGAAACAGTTTTGACAACTGATAGG/ GGTGCATTTTGTGAATACCACATTCTGGATGTTTTATTTTCTCATATTAG	MMBIR/FoSTeS
Patient 4	Hg19	13	75055605	13	74724020	GGTATACTAT/TTCCATCTGCCTCGGGGACTCTGCCCTAGATATGTGCTG GTTGCATTTTCTCACCTTCTGGTTTCTGCTCACCCAGTCAGTGAGGGCTT	MMBIR/FoSTeS
Patient 4	Hg19	13	72875571	13	76467126	CCATCCCTCAGAAAAAAGACTGACAAATACCCACTATTAGTAGT/GGT ATACTATTTCCATCTGCCTCGGGGACTCTGCCCTAGATATGTGCTGGTTG	MMBIR/FoSTeS
Patient 4	Hg19	13	76467126	13	72875571	GGAAAAACCTAGGTTTTATAGACACAAGAGTCTATACCCTGTCCAAGGCA TGCTGTCCAAG/GCATAGGATCTGGAATCAATACTGAACTTGTGATACTG	MMBIR/FoSTeS
Patient 4	Hg19	13	57153641	13	92227571	GTCAATGTGGTCTAATAGGAAAAACCTAGGTTTTATAGACACAAGAGTCT A/TACCCTGTCCAAGGCATGCTGTCCAAGGCATAGGATCTGGAATTCAATA	MMBIR/FoSTeS
Patient 4	Hg19	13	92227571	13	57153641	GTTTAAATGTAGCTGTCAAAGGATCTGTTTTCAGATACTGGTAGTGAAAT TACATTGACTTGAAA/CCAAGTCAATTGATTCACCAATAAATGCAATTTT	MMBIR/FoSTeS
Patient 4	Hg19	13	82293513	13	92227548	AGTATTTTTGTACAGCATATTTTTTAcaAAAATTGCATTTATTG/GTGAA TCAATTGACTTGGTTTCCAAGTCAATGTAATTTCACTACCAGTATCTGA	MMBIR/FoSTeS
Patient 4	Hg19	13	92227548	13	82293513	AGTATTTTTGTACAGCATA/TTTTTAAcaAAAATTGCATTTATTGGTGAA TCAATTGACTTGGTTTCCAAGTCAATGTAATTTCACTACCAGTATCTGA	MMBIR/FoSTeS
Patient 4	Hg19	13	82293493	13	95602658		
Patient 4	Hg19	13	95602658	13	95602658	GAGTTGTTCTGCAACTGTATATAAAGTGTATATAaagtaACACAGTTACTTT ATAACTGTGTACAAA/AAGTATTTTTGTACAGCATAAATATTTCCAAA	MMBIR/FoSTeS
Patient 4	Hg19	13	82293531	13	1.09E+08	GAGTTGTTCTGCAACT/GTATATAAAGTGTATATAaagtaACACAGTTACTTT ATAACTGTGTACAAAAGTATTTTTGTACAGCATAAATATTTCCAAA	MMBIR/FoSTeS
Patient 4	Hg19	13	10859932 6	13	82293531	TTTGTACAGCATAAATATTTCCAACTGTAT/ATACACTTTATATACACTTT ATATACACTTTATATACAGTTACTGTGTTACTCGTTACAGTTGTGTATAT TCAAAATATGAATACACACCACAATTACAG	MMBIR/FoSTeS
Patient 4	Hg19	13	58859492	13	91535853	CTGTAATTGTGGTGTGTATTTCATATTTTGA/ATATACACAAGTAAACGAG TAACACAGTAAACTGTATATAAAGTGTATATAAAGTGTATATAAAGTGTAT	MMBIR/FoSTeS
Patient 4	Hg19	13	91535858	13	58859497	GAAATAATTTATTTTGGAAAATGTAATAATATGCAATAAGTGA/ATGG CCCAGCACTTTCATTTTTAGGTCCTTACACAAAAACAGAAAATACTTGTCT	MMBIR/FoSTeS
Patient 4	Hg19	13	38803920	13	88872650	TTCCTGATTACCAGAAGAAATAATTTATTTTGGAAAATGTAATAATATAT GCAAATAAGTGAATGGC/CCAGCACTTTCATTTTTAGGTCCTTACACAAAA	MMBIR/FoSTeS
Patient 4	Hg19	13	88872640	13	38803930	CATAGAAGATAAAAACCAAGGATGTGCAGTGTACTGAAAGCCA/AAAGAG GTAATGTTACAGGAGAAAAAACAAGAAGGAGCTTATAAAGAGACCCT CTT	MMBIR/FoSTeS

Patient 4	Hg19	13	61834578	13	92103925	GAGGGTCTCTTTATAAGCTCCTTCTTTGTTTTTCTCCTGTAACA/TTACCTCTTTGGCTTTCAGTACACTGCACATCCTTGGTTTTATCTTCTATGCC	MMBIR/FoSTeS
Patient 4	Hg19	13	92103925	13	61834578	TATATATATATACACACACACACACATATATGTATGAGACAT/ATATGTCTCATAATATGAGACATGTGCAGTGATGGGTTGATAGagtGTGCACCCCACTCTAGCCTGGGTGA	MMBIR/FoSTeS
Patient 4	Hg19	13	67927722			TGTCACCCAGGCTAGAGTG/TGGGGTGCACACTCTATCAACCCATCACTGCACATGTCTCATATGTATGAGACATATATGTCTCATAATATATGTGTGTG	NHEJ
Patient 4	Hg19	13	10837425 6			AAAAACAAAATTCAGATAAATTCATAGAAAACCTCCACAA/CCTGTAA GCAATTGACCTGGCAAGTCAATTGTTGAATTGCAGGTCAAAAAGTTTTA A	NHEJ
Patient 4	Hg19	13	91626360	13	1.08E+08	CAGGAGGATGGGCAGGTCAAGGAAAGTGTTAGTAACCTCCTTTCACCTGATCATCACACACATCCCTACCGCAGATTCATGAATATTTAAAACCTTTTT TGA/CCTGCAATTCACAATTGACTTGCCAGGTCAATTGCTTACAGGTTGT GGAAGTTTTCTAT	NHEJ
Patient 4	Hg19	13	10837420 0	13	91626362	TAAAATATTCATGAATCTGCGGTAGGGATGTGTTGTGATGATC/AGGTGAA AGGAGGTTACTAAACACTTTCCTTGACCTGCCATCCTCCTGCCCTAGGA	NHEJ
Patient 4	Hg19	13	91218856	13	1.08E+08	CAGGAGGATGGGCAGGTCAAGGAAAGTGTTAGTAACCTCCTTTCAC/CT GATCATCACACACATCCCTACCGCAGATTCATGAATATTTAAAACCTTTTT	MMBIR/FoSTeS
Patient 4	Hg19	13	10837425 6	13	91218856	TTGGGCTATTAGTTTCTGAGCCTTAACAAAATCATTAGGCTAGAGTAAGC ATGGCA/TCCATACTTACATACTTACGAGTATACACAGTA AACGAGGCTTT TGGCGTGCTCTGTTTTATTCATACATGCTGGCTAGTGATAAATAAGAAACA TCACTGAGGGTT	MMBIR/FoSTeS
Patient 4	Hg19	13	94797420	13	1.04E+08	CTTACATACTTACGAGTATACACAGTA/AACGAGGCTTTTGGCGTGCTCTG TTTATTCATACATGCTGGCTAGTGATAAATAAGAAACATCACTGAGGGTT	MMBIR/FoSTeS
Patient 4	Hg19	13	10438497 7	13	94797424	CTACTCAAGAGGCTGAGGCAGGAGAGTGGCATGAACCTGGGAGG/TGGAT GGACGGATGATGAATGGATGGATGGATAAATATATTAATAATCTATACTT G	MMBIR/FoSTeS
Patient 4	Hg19	13	66300137	13	1.05E+08	TAGATTTTAATATATTTTATCCATCCATCCATTCATCATCCGTCCA/TCCAC CTCCCAGGTTTCATGCCACTCTCCTGCCTCAGCCTCTTGAGTAGCAGGGA	MMBIR/FoSTeS
Patient 4	Hg19	13	10491401 6	13	66300132	ATGAATGTCTTCCACTGACATCGGGCTTTCATTCCTGGCTTTGACCACAG CA/CTTTGGGAGGCTAAGGCGGGTGGATCATTACGTCAGGAGATCGAGAC	MMBIR/FoSTeS
Patient 4	Hg19	13	91220038	13	1.05E+08	TGACATCGGGCTTTCAATTCCTGGCTTTGACCA/CAGCACTTTGGGAGGCT AAGGCGGGTGGATCATTACGTCAGGAGATCGAGACCAGCCTAACCAATA T	MMBIR/FoSTeS
Patient 4	Hg19	13	10491473 6	13	91220038	GGAGATATTTTTGAAGgCaAAaATATACA...CaTGTCTCATCCCACACTGAA/ CCCACTTCTGGGTTCAAGCAATTCTGCCTCAGCCTCCTGAGTAGCTGGG	MMBIR/FoSTeS

Patient 4	Hg19	13	10250427 6	13	1.06E+08	GAAGTGACCTAAGCTGAAAATGGCAAGCTTCCGGAGATATTTTTGAAGGC AAAAATATACACATGTCTC/ATCCACACTGAACCCACTTCCTGGGTTCAA GCAATTC	MMBIR/FoSTeS
Patient 4	Hg19	13	10551891	13	1.03E+08	TGCCTGACTCCATGGGAAAACCATCTCCTTGA/GAAGTAGCCTCCTATGGT CACACAGCTTGTAGGTGGTGAATCAGAGGCTCAGTCAGGTCAGCCGTGC	NHEJ
Patient 4	Hg19	13	87948491	13	1.07E+08	GGTGT CAGAGAAGTGTCTGGCGCTTGGCTGCCTGACTCCATGGGAAATCC ATCTCC/TTGAGAAGTAGCCTCCTATGGTCACACAGCTTGTAGGTGGTGA	NHEJ
Patient 4	Hg19	13	10681066 3	13	87948491	GATGTATTAGTCATAGCCACTCTGGACAGCAATTTTGG/TAACCCACACAG TGTAAGGGAAATATCCAACCTCAGTAGCAACTAGCTTTTTAGTCTCACCT	NHEJ
Patient 4	Hg19	13	35907754	13	1.11E+08	GGATTTAGGGAAACAGAAGCCCCCAGGCAAACCTGGCAGGGATGTATTAG TCATAGCCACTCTGGACAGCAATTT/TGGTAACCCACACAGTGTAAAGGGA AA	NHEJ
Patient 4	Hg19	13	11051282 3	13	35907754	GGTCAGTGTTAGAGGGAGGCAGTGTTAGAGGGAGCCTAAGGCTGGCCTT GAGGCAGATTGACATG/TCAATATTAGGCCTTATTCATTCTTTCTAATTAA T	MMBIR/FoSTeS
Patient 4	Hg19	13	55159296	13	1.1E+08	AATTAGAAAGAATGAATAAGGCCTAATATTGACATGTCAATCTGCCTCAA GGCCAGCCTTAGGCTCCCTCTAACACT/GCCTCCCTCTAACACTGACCCCTC	MMBIR/FoSTeS
Patient 4	Hg19	13	11010586 3	13	55159296	CATGAGGAAATGTGAGTCTCTTTCTTTATAAATTACCCAGTCTTGGGTAG TATCTTTACAGCA/CTAAAGTACCGGCACATCCATAGGATGTGGGCCAAC	MMBIR/FoSTeS
Patient 4	Hg19	13	66304848	13	75065069	CTCACACCAACAGCAAATGAGACCATCTGTTGGCCACATCCT/ATGGATG TGCCGGTACTTTAGTGCTGTAAAGATACTACCCAAGACTGGGTAATTTAT	MMBIR/FoSTeS
Patient 4	Hg19	13	75065069	13	66304848	ATTGGTAATGTCTGGAGAGTTGATGGGTGCATCAACTCTCAATATTA AAAAGA/TTAAAATATGGTCTACAGTCAAAAATGATTTTAAAGTACTACCTCTGTC	MMBIR/FoSTeS
Patient 4	Hg19	13	75065169			ATCTTTTAAATATTGAGAGTTGATGCACCCATCAAC/TCTCCAGACATTAC CAATGTCCCTTTGTAGCAAAAATTGCACATTTGAGAATCACTGTTCTAG	MMBIR/FoSTeS
Patient 4	Hg19	13	82243975			CAAAAATTGCACATTTGAGAATCACTGTTCTAGAGCCACAAGCACATTGG GTATGCTGTC/CAAGGCCTATGCCTTGGCCTCCCAAAGTGCCGAGATTGCA	



Figure 90 : Circos plot du chromosome 13 illustrant les points de cassures chez le fœtus 1 (43).

4.1.5. *Patiente 5*

La cinquième patiente présente un retard de développement, des malformations vertébrales, des malpositions dentaires et une hernie de l’ovaire. Le caryotype a mis en évidence un remaniement de structure sur un chromosome 19 de type dicentrique (Figure 91).

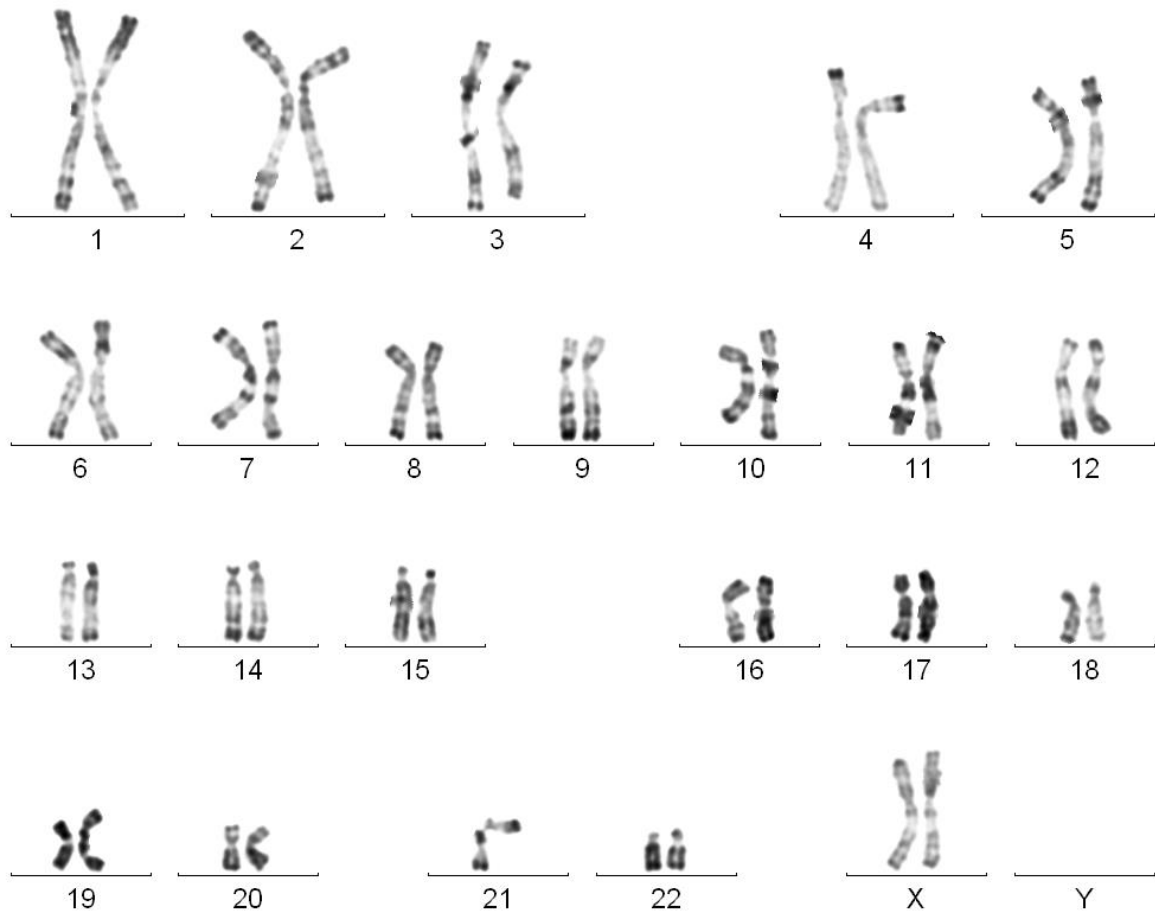


Figure 91 : Caryogramme (en bande R) de la patiente 5 présentant un chromosome 19 dicentrique avec des anomalies structurales.

L'utilisation de plusieurs sondes de FISH a révélé des gains et des pertes de certaines régions au niveau du chromosome 19. Les sondes : RP11-1012F23 en 19q12, RP11-38C1 en 19q13.11, RP11-1101H19 en 19q13.11 et RP11-26D24 en 19p12 sont présentes en 3 copies (Figure 92A) et la sonde WI2-1841M21 en 19q12 en une seule copie (Figure 92B).

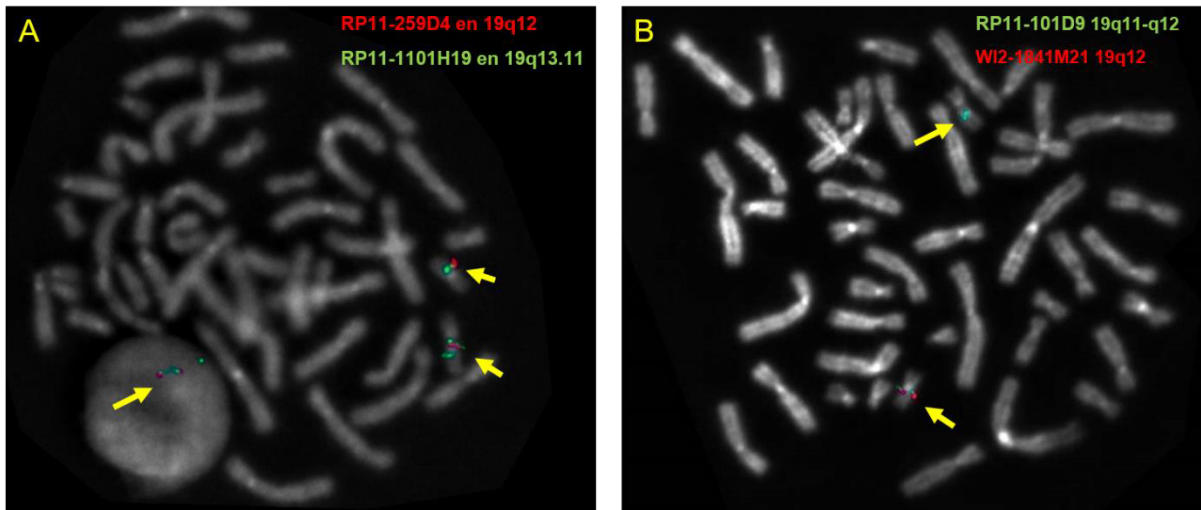


Figure 92 : Images d'hybridation *in situ* en fluorescence chez la patiente 7 A : les sondes RP11-259D4 en 19q12 (rouge) et RP11-1101H19 en 19q13.11 (vert) révèlent un gain de copie. B : les sondes RP101-D9 en 19q11q12 (vert) et WI2-1841M21 en 19q12 (rouge) révèlent une perte de copie.

La CGH-array a révélé quatre gains et une perte au niveau du chromosome 19. Ces déséquilibres englobant plusieurs gènes permettent d'expliquer le phénotype de la patiente (Figure 93).

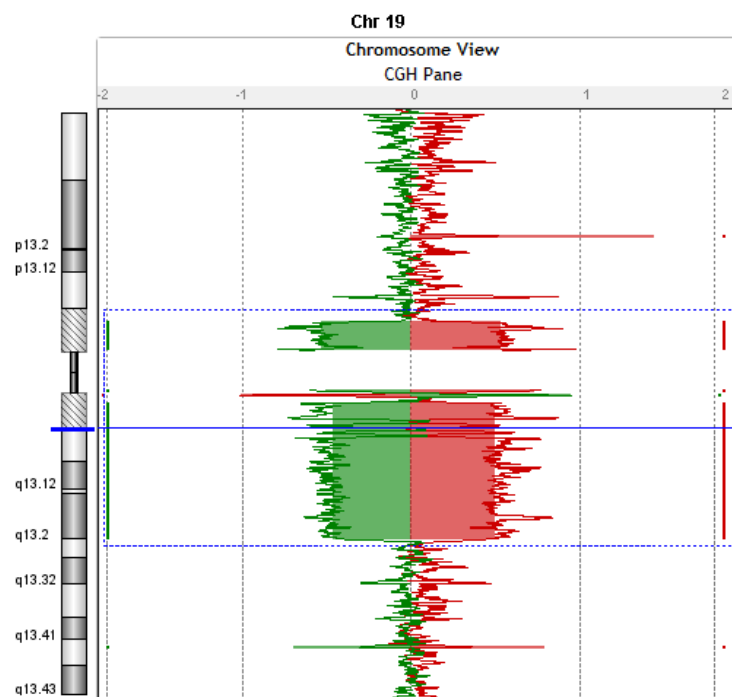


Figure 93 : Profil d'ACPA montrant 4 gains et une perte au niveau du chromosome 19 (puce Agilent).

Le WGS a détecté 18 points de cassure dont 8 gains et 4 pertes. La présence de ces déséquilibres ainsi que l'analyse des split-reads (tableau VII) suggère un mécanisme de chromoanasythesis à l'origine du réarrangement complexe chez la patiente.

Les points de cassure interrompent 3 gènes : *TSHZ3*, *PSG3* et *PSG6*. Jusqu'à présent ces gènes ne sont pas rapportés en pathologie. Les fonctions de *TSHZ3* sont inconnues mais l'expression de ce gène est très élevée dans le néocortex humain en développement. Les souris mutées hétérozygotes pour ce gène montrent des changements fonctionnels niveau des synapses établies par les projections corticales cérébrales (170). Les souris mutées à l'état homozygote présentent une absence de différenciation neuronale et musculaire (171).

Tableau VII : Liste des points de cassure au niveau du chromosome 19 chez la patiente 5. La séquence en Noire désigne la séquence du partenaire A, en rouge celle de B et en vert une séquence insérée.

Patient	Génome	chr (read 1)	position	chr (read 2)	position	Séquence selon les Split read	Mécanisme
5	hg19	19	28069001	19	28492749	/GGAAGCCCATAAAGGCATATGGCAAGGA ACCGAATATCCCCAGATAAAAGCT	MMBIR/FoSTeS
5	hg19	19	28492990	19	28069005	GTGAACAAGCACCCACTGGTAACGATTGC ATCCAGAGTTGTTTCAAAAATTTCCAAGAG GAA/GCCCATAAAGGCATATGGCAAGGAA CCGAATATCCCAG	MMBIR/FoSTeS
5	hg19	19	28733386			Coordonnee imprecise	
5	hg19	19	28930022	19	29496147	GAAATGTACAAAATAAGCCGTAATCCAA CTGTCCAGCACCTGTAATC/GCAGGTGCTG AAGCCTTGCTGGAACTCAGTAACTCTCA GGGATCATTACAA	
5	hg19	19	23777729			Position imprécise	
5	hg19	19	23784095			TTTTGTATATAAATGCAGCAAATGAACAT GGCAGTGCTTGCTGTGTAGCAGATGCTTC ATGATAAGCCATAAGTATTCCTCCTGGAGT TAGTTTGTGTA/GAC	
5	hg19	19	29496147	19	28930022	AGATACTGAGTTCCAGCAAGGCTTCAGC ACCTGCGATTACAGGTGCT/GGACAGTAG GATTACGGCATAGTTATGTACAATTCAGG AAGGCAGGAGAAGCG	
5	hg19	19	31166089	19	28733386	ATCTATAGGACAGTGCTTATTGATCTTGGA ATCCGTTTGCCGCACCTCTAAGGAACTGG ATATGCCATT/GTTGTCTCATCTGTCCA TGTGGGTGATTG	
5	hg19	19	23932111	19		Position imprecise	

5	hg19	19	23932480			Position imprécise	
5	hg19	19	23982262			Position imprécise	
5	hg19	19	23982632			Position imprécise	
5	hg19	19	24049961	19		AAAAGAAAGGCTCAACGCTGTGAGTTGAA TCTAC/ATGTATGGGCTTATGGACACAGG AGCTGATGTGTCAGTAATATCCAGTAAGG ATTGACCCCCATCCT	
5	hg19	19	24404706	19	27830055	AAATAAAGGCTTCACTCTG/ATACTACACT TCCATCTAGTTTTATGTGAAGATAAACAGT GTGTTTAAACTGTCTTTGTAGTATCTGC AAGTGGATATT	
5	hg19	19	31684717			Point de cassure douteux	
5	hg19	19	31924693	19		CACAAGCCTGGGTTTCAGGGCTTAGTTTT TCAGGATGGGAACC/TGAAAGTTCTAGAC TTacGTCTACTCAGGAGGCTaAGGCAGAA GAATCACTGAACCC	
5	hg19	19	23562558	19		Douteux	
5	hg19	19	23562678	19		Douteux	
5	hg19	19	43449338	19		AACACAGCTCTCATAAAAAAGTTAGTACTT CTGCGGTACATGTCATTATGGCTGATAACT TAATTTATTTCTAGGAATCT/CATACAAATT CTGGAACCATC	NHEJ
5	hg19	19	43781616	19		Position imprecise	
5	hg19	19	32358312	19		Position imprecise	
5	hg19	19	32669892	19		CGACAGTTCCAAATGCAAATTTTCATCAAGT CTCTGCCCTCCCAGGACCCCTACGTGGCTC CCCAGGGTCCCACAGAAC/TTTACAAAGG AGTCAGCTATAGG	NHEJ
5	hg19	19	43425622	19		Position imprecise	
5	hg19	19	21414611	19		Position imprecise	
5	hg19	19	32990693	19		Position imprecise	

5	hg19	19	23934643	19	23784095	CATTCTTGCCTCAGTCTCCTGAATAGCT GGGACTACAGGCGCCTGCCACCA/GGCAG CCAGTTAGTCTACAACTAACTCCAGGAG GAATACTTATGGCT
5	hg19	19	27830055	19		Position imprecise

Tableau III : Résultats du NGS et de la caractérisation des points de cassure chez les cinq patients présentant un déséquilibre au niveau du caryotype

Patient présentant un déséquilibré au niveau du caryotype							
Patient	Caryotype	ACPA	WGS qualité du Run	WGS: Nombre de point de cassure	Gène interrompu	Éléments répétés présents au niveau du point de cassure	Mécanisme (dédit)
			Profondeur moyenne				
1	46,XY,add(12)(q23q24)	arr[hg19] 12q23.1q24.13(101500368_113290066)x3	12,7X	1	<i>ANO4</i>	-	?
					<i>RPH3A</i>	hAT-transposon	
2	46,XX,der(18)t(16;18)(q21;p11)mat	arr [hg19] 16q22.2q24.3(70607066_90155062)x3,18p11.32p11.22(136226_9091008)x1	10,7X	1	-	hAT-transposon-Alu	?
3	47,XY,del(4)(q?)	arr [hg19] (X)x1,(Y)x2,4q21.1q22.3 (77309280_97798507)x1	30X	2	<i>CCDC158</i>	LINE	MMBIR/FoSTeS
					-	-	
4 (foetus 1)	rob(13;15) avec anomalie de structure au niveau du dérivé du chromosome 13	Nombre de CNVs: 7 gains et 4 pertes	12,93X	57	<i>STARD13</i>		Chromoanasythesis
					<i>NBEA</i>		
					<i>SMAD9</i>		
					<i>FAM155A</i>		
					<i>FGF14</i>		
					<i>RBM26</i>		
					<i>GPC5</i>		
<i>GPC6</i>							
5	chromosome 19 dicentrique	Nombre de CNVs: 7 gains et 4 pertes	7,53X	18	<i>TSHZ3</i>		Chromoanasythesis
					<i>PSG6</i>		
					<i>PSG3</i>		

En résumé, nous avons étudié cinq remaniements chromosomiques déséquilibrés. Dans tous les cas, nous avons précisé la taille du déséquilibre et confirmé son caractère pathogène. Pour trois d'entre eux, un mécanisme de type réplcatif était à l'origine du remaniement. Deux remaniements se sont avérés être des chromoanasythesis.

4.2. Caractérisation moléculaire de remaniements chromosomiques apparemment équilibrés chez des patients atteints de déficience intellectuelle et/ou de malformations congénitales

Le tableau VIII résume les quatre patients présentant un RCAE au niveau du caryotype.

4.2.1. Patient 6

Ce patient, âgé de 5 ans, est suivi dans le cadre d'une déficience intellectuelle et d'un retard de langage associés à une dysmorphie faciale avec un visage triangulaire, une macroglossie, un micropénis et une ectopie testiculaire (Figure 94). Il a également présenté des crises d'épilepsie, accompagnée par une faiblesse musculaire. Cette dernière a disparu avec le contrôle des crises d'épilepsie.



Figure 94 : Image du patient 6.

Le caryotype constitutionnel a montré une inversion péricentrique apparemment équilibrée d'un chromosome 8: 46,XY,inv(8)(p2?3;q23) dans toutes les métaphases analysées (Figure 95). Les caryotypes parentaux étaient normaux.

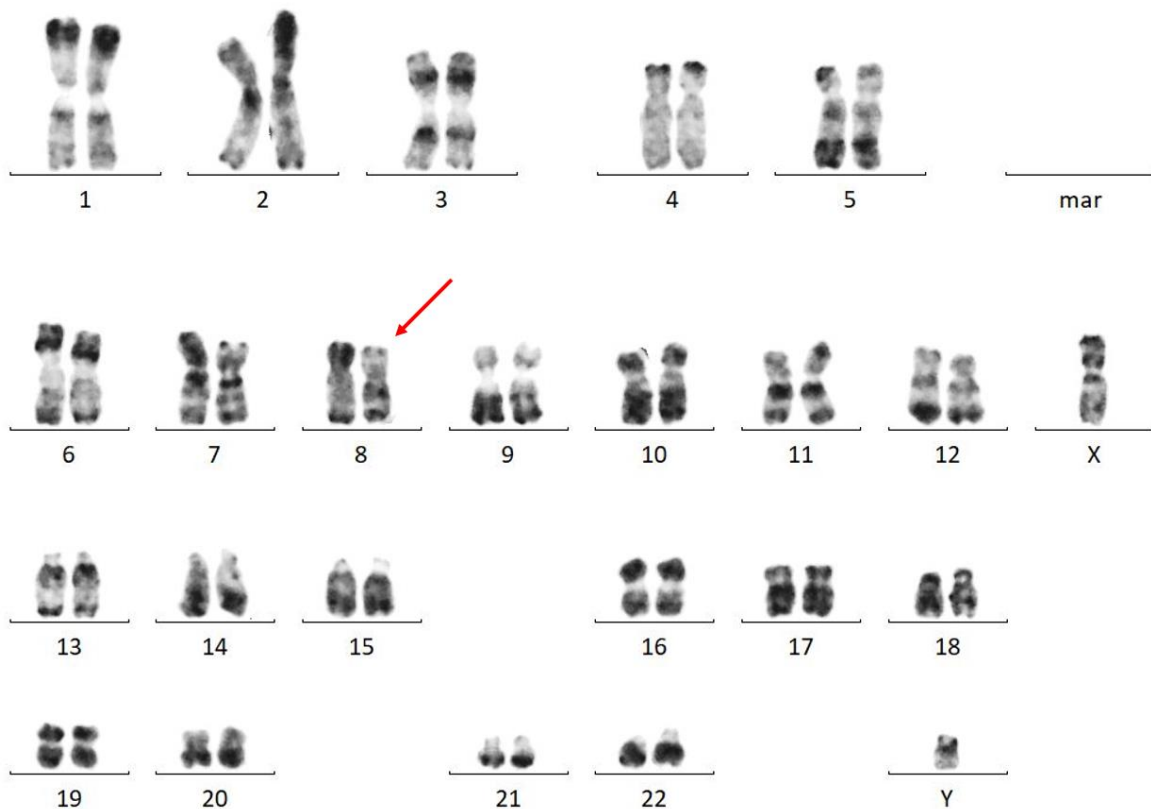


Figure 95 : Caryogramme (en bande R) du patient 6 montrant une inversion péricentrique au niveau d'un chromosome 8 : 46,XY,inv(8)(p2?3;q23) .

L'ACPA a montré une délétion de 7,6 Mb au niveau du bras court du chromosome 8, en chr8:31490519_39144107 (Hg19) (Figure 96) :

46,XY,inv(8)(p2?3;q23).arr[GRCh37] 8p12p11.22(31490519_39144107)x1.

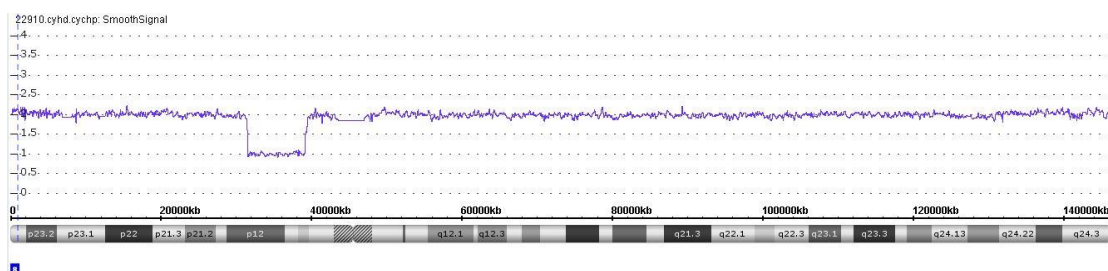


Figure 96 : Profil d'ACPA montrant la perte d'une copie de la région 8p11.22p12 (puce Affymetrix).

Cette délétion comporte 36 gènes (Tableau IV) et est rapportée pathogène d'après les différentes bases de données.

Tableau IV : Liste des gènes impliqués dans la délétion 8p11.22p12 chez le patient 6.

Chromosome	Région	Taille (kbp)	Gènes
8p11.22p12	31490519 - 39144107	7653.588	NRG1, NRG1-IT3, FUT10, MAK16, TTI2, RNF122, DUSP26, UNC5D, KCNU1, ZNF703, ERLIN2, LOC728024, PROSC, GPR124, BRF2, RAB11FIP1, GOT1L1, ADRB3, EIF4EBP1, ASH2L, STAR, LSM1, BAG4, DDHD2, PPAPDC1B, WHSC1L1, LETM2, FGFR1, C8orf86, RNF5P1, TACC1, PLEKHA2, HTRA4, TM2D2, ADAM9, ADAM32

Une revue de la littérature montre que le gène *FGFR1* (# MIM 136350), autosomique dominant, explique la majorité du phénotype chez le patient.

Une technique FISH a été réalisée en utilisant deux sondes, une sonde pour la région télomérique du bras court du chromosome 8 marquée en rouge et une autre sonde localisée dans la région délétée, RP11-100B16 (8p11.23p11.22 ; chr8:38239682_38403260), marquée en vert (Figure 97). Cette technique a confirmé que la délétion est localisée sur le même chromosome 8 dérivé porteur de l'inversion.

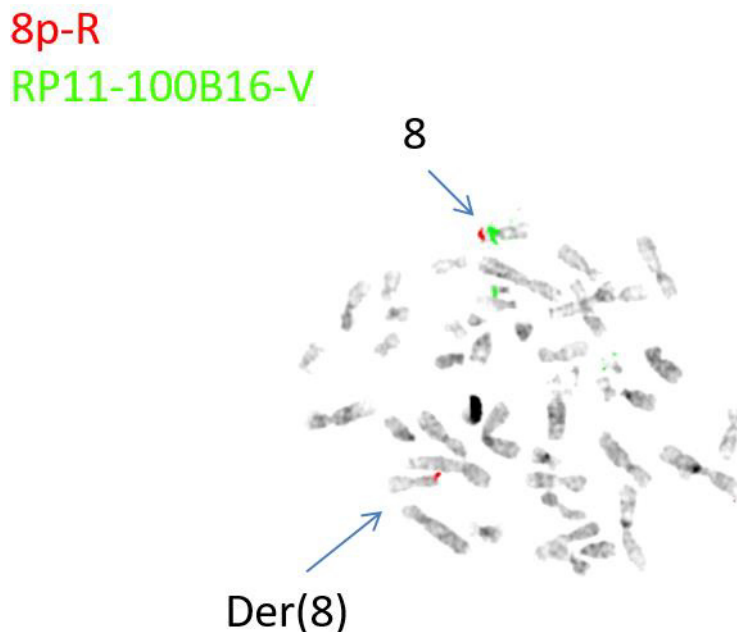


Figure 97 : Hybridation *in situ* en fluorescence chez le patient 6 pour les sondes: 8pter (Cytocell) en rouge et RP11-100B16 en vert. La délétion de la région 8p11.22p12 est localisée sur le chromosome 8 porteur de l'inversion péricentrique.

L'analyse des données de WGS a permis de détecter une diminution de la profondeur des reads au niveau de la région chr8:31488322_39144893 (hg19) du chromosome 8, confirmant la délétion détectée par ACPA. De plus, deux points de cassures d'inversion ont été mis en évidence au niveau du chromosome 8 : le premier point de cassure en chr8:9435071_9435072 en 8p23.1 et le deuxième en chr8:97861839_97861840 en 8q22.1 (Figure 98). Ces deux points de cassures confirment l'inversion péricentrique détectée par le caryotype.

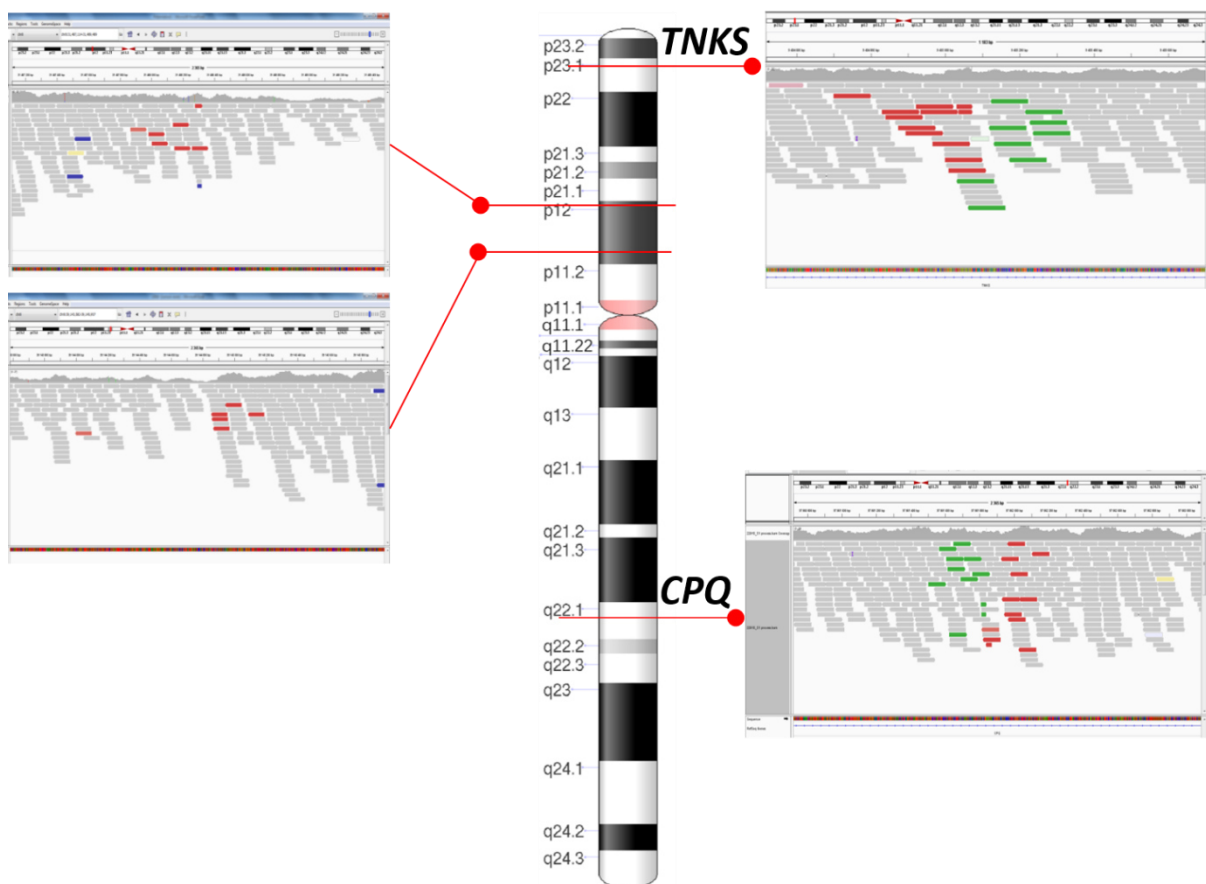


Figure 98 : Image illustrant les points de cassures détectés sur le chromosome 8 par WGS chez le patient 6 : à gauche points de cassure de la délétion ; à droite : points de cassure de l'inversion, interrompant les gènes *TNKS* et *CPQ*.

Les deux points de cassures de l'inversion interrompent deux gènes : le gène *TNKS* (# MIM 603303) au niveau de l'intron 1, et le gène *CPQ* (# MIM 618754) au niveau de l'intron 3. La formation d'un transcrit de fusion a été exclue par des études in silico. Aucun de ces gènes n'a été rapporté en pathologie constitutionnelle à ce jour.

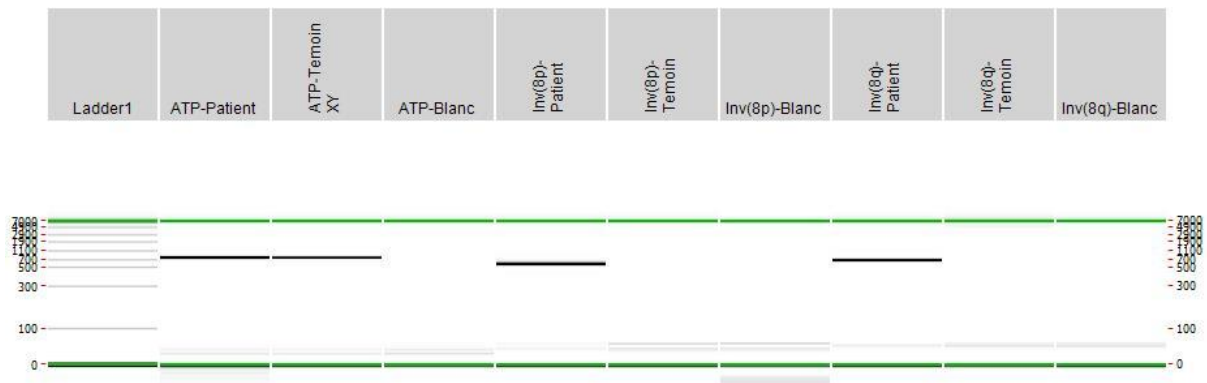


Figure 99 : Amplification des fragments de jonction des points de cassures chez le patient 6 par PCR (visualisation Labchip) pour les points de cassures de l'inversion et une séquence de contrôle (ATP1A3 exons 7-8).

Les deux fragments de jonction ont été amplifiés par PCR puis séquencés (Figure 99), ce qui a permis de préciser les positions exactes des points de cassure sur le chromosome 8: chr8:9435071_9435072 (hg19), et chr8:97861838_97861841 (hg19) (Figure 100). Il existe une microhomologie de deux bases au niveau de chaque jonction (AC et GT respectivement). L'examen attentif de l'orientation des reads au sein des paires aberrantes a mis en évidence une orientation inattendue des reads. En effet, dans le cas d'une inversion, on s'attend théoriquement à voir les deux reads d'une même paire orientés sur le même brin du génome de référence. Or pour le patient 6, les reads des paires aberrantes de l'inversion étaient alignées l'un sur le brin sens, l'autre sur le brin anti-sens. En fait, le point de cassure proximal de l'inversion est survenu au niveau d'une région qui est le siège d'une inversion polymorphe, rapportée dans une séquence alternative du génome hg19 : chr8_KZ208915v1_fix:1-6367528 : chr8_KZ208915v1_fix:3773905_3773906.

inv(8p) chr8:9435071(/73)::97861838(/40)

Chr8- TGGTTGTTTAAAAGTGTGTAGCACCTCCTCCTGCTCTCCTCTGCCTCCT
der (8p) TGGTTGTTTAAAAGTGTGTAGCACTTCCCAACATATGGAGCAGCACTGG
Chr8- ACTCTGACTTCCTTTCTGTCATACTTCCCAACATATGGAGCAGCACTGG

inv(8q) chr8:9435072(/70)::97861841(/39)

Chr8+ AGGCAGAGGAGAGCAGGAGGAGGTGCTACACACTTTTAAACAACCAGAT
der (8) AGGCAGAGGAGAGCAGGAGGAGGTATGACAGAAAGGAAGTCAGAGTGGAA
Chr8+ GTGCTGCTCCATATGTTGGGAAGTATGACAGAAAGGAAGTCAGAGTGGAA

Figure 100 : Séquences des fragments de jonction et coordonnées des points de cassures chez le patient 6. En rouge: séquence de référence sur le bras court du chromosome 8 ; en bleu : séquence de référence sur le bras long du chromosome 8. Le séquençage Sanger a montré une homologie de deux bases au niveau des points de cassures.

Le point de cassure sur le bras court du chromosome 8, qui interrompt le gène *TNKS*, est situé dans une région LTR (Long terminal repeat). Le point de cassure sur le bras long du chromosome 8, qui interrompt le gène *CPQ* est situé dans un transposon, de type Arthur1, de la famille hAT-Tip100. Ces deux répétitions font partie de la famille des rétrotransposons.

del(8): (split read) chr8:31488322::39144893

Chr8+ TTTCCTGGTCTAACATTTGATGTATTTCTCTATGGTTCAAAGCATAACC
der (8) TTTCCTGGTCTAACATTTGATGTACTTTGATTAATTTGCATATATTTGT
Chr8+ TGACTCATTGGTTGTTTCAGGAGCACTTTGATTAATTTGCATATATTTGT

Figure 101 : Séquence du fragment de jonction de la délétion d'après split-read au niveau du bras court du chromosome 8 chez le patient 6.

La séquence de jonction, et donc la taille exacte, de la délétion a été définie par étude des split-reads (Figure 101). Les points de cassures de la délétion sont situés au niveau de séquences répétées. Le point de cassure proximal (chr8:31488322) est localisé dans un SINE (Short interspersed nuclear elements). Le point de cassure distal (chr8:39144893) est localisé dans un LINE (Long interspersed nuclear elements). Le mécanisme le plus probable de ces deux remaniements est un NHEJ.

4.2.2. *Patient 7*

Le patient est âgé de 18 mois et a été adressé dans le cadre d'une hypotonie et d'une faiblesse musculaire associées à une croissance excessive, des crises d'épilepsie et des troubles visuels (Figure 102). L'imagerie cérébrale a mis en évidence une dilatation des ventricules latéraux avec une colpocephalie marquée, une hypodensité diffuse avec perte de volume de la substance blanche périventriculaire et une hypogénèse du corps calleux. L'échographie abdominale a montré une hernie inguinale bilatérale.



Figure 102: Photo du patient 7.

Le caryotype a révélé une inversion péricentrique apparemment équilibrée au niveau d'un chromosome 11 : 46,XY,inv(11)(p13;q12) dans toutes les métaphases analysées (Figure 103). Les caryotypes parentaux étaient normaux.

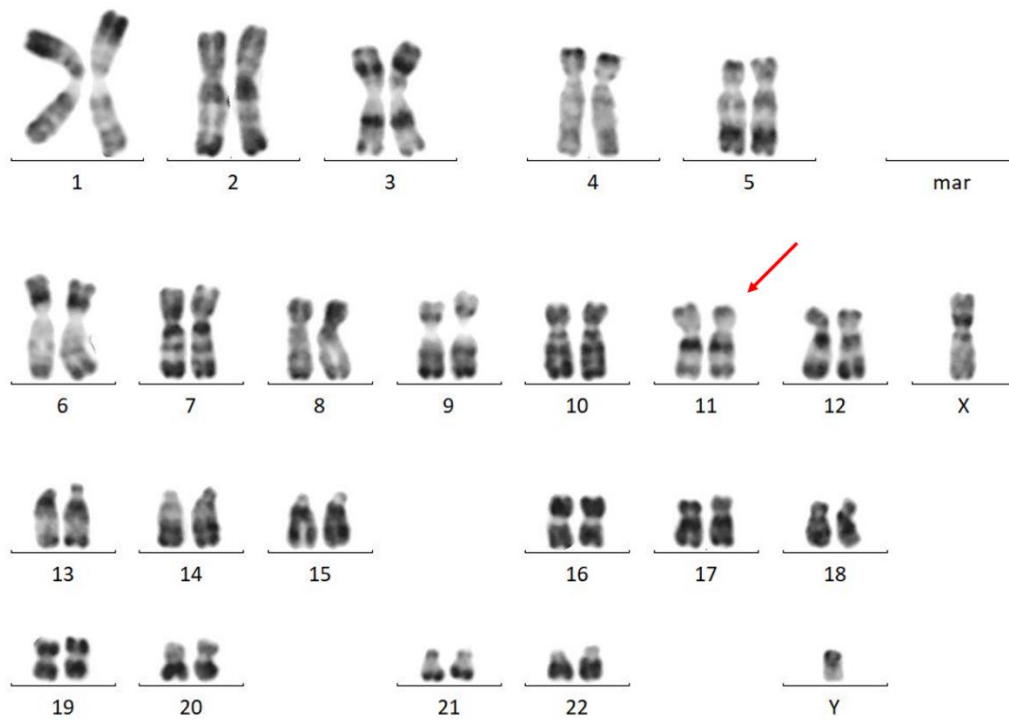


Figure 103 : Caryogramme (en bande R) du patient 7 présentant une inversion péricentrique au niveau d'un chromosome 11.

L'ACPA n'a détecté aucun CNV au niveau de la région présumée des points de cassure de l'inversion, ni de CNV pathogène sur le reste du génome.

Le WGS a permis de caractériser les deux points de cassure de l'inversion sur le chromosome 11 (Figure 104).

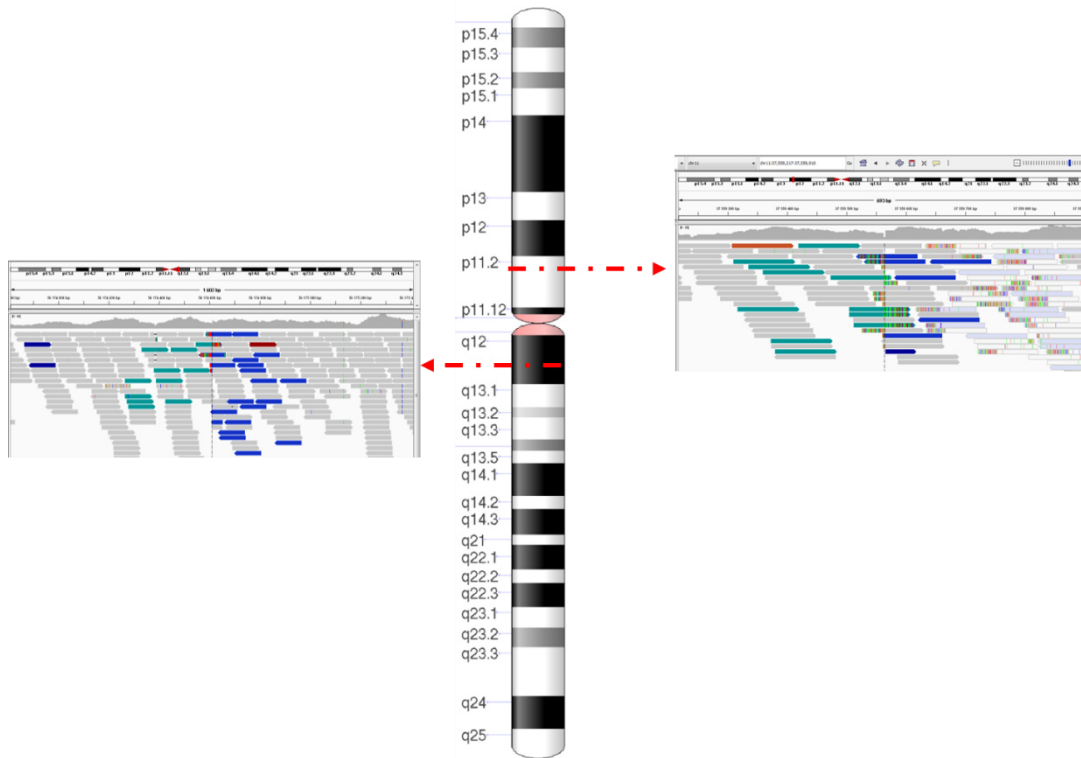


Figure 104 : Image illustrant les points de cassures de l'inversion sur le chromosome 11 détectés par WGS chez le patient 7.

L'amplification par PCR (Figure 105a) suivie du séquençage Sanger des fragments de jonction a permis de préciser les positions des points de cassure en chr11:37581115_37581131 et en chr11:55942089_55942089 (hg19) (Figure 105b). On note, au niveau de la jonction proximale, que 5 bases GAAAA (représentées en vert) ont été insérées et que 6 bases CAGTGG (représentées en gras) localisées en chr11:37581113_37581117 ont été remplacées par 4 bases AAAA (représentées en vert).



Figure 105 : a- Amplification des fragments de jonction des points de cassures chez le patient 3 par PCR (visualisation Labchip) pour le der (11p), le der (11q) et une séquence de contrôle (ATP1A3 exons 7-8). **b-** Séquences des fragments de jonction et coordonnées des points de cassure. En bleu séquence de référence sur le chromosome 11p, en rouge: séquence de référence sur le chromosome 11q, en vert, insertions nucléotidiques.

Le point de cassure sur le bras court du chromosome 11 est situé dans une région SINE. Des microhomologies de part et d'autre du point de cassure suggèrent un mécanisme basé sur la réplication de type FoSTeS/MM-BIR expliquant ce réarrangement chromosomique.

Les points de cassures de l'inversion n'interrompent aucun gène. L'exploration de la région de part et d'autre des points de cassure n'identifie pas de gène qui pourrait expliquer le phénotype du patient par un effet de position.

La profondeur moyenne du séquençage chez le patient 6 est de 13,4X, avec un score Q30>91,45%. Le score Q30 indique le taux d'erreur du séquençage nucléotidique qui correspond à une erreur pour 1000 bases séquencées. Cette profondeur de 13,4X est largement

suffisante pour détecter les variants structuraux mais n'est pas assez fiable pour détecter les variants nucléotidiques. Avec une profondeur de 20X, 5 à 15% des variants nucléotidiques hétérozygotes et 1 à 4% des variants homozygotes sont manqués (172). L'analyse des SNVs a toutefois été réalisée pour ce patient à partir des données de WGS, sur un panel bio-informatique de gènes impliqués dans la déficience intellectuelle. Cette stratégie nous a permis d'identifier un variant au niveau du gène *NSD1* (#MIM 606681): c.2954_2955delCT (NM_0224554) (Figure 106). Cette délétion provoque un décalage du cadre de lecture, l'acide aminé serine est substitué par l'acide aminé cystéine et crée un codon stop prématuré, 25 positions en aval, noté p.Ser985CysfsX25. Ce variant a été confirmé par séquençage Sanger (Figure 107).

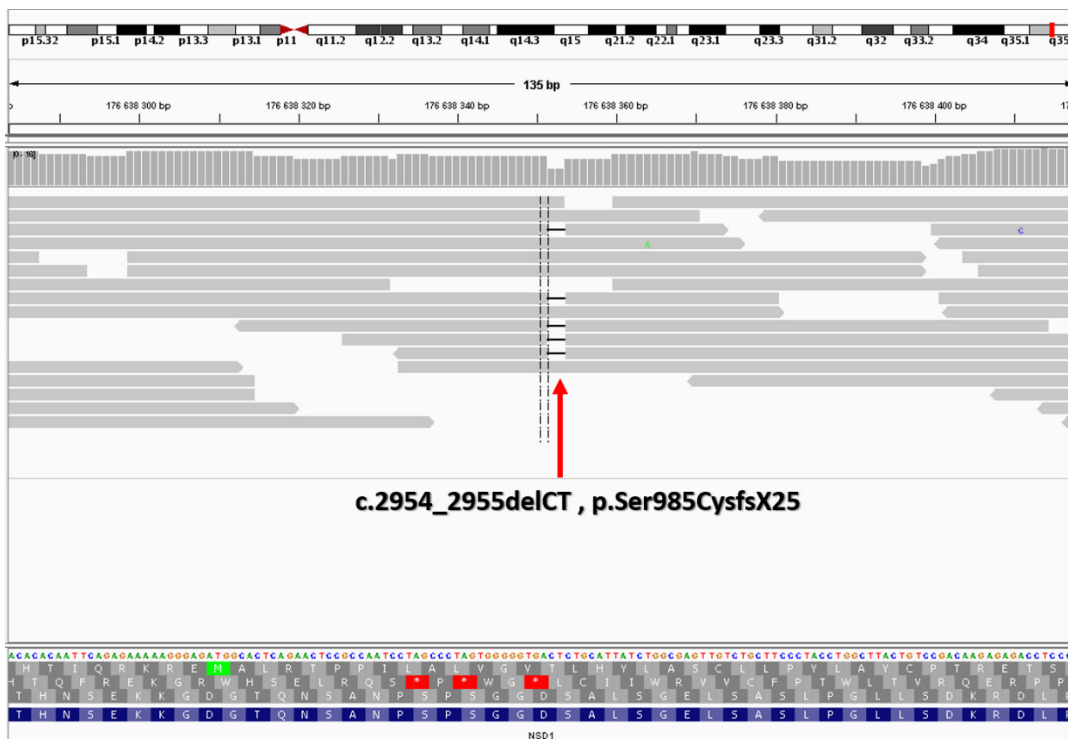


Figure 106 : Visualisation par IGV de la délétion c.2954_2955delCT au niveau du gène *NSD1*.

Cette variation est prédite pathogène d'après plusieurs programmes de prédiction et elle a été rapportée plusieurs fois comme pathogène dans la base de données ClinVar et dans la littérature (173). Le gène *NSD1*, autosomique dominant, est impliqué dans le syndrome de SOTOS.

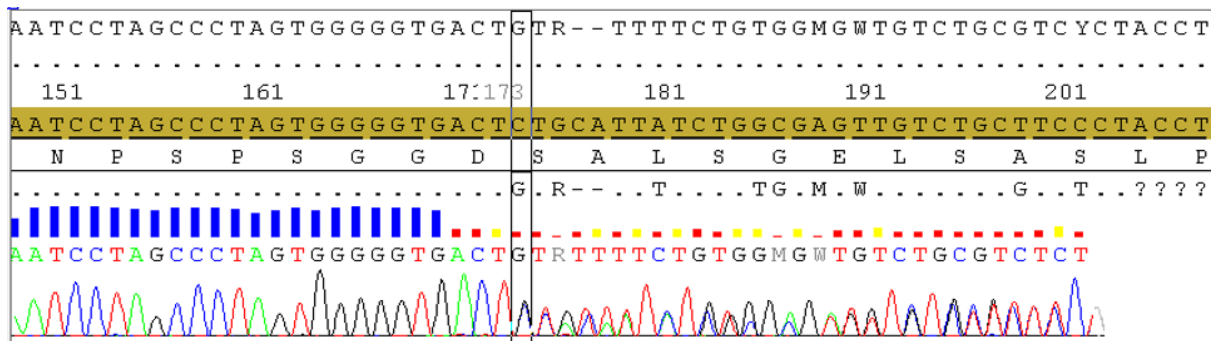


Figure 107 : Chromatogramme du séquençage de l'exon 5 du gène *NSD1*, confirmant la délétion de deux bases CT par la méthode Sanger.

Cette variation explique le phénotype chez le patient et élimine l'implication de l'inversion péricentrique du chromosome 11 dans le phénotype du patient. Cette variation doit être recherchée chez les parents pour préciser le caractère hérité ou *de novo*.

4.2.3. *Patient 8*

Le patient 8 est né avec un poids de 2600g et une taille de 49 cm et a présenté une hémorragie digestive en période néonatale. Deux ans plus tard, il a été adressé pour l'évaluation d'une déficience intellectuelle, associée à une dysmorphie faciale (visage triangulaire, petit front, oreilles larges et décollées) et une implantation particulière du pouce droit.

Le caryotype a révélé une translocation réciproque apparemment équilibrée entre les bras longs d'un chromosome 1 et les bras courts d'un chromosome 2 : 46,XY,t(1;2)(q21;p13) dans toutes les métaphases analysées (Figure 108). Les caryotypes parentaux étaient normaux.

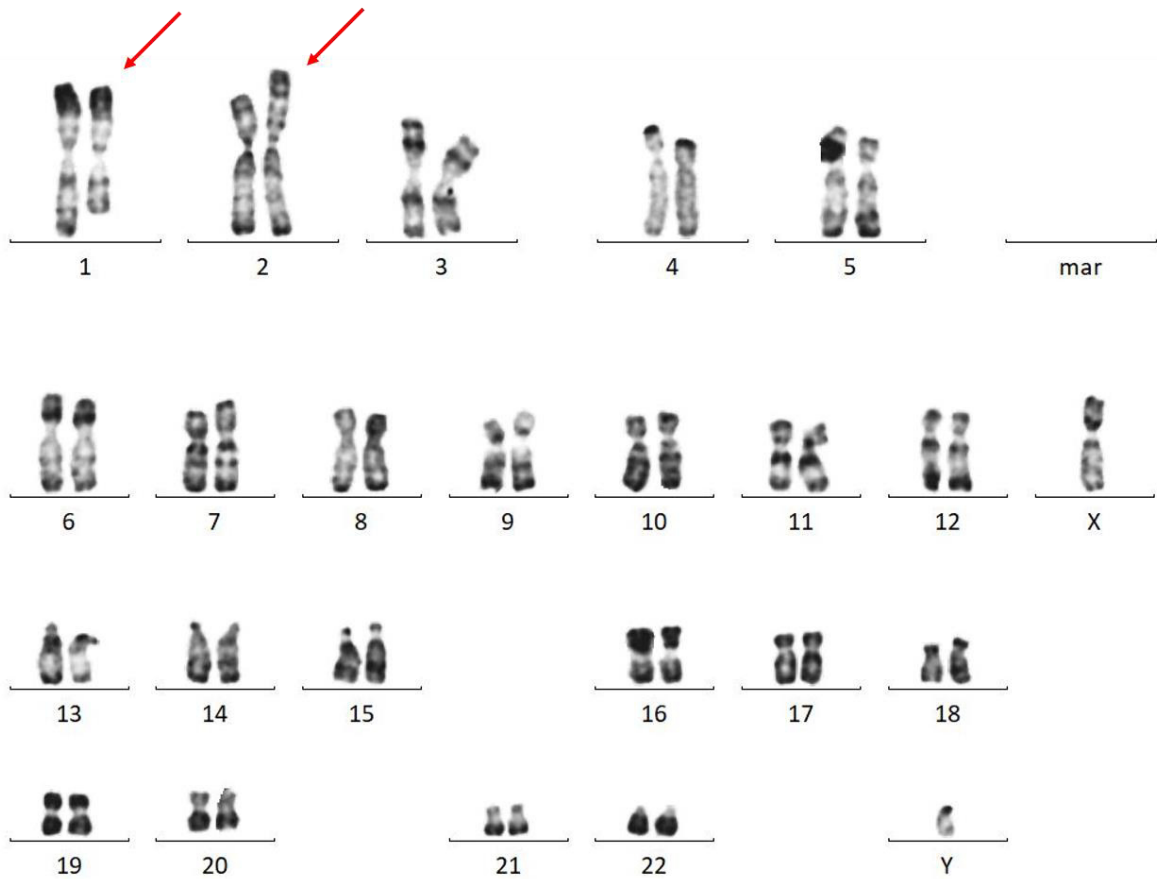


Figure 108 : Caryogramme (en bande R) du patient 8 présentant une translocation réciproque entre les bras longs d'un chromosome 1 et les bras courts d'un chromosome 2 : 46,XY,t(1;2)(q21;p13).

L'ACPA a détecté deux pertes au niveau du bras court du chromosome 2 : une perte de 7 Mb localisée en chr2:30908598_37914388 en 2p22.2p23.1 et une perte de 1,1 Mb située en chr2:50598357_51707702 Mb en 2p16.3 (Figure 109). Aucun CNV pathogène n'a été détecté sur le chromosome 1 ni sur d'autres chromosomes.

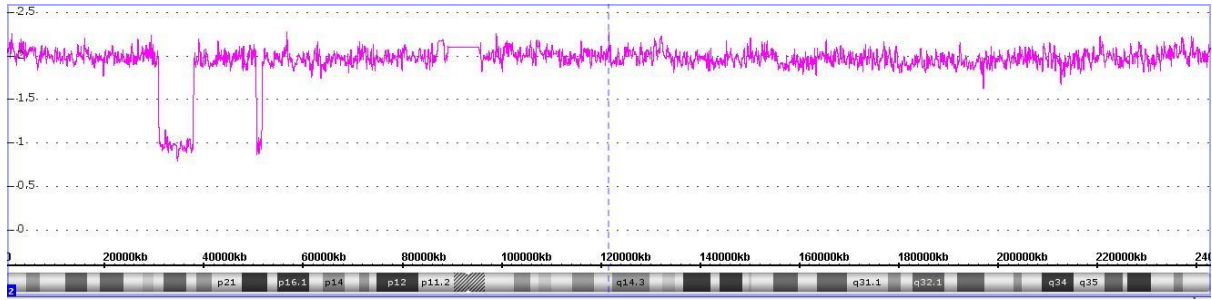


Figure 109 : Profil d'ACPA montrant les deux délétions détectées au niveau du bras court du chromosome 2 en 2p16.3 et 2p22.2p23.2 chez le patient 8 (puce Affymetrix).

La délétion de 7 Mb en 2p22.2p23.1 chr2:30908598_37914388 (Hg19) explique le phénotype du patient. Cette délétion englobe 27 gènes OMIM dont trois dominants : *SPAST*, *NLRC4* et *EIF2AK2*. Des délétions plus petites que celle-ci sont citées pathogènes dans les bases de données (Decipher, ClinVar).

Un WGS a été effectué chez ce patient afin de comprendre le mécanisme de ce réarrangement chromosomique. Cette technique a détecté cinq points de cassures dont trois délétions.

L'amplification des 5 fragments de jonction (figure 110) et le séquençage des fragments de jonction par méthode Sanger a permis de préciser les positions des points de cassure représentés dans la figure 111.

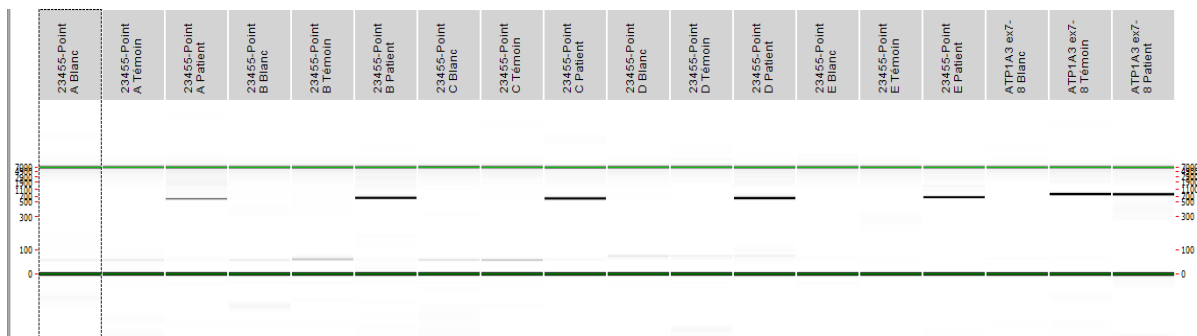


Figure 110 : Amplification des 5 fragments de jonction des points de cassures chez le patient 8 par PCR (visualisation Labchip) pour le der(1) et le der(2) et une séquence de contrôle (ATP1A3 exons 7-8).

der (1) chr1:178764844::178926138

Chr1+ TACAGGTGTAGCCACCACACCTGGACTGTATCTTATTAAAATATTTCTG
der (1) TACAGGTGTAGCCACCACACCTGGCATAACAGGAGAATTAATACTCAG
Chr1- ACTTTCAGATCCATGACCTCTCTTCATAACAGGAGAATTAATACTCAG

der (1) chr1:178764842::chr2:45380782

Chr1- AATATTTTAATAAGATACAGTCCAGGTGTGGTGGCTACACCTGTAATCC
der (1) AATATTTTAATAAGATACAGTCCACATTGTGCTCTTGATCTTCCCAACA
Chr2+ GTTCTATGCCTGTTGTTTGACATACATTGTGCTCTTGATCTTCCCAACA

der (1) chr2:50596388::30891889

Chr2+ CTATAAAAGGACATTTTAAAAAAACCAAAAACCCGAATGAATACATTT
der (1) CTATAAAAGGACATTTAAAATAAATGACAAAAATAGCATAAAGAATGGG
Chr2- TCTCTTTACAACAGGTTAAAATATATGACAAAAATAGCATAAAGAATGGG

der (2) chr1:178926135::chr2:37914466

Chr1- TCAGATCCATGACCTCTCTTCATAACAGGAGAATTAATACTCAGGTAG
der (2) TCAGATCCATGACCTCTCTTCATAATCTCCAAAAGATGGTCAGCAGGG
Chr2+ ATTATATAGACTCTAGTACCCACCATCTCCAAAAGATGGTCAGCAGGG

der (2) chr2:45359902::51707044

Chr2+ CACAGGAGAGCAGCAGCCGTGTCCTTAGCCCCATCACACCAGGGCCATG
der (2) CACAGGAGAGCAGCAGCCGTGTCCGTGATCAGTTTTCAATACTTAGTTG
Chr2+ ACTGCCTTCCTATCAAGATCTCCCGTGATCAGTTTTCAATACTTAGTTG

Figure 111 : Séquences des 5 fragments de jonction et coordonnées (hg19) des points de cassures chez le patient 8. En bleu et en rouge (haut et bas) : séquences de référence, séquences de jonction au milieu. Les microhomologies sont soulignées.

Ceci a permis de définir les points de cassure aux localisations suivantes :

- Point de cassure au niveau du chromosome chr1:178764844_178764842 (hg19) en 1q25.2,
- Point de cassure au niveau du chromosome chr1:178926138_178926135 (hg19) en 1q25.2,
- Point de cassure au niveau du chromosome chr2:30891889_37914466 (hg19) en 2p22.1 indiquant une délétion de 7Mb ,
- Point de cassure au niveau du chromosome chr2:45359902_45380782 (hg19) en 2p22.1 indiquant une délétion de 20kb,
- Point de cassure au niveau du chromosome chr2:50596388_51707044 (hg19) en 2p16.3 indiquant une délétion de 1,1 Mb.

Tous ces points de cassures sont localisés dans des éléments répétés de type SINE ou LINE, de familles différentes. Le mécanisme probable est un NHEJ. La figure 112 montre la construction du remaniement à partir des données générées par le WGS. Elle permet d'évoquer un remaniement de type chromoanagenesis.

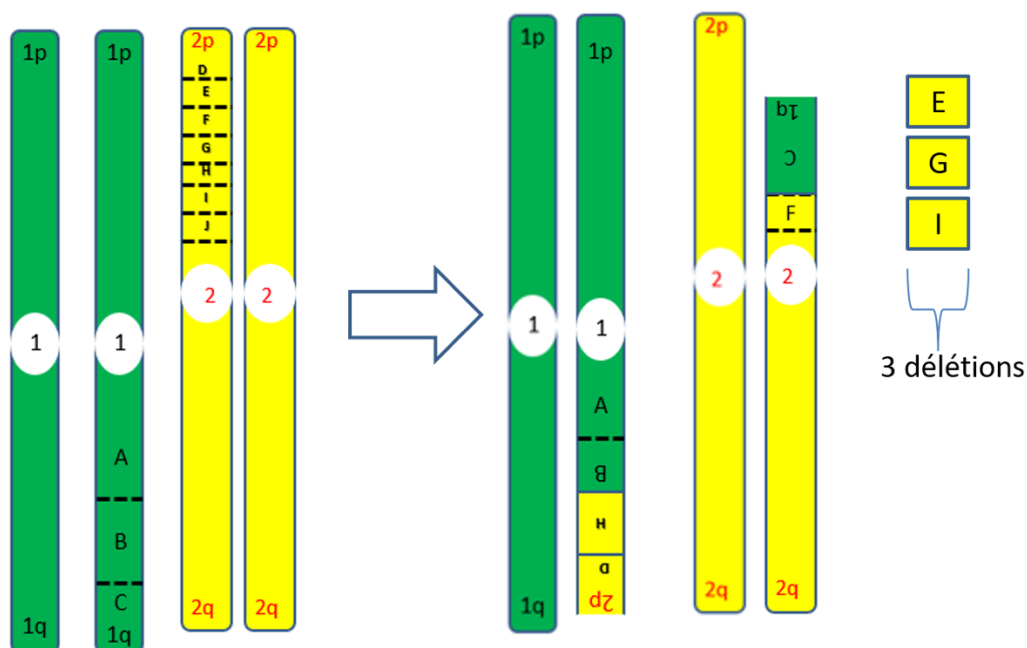


Figure 112 : Construction des chromosomes dérivés 1 et 2 d'après les données générées par le WGS chez le patient 8.

4.2.4. Patient 9 (Fœtus 2)

Le patient 9 est un fœtus interrompu à 31 SA suite à la découverte d'un remaniement chromosomique complexe lors d'une amniocentèse réalisée pour dépistage de la trisomie 21 positif (test combiné : 1/172). Il ne présentait aucun signe d'appel échographique. L'examen foetopathologique a montré un fœtus de sexe masculin, dont les données biométriques sont conformes à un âge fœtal de 33 à 34 SA au lieu de 31 SA (poids de 1940 g, longueur vortex-talon de 42,5 cm, longueur vortex-coccyx de 29 cm, périmètre cranien de 29,5 cm et longueur du pied de 7 cm). L'examen externe n'a révélé aucune malformation ni dysmorphie faciale. L'examen interne montre des remaniements congestifs au niveau du foie, de la rate et des surrénales.

Le caryotype a révélé une translocation complexe entre les chromosomes 1, 3, 6 et 13 (Figure 113). Les caryotypes parentaux sont normaux.

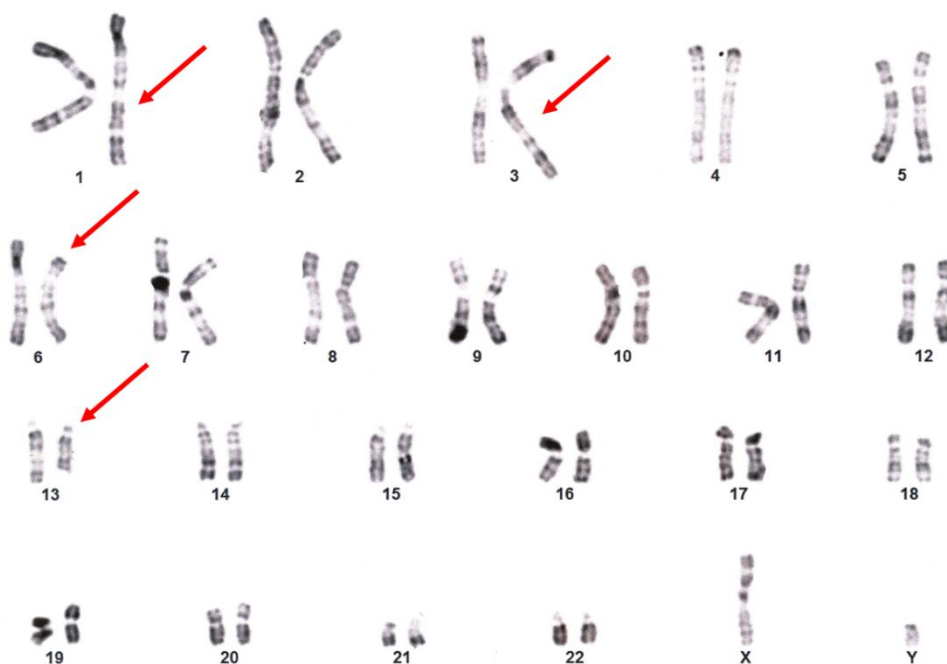


Figure 113 : Caryogramme (en bande R) du patient 9 présentant un réarrangement complexe entre les chromosomes 1, 3, 6 et 13.

Une technique de FISH avec les peintures chromosomiques des chromosomes 1, 3, 6 et 13 a confirmé les réarrangements trouvés au caryotype. La peinture du chromosome 1 a confirmé

la présence d'un fragment chromosomique du chromosome 1 sur le dérivé 3 (Figure 114A). La peinture du chromosome 3 a montré la présence d'un fragment du chromosome 3 sur le dérivé 13 (Figure 114B). La peinture du chromosome 6 a montré la présence d'un fragment du chromosome 6 sur le dérivé 3 et la peinture du chromosome 13 a montré la présence d'un fragment chromosomique du chromosome 13 sur les dérivés 1, 3 et 6 (Figure 114C).

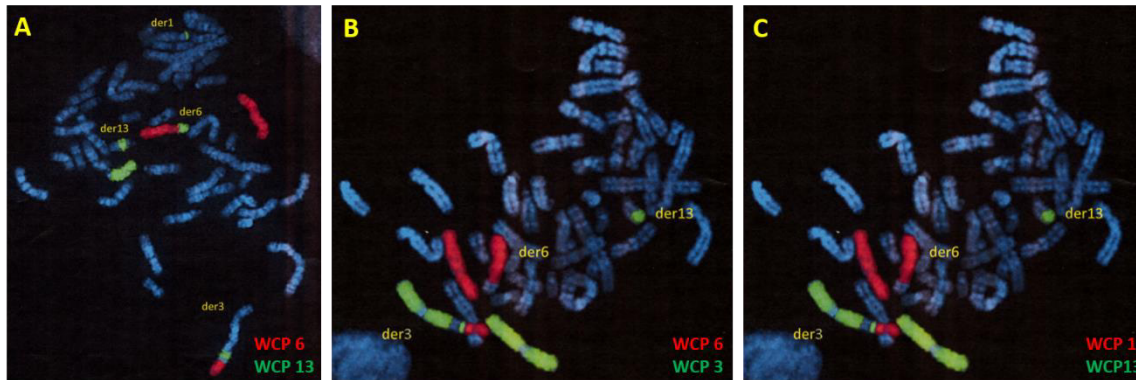


Figure 114 : Images d'hybridation *in situ* en fluorescence chez le patient 9. A : peinture chromosomique des chromosomes (WCP) 6 (en rouge) et 13 (en vert). B : WCP 6 (en rouge) et WCP 3 (en vert). C : WCP 1 (en rouge) et WCP 13 (en vert).

L'analyse du WGS a permis de détecter 9 points de cassures (hg38) (Figure 115):

- Un point de cassure sur le chromosome 1 localisé en chr1:144223272_144223272 en 1q21.1 ;
- Un point de cassure sur le chromosome 1 localisé en chr1:146718944_146741334 en 1q21.1 ;
- Un point de cassure sur le chromosome 3 localisé en chr3:159391328_159391325 en 3q25.33 ;
- Un point de cassure sur le chromosome 6 localisé en chr6:50438413_50438417 en 6p12.3 ;
- Un point de cassure sur le chromosome 13 localisé en chr13:43344835_43344834 en 13q14.11 ;
- Un point de cassure sur le chromosome 13 localisé en chr13:70702695_70702694 en 13q21.33 ;
- Un point de cassure sur le chromosome 13 localisé en chr13:79659509_79659508 en 13q31.1 ;

- Un point de cassure sur le chromosome 13 localisé en chr13:85841126_85841138 en 13q31.1.

Aucun CNV n'a été détecté au niveau du remaniement en dehors d'une délétion de 22kb sur le chromosome 1 en chr1:146718944_146741334. Les points de cassure interrompent trois gènes : *ENOX1*, *IQCJ-SCHIP1* et *HYDIN2*. Jusqu'à présent ces trois gènes ne sont pas liés à une maladie constitutionnelle. *ENOX1* (#610914) joue un rôle dans le transport d'électrons des membranes plasmiques qui sont impliqués dans diverses fonctions, comme la défense cellulaire, le contrôle de la croissance et de la survie des cellules. La mutation de ce gène à l'état hétérozygote chez la souris entraîne une augmentation de la densité minérale osseuse et une augmentation de la quantité de graisse corporelle. *IQCJ-SCHIP1* est un des rares gènes dans le génome humain, dont la transcription couvre deux phases ouvertes de lecture adjacentes indépendantes. L'ARNm de ce transcrite est très exprimé dans le cerveau (174). Les duplications et les délétions de la région 1q21.1 sont caractérisées par une micro ou macrocéphalie, des troubles neurodéveloppementaux et des anomalies congénitales de pénétrance incomplète et d'expressivité variable (175). Il a été proposé que le gène *HYDIN2*, présent dans ces CNVs, pourrait être à l'origine des anomalies de la croissance cérébrale chez les patients.

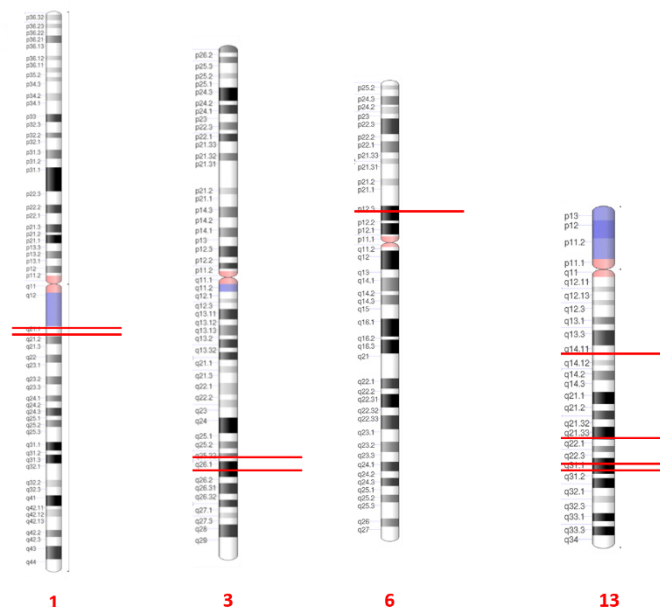


Figure 115 : Schéma illustrant les points de cassure présent chez le patient 9 sur les idéogrammes des chromosomes 1, 3, 6 et 13 impliqués dans le réarrangement chromosomique.

La séquence des jonctions a été déterminée par l'analyse des splits-reads et le remaniement a été reconstitué à partir des données de WGS (Figure 116). A noter que l'implication de l'hétérochromatine constitutive du chromosome 1, en particulier au niveau de duplications segmentaires, a nécessité de faire une correction sur l'orientation du fragment chr1:146741334_qter et a également entraîné un degré d'incertitude sur la position de certains points de cassure, puisque des séquences étaient présentes sur différentes positions sur le chromosome 1 dans une région de 2 Mb.

L'analyse des split-reads a identifié de rares microhomologie de 4 bases ou moins au niveau des jonctions, ce qui est plutôt en faveur d'un mécanisme de NHEJ et permet d'évoquer un remaniement de type chromothripsis.

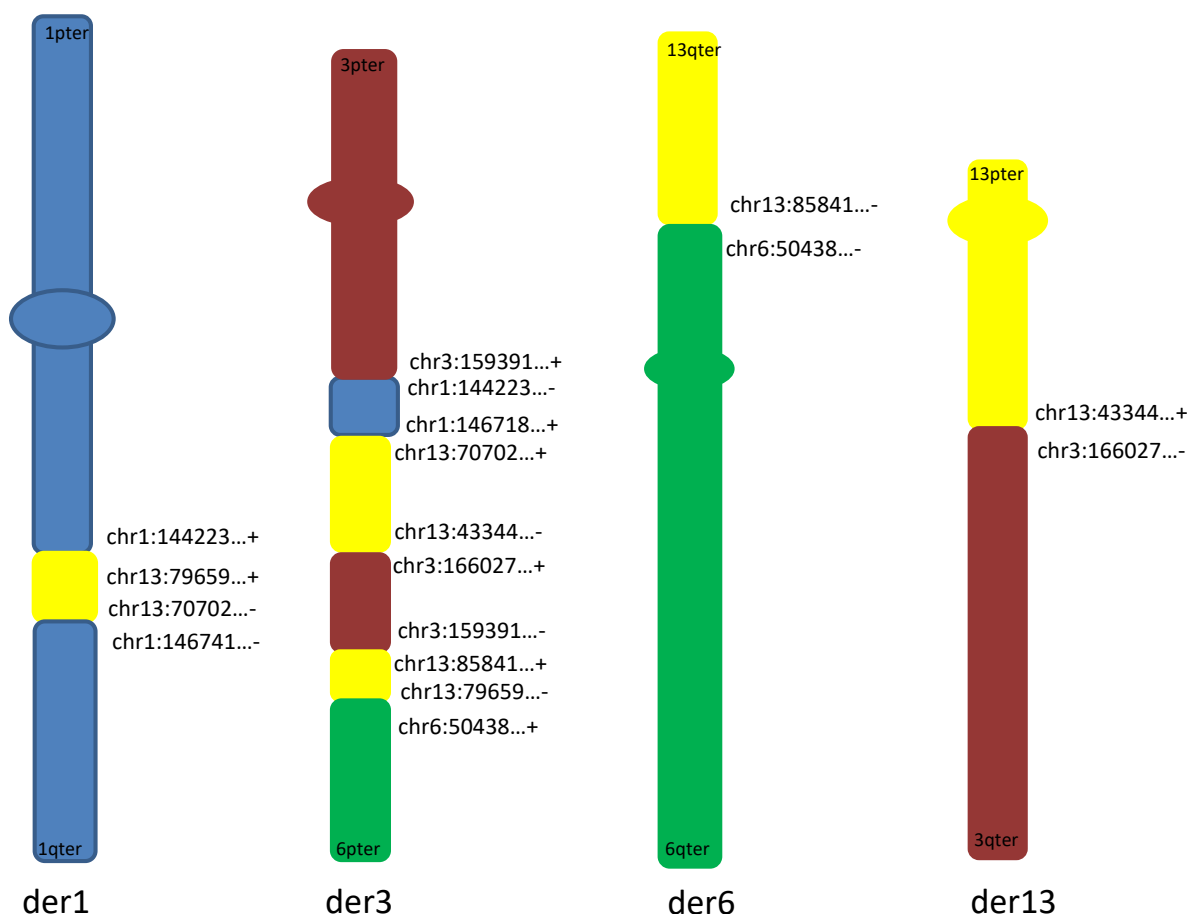


Figure 116 : Schéma du remaniement chromosomique impliquant les chromosomes 1, 3, 6 et 13 d'après la combinaison des données générées par le WGS, le caryotype et la technique de FISH chez le patient 9.

Tableau X : Liste des fragments de jonctions identifiés par l'analyse des splits-reads chez le patient 9. La séquence en Noire désigne la séquence du partenaire A et en rouge celle de B.

Patient	Genome	Chr (read 1)	Position	Chr (read 2)	Position	Séquence selon les Split read	Mécanisme
9	Hg38	6	50438413	13	79659508	TCGTAAGTCTAGAAAGTACATCCTATTACAGGATTTAAAGTAACTTTCCCATCT/ <u>GAA</u> ATGCCAATCTTAAAGGTAGCTCTTGGCTCC TGCTGGGTACAGTA	FoSTeS ou MMBIR
9	Hg38	6	50438413	13	79659504	AGTCTAGAAAGTACATCCTATTACAGGATTTAAAGTCACTTTCCCATCT/ <u>GAA</u> CTGCCAATCTTAAAGGTAGCTCTTGGCTCCTGCTG GGTACAGTATGCCT	FoSTeS ou MMBIR
9	Hg38	6	50438417	13	85841138	CTTATAGCCTCATTGTTGCTTTTACTGAAATTTCAAGAACCAACAGCAA/AGGACTTGAATATATATGTATTATATATGATTAGGTTCC ACCTACATTGAACA	NHEJ
9	Hg38	6	50438417	13	85841138	TCAATGTAGGTGAACCTAATCATATATAATACATATATATTCAAGTCCT/TTGCTGTTGGTTCTTGAAATTCAGTAAAGCAACAAT GAGGCTATAAGCAA	NHEJ
9	Hg38	13	85841130	3	59391329	AAAATAATTAATATTACCATTGACTTTAAGAAGCTTTCTTTCTTTTCAGTCT/ <u>TCTTTTCTTGATTGCAGGGACTTAACATATAAT</u> ACTAAAGACATT	FoSTeS ou MMBIR

9	Hg38	13	85841126	3	159391325	TATTACCATTGACTTTAAGAAGCTTTCTTCTTTTCA/GTCTTCTTTCTGTATTGCAGGGACTTAACATATAATACTAAAGACATT TTGAGTTAAAAA	FoSTeS ou MMBIR
9	Hg38	3	159391328			GATAGCAATCCTGATGAATTCATGAATTAATAATTTTCATCCAAATGTTAGATTCAAGGTGGCAGTTCAGGGTCT/AGAACCTAG AAACCCTCTGCTAA	NHEJ
9	Hg38	13	43344835	3	166027379	TGCCAAGATGAGAAAATACACTTTATTCTTGTCTCTCAAAGACAGAGCATA/TCTCGGTATTTGAAATACATGAGAAAATATT GTTAATGACAGTCCC	NHEJ
9	Hg38	3	166027312	13	43344836	AAATTTATGAGGTACATGTGCTATTTGGATAAATGCACACAATGTGTAATCATCAAATTAATGTA/GTCCAACCTAGTGGATATATTT TCAGAACTCAGCAA	NHEJ
9	Hg38	3	166027310	13	43344834	TAAATGCACACAATGTGTAATCATCAAATTAATG/TAGTCCAACCTATTTGATATATTTTCAGAACTCAGCAAGTTAGCTTCTCTGAG ACTGACTGCCATAT	NHEJ
9	Hg38	13	43344835	3	166027379	TACACTTTATTCTTGTCTCTCAAAGACAGAGCATA/TCTCGGTATTTGAAATACATGAGAAAATATTGTTAATGACAGTCCCCC TGCTGTGCTATCAA	NHEJ
9	Hg38	13	70702695	1	146741334	TATAAGAAGAATTAATTTTATAGGATCAGGGTTTTCTTTGTCCTTGCACTCGA/CTAGGCTCCCATCCTCTCACAGAACTGGGAGT AGGGACTTACCTT	NHEJ

9	Hg38	13	70702695	1	146741334	GAATTAATTTTATAGGATCAGGGTTTTCTTTGTCTTGCACTCGA/CTAGGCTCCCATCCTCTCACAGAAACTGGGAGTAGGGACTT CACCTTCTTGTTG	NHEJ
9	Hg38			13	70702694	CATCCCTTGTAAAGTTGGATTCTAGGTATTTATTCTCTT/GATCATTGGTAAGAAACACTGCGGTCCAAAACAGCCATAGAATGAC AAAGTCACAAGGCA	NHEJ
9	Hg38	13	79659509			CAATGGGATGACATTTCCGAAGTGGTAATTACCAAAGGATTCCTGTTGATGCCTTTTGGAAGT/TATGACCCATCTTGAGGAAGA TAAATCTGGTTCT	NHEJ
9	Hg38	3	159391328			GAATTAATAATTTTATCCAAATGTTAGATTTCAAGGTGGCAGTTCAGGGTCT/AGAACCTAGAAACCCTCTGCTAAAGCAAACC ATAAACCTAGAAAG	NHEJ

Tableau VIII : Résultats du NGS et de la caractérisation des points de cassure chez les quatre patients présentant un RCAE au niveau du caryotype

Patient présentant un remaniement équilibré au niveau du caryotype							
Patient	Caryotype	ACPA	WGS qualité du Run	WGS: Nombre de point de cassure	Gène interrompu	Eléments répétés présents au niveau du point de cassure	Mécanisme (dédit)
			Profondeur moyenne				
6	46,XY,inv(8)(p2?3;q23)	arr[hg19] 8p12p11.22(31490519_39144107)x1	12,7X	3	<i>TNKS</i>	LTR	NHEJ
					<i>CPQ</i>	hAT-Tip100	
						SINE - LINE	
7	46,XY,inv(11)(p13;q12)	Absence de CNV pathogène	13.4X	2	-	-	MMBIR/FoSTeS
8	46,XY,t(1;2)(q21;p13)	arr[hg19] 2p22.2p23.1(30908598_37914388)x1, 2p16.3 (50598357_51707702)x1	13.3X	5	<i>RALGPS2</i>	SINE - SINE	chromothripsis
						SINE - SINE	
					<i>NRXN1</i>	- - LINE	
						SINE - -	
9	une translocation complexe entre les chromosomes 1, 3, 6 et 13.	FISH WCP des chromosomes 1 6 et 13	-	9	<i>ENOX1</i>	-	chromothripsis
					<i>IQCJ-SCHIP1</i>		
					<i>HYDIN2</i>		

4.3. Discussion

Nous avons caractérisé chez neuf patients (7 cas post-nataux et 2 cas pré-nataux) des remaniements chromosomiques équilibrés ou déséquilibrés au niveau du caryotype. Ces remaniements étaient survenus *de novo* à l'exception d'un seul cas hérité (Patiente 2). Cette étude confirme que le WGS est un outil rapide et efficace pour la caractérisation des points de cassure. Nous avons détecté beaucoup plus de points de cassures que ce qui était suspecté par le caryotype, ce qui est concordant avec la littérature (176).

Concernant les remaniements déséquilibrés, le bénéfice du WGS par rapport au caryotype et à l'ACPA est de permettre la description à la base près de la structure de ces remaniements. Par exemple, pour le patient 1, nous avons pu ainsi préciser l'orientation de la duplication, en tandem. Cette duplication, impliquant le gène *PTPN11*, est responsable du phénotype du patient 1.

Chez la patiente 2, le dérivé de la translocation maternelle, détecté par les différentes techniques est responsable du phénotype.

Le patient 3 présente à la fois une anomalie de structure et une anomalie de nombre. La délétion interstitielle du bras long du chromosome 4 est responsable du phénotype chez le patient. La présence de deux chromosomes Y ne permet pas d'expliquer le phénotype.

Chez les patients 4 et 5, l'ACPA a révélé plusieurs délétions et duplications au niveau d'un chromosome, suggérant un chromoanasyndromes. Le WGS est le seul moyen efficace pour vérifier cette hypothèse. Chez le patient 4 (fœtus 1), le WGS a détecté 57 points de cassure au niveau du bras long du chromosome 13. Le WGS a détecté 18 points de cassure au niveau du chromosome 19 chez la patiente 5. Cependant, pour ce remaniement, l'implication du centromère ne nous a pas permis de caractériser l'ensemble des points de cassures. Le caryotype et la seule technique qui permette de détecter la présence d'un chromosome dicentrique. Ces délétions et duplications impliquent plusieurs gènes. Des études plus poussées vont être effectuées afin d'établir une corrélation génotype - phénotype.

Dans le cas des anomalies équilibrées, le WGS a présenté le grand avantage de caractériser les points de cassure chez les 4 patients en une seule technique.

Chez le patient 6, le WGS nous a permis de localiser tous les points de cassures et de détecter la présence d'un déséquilibre cryptique chez le patient permettant d'expliquer le phénotype.

Chez le patient 7, après avoir caractériser les points de cassure de l'inversion, nous avons utilisé les données de WGS pour rechercher les variations nucléotidiques. Nous avons détecté une variation pathogène, secondairement confirmée par la méthode de séquençage Sanger : une délétion de 2 deux bases, c.2954_2955delCT (NM_0224554), au niveau du gène *NSDI*. Ce gène est responsable du syndrome de SOTOS et explique le phénotype chez le patient. Pour les patients 8 et 9, le WGS a identifié beaucoup plus de points de cassure que ce qui était attendu d'après le caryotype, mettant ainsi en évidence des remaniements de type chromoanagenesis.

Pour le remaniement hérité chez la patiente 2, un conseil génétique est recommandé à la famille et un diagnostic prénatal est souhaitable avant chaque grossesse. Pour les autres familles, il ne faut pas éliminer une anomalie chromosomique en mosaïque germinale chez l'un des parents qui peut être l'origine d'une récurrence de l'anomalie chromosomique. A noter que pour le syndrome de Sotos, les variations peuvent être héritées dans 5% des cas. Dans ce syndrome la pénétrance est complète mais les patients présentent une grande variabilité d'expression (177).

5. Discussion

Cette étude a permis la caractérisation moléculaire de remaniements chromosomiques équilibrés ou déséquilibrés, hérités ou *de novo* en prénatal et postnatal. Ces anomalies chromosomiques sont impliquées dans de nombreuses maladies y compris les troubles de la reproduction et les anomalies du développement. Dans certains cas, grâce à la complémentarité des techniques, les points de cassure ont été élucidés. Dans quelques cas, le point de cassure n'a pas pu être identifié. Ceci peut être dû au fait que le génome de référence, même avec la plus récente version GRCh38.p13, présente encore des « lacunes » (ou gap) (178). Le point de cassure peut également être localisé dans un élément répété, et ainsi ne pas pouvoir être aligné de façon fiable sur le génome de référence. Les éléments répétés représentent plus que 50% du génome humain (Lander et al., 2001). Ils sont particulièrement présents dans les régions centromériques (constituées de séquences α satellites), les régions riches en hétérochromatine, les régions télomériques ou bien les duplications segmentaires. Le WGS peut être mis en difficulté pour détecter des points de cassure dans ces régions et nécessite l'utilisation d'autres approches afin de les identifier.

D'autres approches ont été développées afin de résoudre le problème des régions répétitives, soit en modifiant la préparation des librairies, avec les techniques *mate pair* ou *linked read*, soit en utilisant de nouvelles technologies de séquençages appelées « Long reads » (180). D'autres types de librairies peuvent être utilisées afin de construire de longs fragments tels que le *mate pair*. Cette technique consiste à fragmenter l'ADN en de grands fragments de l'ordre de 3 à 10 Kb, puis de circulariser ces fragments autour d'une séquence adaptatrice par ligation. Un clivage enzymatique des fragments circularisés est effectué afin d'obtenir un fragment court linéaire composé de la séquence adaptatrice et des deux extrémités du grand fragment initial apte à être séquençé (181). En pratique, ces librairies utilisent à la fin un séquençage classique short-read. Huit à 9 % des remaniements ne peuvent pas être caractérisés en utilisant ce type de librairie (182).

Une autre approche, *linked read*, proposée par la société 10X Genomics, consiste à inclure de grands fragments d'ADN, de l'ordre de 60 à 100 Kb, dans une gouttelette. Dans chaque gouttelette, chaque fragment est découpé en plus petits morceaux auxquels s'ajoute une séquence d'ADN spécifique à chaque gouttelette. Celle-ci permettra après le séquençage, d'associer les petits fragments de chaque gouttelette et ainsi de reconstituer des fragments de grande taille *in silico* (183). L'équipe de cytogénétique des Hospices Civils de Lyon a contribué

à une étude de comparaison entre cette approche avec la technique *short read*. Les deux techniques ont pu détecter les remaniements chromosomiques chez 10 des 13 patients inclus dans l'étude. Cependant, la technique linked-read a pu détecter un point de cassure localiser dans un LINE, au niveau du chromosome X (183), qui n'avait pas été identifié par la technique short-read.

D'autres approches ont été développées pour séquencer des fragments de grande taille, ou *long read*, la « Single Molecule Real Time Sequencing (SMRT) » par la société Pacific Biosciences et le séquençage Nanopore de « Oxford Nanopore Technologies (ONT) » (184). La principale limitation de ces technologies est leur taux d'erreur élevée par base distribuée aléatoirement sur les séquences. Une profondeur de lecture élevée est nécessaire pour dépasser ces erreurs, rendant ces techniques très coûteuses (183). Le principe de la SMRT est basé sur la fragmentation de l'ADN de 10 à 20 kb. Après réparation de leurs extrémités, les fragments doubles brins sont circularisés par des adaptateurs en formes d'épingle à cheveux. Ce modèle s'appelle « SMRTbell ». Une fois généré, l'ADN circulaire est incorporé dans un puits dans lequel une ADN polymérase est liée et permet de synthétiser le brin complémentaire. L'incorporation des bases azotées fluorescentes génèrent des signaux fluorescents qui vont établir la séquence d'ADN (184). Récemment, le développement de la technologie CCS (circular consensus sequencing), permettant le séquençage de la même molécule d'ADN plusieurs fois, a amélioré la précision des séquences de 99.8% avec une longueur moyenne de 13,5 Kb (185). La deuxième approche, ONT, utilise des fragments d'ADN linéaires de l'ordre d'un à plusieurs centaines de Kb, à laquelle s'ajoute des adaptateurs pré-chargés par une protéine motrice. Des protéines de liaison sont liées à ces fragments, qui lorsqu'ils sont chargés sur une *flowcell*, composée de milliers de nanopores, permettent au fragment d'ADN de se lier au nanopore. L'adaptateur de séquençage s'insère dans l'ouverture du nanopore et la protéine motrice dirige l'ADN dans le nanopore et module sa vitesse de passage dans le pore. Le passage de l'ADN provoque des perturbations caractéristiques du courant. Cette perturbation, correspond à une séquence spécifique, permettant d'identifier la séquence d'ADN (184).

La cartographie optique (Bionano) est une autre approche pour identifier les variants structuraux. Elle n'est pas basée sur le séquençage. Son principe consiste à établir une carte optique du génome basée sur l'ordre et la taille des fragments obtenus après marquage enzymatique fluorescent. Après avoir extrait l'ADN à haut poids moléculaire, de l'ordre de 200 à 400 Kb, une enzyme de restriction, endonucléase, clive un brin d'ADN et permet l'incorporation d'un fluorochrome au niveau d'un motif spécifique. Les molécules sont

linéarisées dans des nanocanaux, permettant la prise d'images qui seront converties en molécule d'ADN. Un assemblage *de novo* est effectué afin de construire la carte optique du génome étudié. Cette carte est comparée à une carte de référence élaborée à des génomes de référence. Cette comparaison permet de mettre en évidence des variations de structure équilibrées et des CNVs. L'équipe de cytogénétique de l'Hospices Civils de Lyon, a contribué à une étude en utilisant la cartographie optique dans l'identification des différents types d'anomalies chromosomiques y compris les aneuploïdies, les CNVs ainsi que les anomalies chromosomiques équilibrées (186). Cette approche permet une étude pangénomique et permet de détecter en un test unique les variants structuraux et les CNVs et pourrait compléter ou même le caryotype et l'ACPA.

Dans la majorité des cas, et comme décrit dans la littérature (187), la plupart des points de cassures sont localisés dans des éléments répétés. Cette étude a souligné l'intérêt du WGS dans la compréhension des mécanismes des remaniements chromosomiques. Le mécanisme le plus fréquent est le NHEJ. Le WGS a mis en évidence des remaniements complexes de type chromoanagenesis. C'est le seul outil efficace pour la détection et la caractérisation de ces remaniements. Nous avons mis en évidence deux cas de chromoanasythesis et deux cas de chromothripsis. Dans la littérature, les chromoanagenesis, comme les RCAE, peuvent être associés à un phénotype normal, et malgré leur complexité ces remaniements peuvent être transmis (188). Ces événements peuvent conduire à l'interruption de gènes ou bien à la modification de l'expression des gènes par effet de position. Comme plusieurs gènes sont impliqués, le séquençage d'ARN à haut débit est souhaitable, mais pose la question du tissu le plus pertinent à utiliser ou même le type cellulaire dans un tissu donné du fait de l'expression cellulaire différentielle dans chaque type cellulaire.

La perturbation de l'organisation chromatinienne via l'altération des domaines associés topologiquement (TAD) peut entraîner une modification de l'expression des gènes (189). L'analyse de l'organisation spatiale de la chromatine dans une cellule, par le séquençage Hi-C, peut contribuer à élucider la pathogénicité des variants structuraux, en particulier dans les régions non codantes du génome humain. Cette technique quantifie le nombre d'interactions entre les loci génomiques, dont la fréquence de contact entre les paires de locus est fortement corrélée à la distance (190). Cette distance peut être proche en structure tridimensionnelle mais très éloignée dans le génome linéaire.

Cette étude nous a permis d'identifier deux gènes candidats : *LACTB2* et *DNAH5* impliqués dans l'infertilité masculine. Dans le cadre des anomalies du développement des gènes candidats ont été retrouvés et nécessitent des études fonctionnelles afin de confirmer leur implication dans le phénotype des patients : *STARD13*, *NBEA*, *FAM155A*, *RBM26*, *GPC5* chez le patient 4 présentant un cas de chromoanasythesis au niveau du chromosome 13 ; *TSHZ3*, *PSG3* et *PSG6* chez le patient 5 présentant un cas de chromoanasythesis au niveau du chromosome 19 et les gènes *ENOX1*, *IQCJ-SCHIP1* et *HYDIN2* chez le patient 9 présentant un cas de chromothripsis impliquant plusieurs chromosomes. En plus des gènes interrompus par les points de cassures, le phénotype des patients peut être dû à un effet de position.

L'étude d'une cohorte plus grande de patients présentant un trouble spermatique associé à un RCAE sera nécessaire pour confirmer la pertinence de l'approche par séquençage génome entier avec analyse des points de cassure et recherche de SNVs en routine diagnostique. Elle permettrait également d'identifier de nouveaux gènes impliqués dans l'infertilité masculine Cette exploration permettrait d'améliorer la prise en charge et d'éviter des traitements inutiles. Par exemple, les mutations du gène *AURKC*, impliqué dans la spermatogenèse, induisent un blocage de la première division méiotique, entraînant la production de spermatozoïdes tétraploïdes macrocéphales (191). Dans ce cas, une injection intracytoplasmique de spermatozoïde dans l'ovocyte ne permettra pas d'obtenir de grossesse.

La complémentarité des techniques nous a aidé à l'identification des points de cassures. La technique de FISH nous a permis de localiser les points de cassure chez deux des trois patients présentant des troubles de la reproduction.

Nous avons montré que le RCAE peut être associé à un variant nucléotidique pathogène responsable du phénotype. Un variant pathogène dans le gène *NSDI* a été mis en évidence en plus de l'inversion péricentrique du chromosome 11 chez le patient 7. Chez le patient 3 du groupe infertilité présente deux variants hétérozygotes du gène *DNAH5* interprétés comme candidats potentiels dans le trouble de la reproduction du patient en plus de la translocation réciproque. Seule l'approche par WGS, en utilisant plusieurs outils bio-informatiques, permet l'étude des anomalies nucléotidiques, des CNVs et des variants structuraux en une seule technique. Avec le progrès technologique et les avancées dans les outils bio-informatiques, la question se pose de savoir si le WGS va remplacer les autres techniques ? Dans cette étude malgré l'avantage du WGS, l'apport d'autres techniques était nécessaire pour caractériser les remaniements. Ces nouvelles technologies engendrent beaucoup de données que nos

connaissances actuelles ne permettent pas toujours d'interpréter. L'examen clinique et la corrélation génotype – phénotype reste la base afin de conclure sur la pathogénicité de ces variants.

6. Conclusion

L'avènement des nouvelles technologies, en particulier le séquençage haut-débit, a révolutionné la génétique. En effet un nombre croissant de gènes ont été identifiés comme responsables d'anomalies congénitales et de troubles de la reproduction. L'identification de ces gènes n'est que le début d'une recherche afin d'approfondir nos connaissances dans le processus et l'implication de ces gènes dans la pathologie.

La gamétogenèse est un processus très complexe nécessitant l'implication d'un grand nombre de gènes. Théoriquement, une variation pathogène dans un de ces gènes entraîne une infertilité. Certaines anomalies génétiques dans l'infertilité sont récurrentes, mais des cas idiopathiques peuvent résulter d'une cause génétique chez le patient. Notre étude a mis en évidence deux gènes candidats, *LACTB2* et *DNAH5* impliqués dans l'infertilité masculine. Malgré les limites rencontrées lors des analyses des variations nucléotidiques du fait de bases de données contrôles peu adaptées au phénotype d'infertilité le séquençage haut-débit semble être un outil prometteur pour le diagnostic de l'infertilité masculine. Des études englobant un grand nombre de patients présentant une altération spermatique sont nécessaires ainsi que la construction d'une base de données spécifiques à l'infertilité.

Dans les anomalies congénitales, un grand nombre de gènes a déjà été rapporté. La caractérisation moléculaire des points de cassures chez neuf patients a mis en évidence des gènes candidats, *STARD13*, *NBEA*, *FAM155A*, *RBM26*, *GPC5*, *TSHZ3*, *PSG3*, *PSG6*, *ENOX1*, *IQCJ-SCHIP1* et *HYDIN2*, qui peuvent être responsable du phénotype chez les patients. En plus cette étude a mis en évidence la présence de *chromoanagenesis*, comme mécanisme responsable des remaniements chromosomiques chez ces patients. La connaissance de ces mécanismes et de ses origines est nécessaire dans un objectif de conseil génétique. Avec les progrès technologiques et l'arrivée des séquenceurs de troisième génération permettant de séquencer de longs fragments d'ADN, la question du transfert en diagnostique de ces techniques est posée.

7. Annexe 1 :

The enrichment of breakpoints in late-replicating chromatin provides novel insights into chromoanagenesis mechanisms.

1 **Title page**

2 The enrichment of breakpoints in late-replicating chromatin provides novel insights into
3 chromoanagenesis mechanisms

4

5 **Authors**

6 Nicolas Chatron ^{1,2,3*}, Giuliana Giannuzzi ³, Pierre-Antoine Rollat-Farnier ^{1,4}, Flavie Diguët ^{1,2}, Eleonora
7 Porcu ³, Tony Yammine ^{1,2}, Kevin Uguen ⁵, Zohra-Lydia Bellil ^{1,2}, Julia Lauer Zillhardt ⁶, Arthur Sorlin
8 ⁷, Flavie Ader ⁸, Alexandra Afenjar ⁹, Joris Andrieux ¹⁰, Claire Bardel ⁴, Eduardo Calpena ¹¹, Sandra-
9 Chantot-Bastaraud ¹², Patrick Callier ⁷, Nora Chelloug ¹³, Emilie Chopin ¹⁴, Marie-Pierre Cordier ¹,
10 Christèle Dubourg ¹⁵, Laurence Faivre ¹⁶, Françoise Girard ¹⁷, Solveig Heide ¹⁸, Yvan Herenger ¹⁹, Sylvie
11 Jaillard ²⁰, Boris Keren ¹⁸, Samantha J. L. Knight ²¹, James Lespinasse ²², Laurence Lohmann ²³, Nathalie
12 Marle ⁷, Reza Maroofian ²⁴, Alice Masurel-Paulet ¹⁶, Michèle Mathieu-Dramard ²⁵, Corinne Metay ²⁶,
13 Alistair T. Pagnamenta ²¹, Marie-France Portnoï ¹², Fabienne Prieur ²⁷, Marlène Rio ²⁸, Jean-Pierre
14 Siffroi ¹², Stéphanie Valence ²⁹, Jenny C. Taylor ²¹, Andrew O. M. Wilkie ¹¹, Patrick Edery ^{1,2}, Alexandre
15 Reymond ³, Damien Sanlaville ^{1,2}, Caroline Schluth-Bolard ^{1,2}

16

17 **Authors' affiliations**

- 18 1. Service de Génétique, Hospices Civils de Lyon, Lyon, France
- 19 2. Equipe GENDEV, CRNL, INSERM U1028, CNRS UMR5292 UCBL1, Lyon, France
- 20 3. Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland
- 21 4. Cellule Bioinformatique Plateforme NGS, Hospices Civils de Lyon, Lyon, France
- 22 5. Service de Génétique, CHU Brest, Brest, France
- 23 6. Unité de Diagnostic Préimplantatoire, Laboratoires de Diagnostic Génétique, Hôpitaux
24 Universitaires de Strasbourg, Strasbourg, France
- 25 7. Laboratoire de génétique chromosomique et moléculaire, CHU de Dijon, Dijon, France
- 26 8. AP-HP, Département de Génétique médicale, Cardiogénétique et myogénétique moléculaire et
27 cellulaire, Hôpital de la Pitié-Salpêtrière, Paris, France
- 28 9. AP-HP, Département de Génétique médicale, UF de Génétique clinique, hôpital d'Enfants
29 Armand Trousseau, Paris, France
- 30 10. CHRU de Lille, Service de Génétique, Hôpital Jeanne de Flandre, 59037 Lille Cedex, France

- 31 11. Clinical Genetics Group, MRC Weatherall Institute of Molecular Medicine, University of
32 Oxford, Oxford OX3 9DS, UK
- 33 12. AP-HP, Département de Génétique médicale, UF de Génétique chromosomique, hôpital
34 d'Enfants Armand Trousseau, Paris, France
- 35 13. Laboratoire de cytogénétique, Service de génétique médicale, CHRU Tours, Tours, France
- 36 14. Centre de Biotechnologie Cellulaire, Hospices Civils de Lyon, Lyon, France
- 37 15. Service de Génétique Moléculaire et Génomique médicale, CHU Rennes, et IGDR, CNRS
38 UMR 6290, Université de Rennes 1, Rennes, France
- 39 16. Centre de génétique, Hôpital d'enfants, CHU de Dijon, Dijon, France
- 40 17. Laboratoire de cytogénétique constitutionnelle et prénatale, CHU Strasbourg, France
- 41 18. APHP, Département de Génétique et Centre de Référence Déficiences Intellectuelles de Causes
42 Rares, Hôpital de la Pitié-Salpêtrière, Paris 75651, France
- 43 19. Service de génétique médicale, CHRU Tours, Tours, France
- 44 20. Univ Rennes, CHU Rennes, INSERM, EHESP, IRSET (Institut de recherche en santé,
45 environnement et travail) – UMR_S 1085, F-35000 Rennes, France
- 46 21. Oxford NIHR Biomedical Research Centre, Wellcome Centre for Human Genetics, University
47 of Oxford, Oxford, UK
- 48 22. Service de génétique, CH Métropole Savoie, Chambéry, France
- 49 23. Département de génétique, Laboratoire CERBA, Saint-Ouen l'Aumône, France
- 50 24. Institute of Neurology, University College London, London, UK
- 51 25. Service de génétique clinique et oncogénétique, CHU Amiens-Picardie, Amiens, France
- 52 26. AP-HP, Hôpital Henri Mondor, Créteil, France
- 53 27. Service de Génétique Clinique, Chromosomique et Moléculaire, CHU Hôpital Nord, Saint-
54 Etienne, France
- 55 28. AP-HP, Service de Génétique Médicale, Hôpital Necker-Enfants Malades, Paris, France
- 56 29. Sorbonne Université, GRC n°19, pathologies Congénitales du Cervelet-LeucoDystrophies, AP-
57 HP, Hôpital Armand Trousseau; Hôpital Armand Trousseau, APHP, GHUEP, Service de
58 Neurologie Pédiatrique, France

60

61 ***Corresponding author:**

62 Dr Nicolas CHATRON

63 Service de Génétique, Centre de Biologie Pathologie Est, 2^{ème} étage

64 Groupement Hospitalier Est, Hospices Civils de Lyon

65 69677 Bron Cedex, France

66 Tel : +33471129640

67 E-mail : nicolas.chatron@chu-lyon.fr

68

69 **Keywords:**

70 chromoanagenesis, chromothripsis, replication-timing, lamina-associated domains, chromosomal
71 breakpoint

72

73 **Running title**

74 Chromoanagenesis breakpoints enrichment in late-replicating chromatin.

75 **Abstract**

76 The rise of pangenomic molecular assays allowed uncovering complex rearrangements named
77 *chromoanagenesis* that were hypothesized to result from catastrophic shattering events. Constitutional
78 cases have typically been reported individually preventing identification of common features and
79 uncovering the mechanisms at play. We characterized 20 new *chromoanagenesis* and discovered yet
80 undescribed features. While literature differentiates *chromothripsis* and its shattering event repaired
81 through non-homologous end joining from *chromoanagenesis* born to aberrant replicative processes,
82 we identified shattered chromosomes repaired through a combination of mechanisms. In particular, three
83 samples present with “rearrangement hubs” comprising a fragmented kilobase-long sequence threaded
84 throughout the rearrangement.

85 To assess the mechanisms at play, we merged our data with those of 20 published constitutional complex
86 chromosomal rearrangement cases. We evaluated if the distribution of their 1032 combined breakpoints
87 was distinctive using bootstrap simulations and found that breakpoints tend to keep away from
88 haplosensitive genes suggesting selective pressure. We then compared their distribution with that of
89 13,310 and 468 breakpoints of cancer complex chromosomal rearrangements and constitutional simple
90 rearrangement samples, respectively. Both complex rearrangement groups showed breakpoint
91 enrichment in late replicating regions suggesting similar origins for constitutional and cancer cases.
92 Simple rearrangement breakpoints but not complex ones were depleted from lamina-associated domains
93 (LADs), possibly as a consequence of reduced mobility of DNA ends bound to lamina.

94 The enrichment of breakpoints in late-replicating chromatin for both constitutional and cancer
95 *chromoanagenesis* provides an orthogonal support to the premature chromosome condensation
96 hypothesis that was put forward to explain *chromoanagenesis*.

97 **Introduction**

98 Since the identification of a supernumerary chromosome 21 in Down syndrome (Lejeune et al. 1959),
99 each technological development (chromosome banding, FISH, cytogenetic microarray, massive parallel
100 sequencing, long-read sequencing) has further exposed the complexity and variability of human
101 chromosomes (Caspersson et al. 1970; Cooper et al. 2011; Sudmant et al. 2015). This “entropy”
102 climaxed with the identification of extremely complex chromosomal rearrangements (CCRs) both in
103 cancer (Stephens et al. 2011) and constitutional samples (Kloosterman et al. 2011) christened
104 *chromothripsis*. These CCRs have their breakpoints clustered to a single chromosome, a chromosome
105 arm or a cytoband and their copy number profile oscillating between two states (1 and 2 copies), in
106 contrast to the classical tumorigenesis scenario. The equal amount of inverted and non-inverted
107 breakpoint-junctions is consistent with a single catastrophic shattering event (*-thripsis* in Greek)
108 followed by random patching through non-homologous end-joining coupled to fragment loss (Korbel
109 and Campbell 2013). The identification of different copy number profiles and alternative replicative
110 repairing mechanisms (Microhomology-mediated breakage induced repair (MMBIR) / Fork Stalling and
111 Template Switching (FoSTeS)) prompted the definition of *chromoanasythesis* (Liu et al. 2011) and
112 *chromoplexy* involving several chromosomes and only observed in cancer (Baca et al. 2013), with the
113 three phenomena being grouped under the umbrella term *chromoanagenesis* (Holland and Cleveland
114 2012).

115 There are two main hypotheses for the underlying mechanisms, mostly deriving from cancer cell
116 observation and cellular models. Firstly, a non-spindle bound lagging chromosome is segregated in a
117 micronucleus (Ganem et al. 2009) where DNA replication is deficient (Crasta et al. 2012; Terzoudi et
118 al. 2015) generating two asynchronous compartments, the micronucleus and main nucleus. This
119 replicating and entrapped chromosome undergoes premature chromosome condensation, which is
120 associated with pulverization of chromatin (Kürten and Obe 1975; Obe and Beek 1975). The second
121 model relies on the dicentric chromosomes (Titen and Golic 2008). Dicentric chromosomes can create
122 bridges between the cytokinesis poles that can only be resolved through multiple DNA breaks
123 (Maciejowski et al. 2015), sometimes with entrapment in a micronucleus (Pampalona et al. 2016). Other
124 hypotheses (incomplete apoptosis (Tubio and Estivill 2011), hyperploidy (Mardin et al. 2015) or mobile
125 elements) have also been formulated.

126 The observed continuous rather than multimodal distribution of the number of breakpoints from simple
127 to highly complex rearrangements was suggested to mirror a common origin and mechanism (Storchová
128 and Kloosterman 2016). Would this continuum hypothesis be correct, we would expect a similar
129 distribution of breakpoints in simple and complex events. However, beyond low copy repeats known to
130 be elective sites for recurrent non-allelic homologous recombination events (Stankiewicz and Lupski
131 2002), we are ignorant about possible relationships between genomic context (sequence, genome

132 folding, genome properties) and breakage probability, and potential differences in regard to complexity
133 of the rearrangement.

134 Here we take advantage of three cohorts of rearrangements – simple, constitutional CCRs and cancer
135 CCRs – to assess and compare the genomic features of breakpoints and provide evidence for the
136 previously hypothesized continuum between simple and complex events. We suggest new elements in
137 the potential mechanisms at play.

138

139 **Results**

140 *Chromoanagenesis characterization reveals new molecular features*

141 We recruited 14 unbalanced and 6 balanced *chromoanagenesis* cases. To be included and genome
142 sequenced unbalanced cases had to carry a minimum of 3 non-recurrent non-polymorphic CNVs
143 identified through array-CGH on a single chromosome (potentially chromosome pair) while balanced
144 cases with a minimum of 10 breakpoints were obtained from our translational research study (Schluth-
145 Bolard et al. 2019). Through paired-end short-read genome sequencing, we identified thirteen to hundred
146 breakpoints per rearrangement involving from one to fifteen chromosomes for a total of 682 breakpoints
147 (Table 1 and Figure 1). A detailed description of each case is provided as supplementary data
148 (Supplementary File S1).

149 This characterization effort uncovered noteworthy features. Individual 20 (I-20) carries the most
150 complex *chromoanagenesis* described to date with a single derivative chromosome 14 made of
151 sequences originating from 15 different chromosomes. At several chromosome 14 breakpoints, we
152 observe a “juxtapositions of sequences” from chromosome to chromosome before joining chromosome
153 14 again until the next “juxtaposition”. Two other single-chromosome cases (I-3 and I-7) also present
154 with copy number gains embedded in junctions but all gains are chained and inserted as a single block.
155 I-10 also shows a single chain of gains inserted as a block. This case is unique as the five gains all derive
156 from chromosome 1 but are inserted in a different chromosome (chr17) close to a paracentric inversion
157 (Figure 2). For these three cases (I-3, I-7 and I-10), the number of junctions not implicating a copy
158 number gain boundary is very limited suggesting that a chromosome shattering is an unlikely cause.

159 Of the remaining rearrangements with multiple chromosomes implicated, I-19 carries the second most
160 complex rearrangement with six chromosomes involved. However, his profile is very different from that
161 of individual 20 as chromosomes 4, 13 and 14 all present with a shattered profile with at least 5
162 breakpoints each and many junctions connecting pairs of these three chromosomes suggesting that they
163 all had been shattered. Chromosome 11 is only involved in this rearrangement through a 6 kb duplication
164 inserted within a derivative chromosome. I-16 (chromosomes 8 and 14) and I-18 (chromosomes 3 and
165 10) also harbor more than one shattered chromosome.

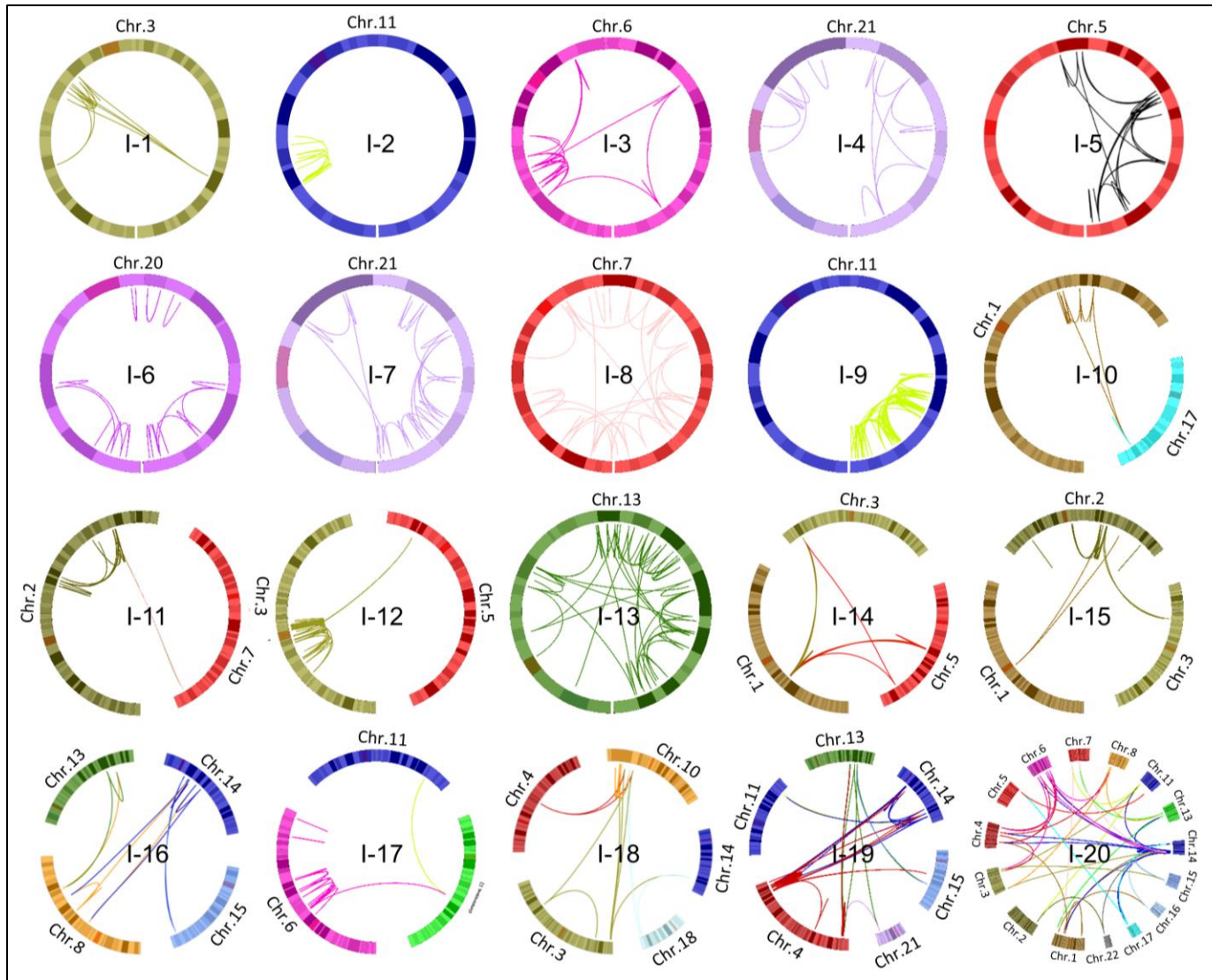
166 Rearrangements of individuals 12, 13 and 15 are a composite of a shattered chromosome and one or two
167 translocations, a situation compatible with a preexisting parental balanced translocation (absent here) or
168 a two-step process during a single gametogenesis.

Individual	No of chromosomes	No of breakpoints	Main chromosome	Copy Number Variant		No of resolved junctions	No of inferred junctions	Inferred mechanism			Inferred mechanism for rearrangement
				Gain in Mb (n gain)	Loss in Mb (n loss)			NHEJ	MMBIR (MMEJ)	Ambiguous	
1	1	14	3 (100%)	0	0	13	4	1	3	0	NA
2	1	16	11 (100%)	2.25 (2)	7.15 (4)	9	7	3	2	2	Chromoanasythesis
3	1	18	6 (100%)	1.43 (7)	6.56 (1)	12	9	1	5	3	Chromoanasythesis
4	1	18	21 (100%)	5.30 (5)	11.73 (4)	9	6	5	1	0	Chromoanasythesis
5	1	23	5 (100%)	0	0	23	18	10	5	3	Chromothripsis
6	1	24	20 (100%)	6.65 (11)	0	14	10	3	3	4	Chromoanasythesis
7	1	25	21 (100%)	4.54 (11)	0.31 (1)	14	14	3	8	3	Chromoanasythesis
8	1	62	7 (100%)	0.76 (15)	19.68 (3)	27	27	13	7	7	Chromoanasythesis
9	1	68	11 (100%)	8.74 (14)	4.40 (1)	52	44	11	23	10	Chromoanasythesis
10	2	13	1 (76,9%)	0.83 (5)	0	8	8	8	0	0	Chromoanasythesis
11	2	18	2 (94%)	0	2.97 (2)	15	9	3	6 (1)	0	Chromoanasythesis
12	2	46	3 (98%)	0	9.00 (5)	48	29	12	9	8	Chromoanasythesis
13	2	56	13 (98%)	35.91 (15)	15.24 (8)	33	31	5	24	2	Chromoanasythesis
14	3	18	1 (72%)	0	0	16	10	7	0	3	Chromothripsis
15	3	23	2 (91,3%)	0	0	18	7	4	3	0	NA
16	4	14	14 (43%)	0	0	14	10	5	1	4	Chromothripsis
17	4	19	6 (84%)	0.25 (1)	14.97 (6)	14	6	3	3	0	NA
18	5	18	10 (56%)	0	0	16	11	7	1	3	Chromothripsis
19	6	84	4 (68%)	0	6.26 (16)	71	28	8	12 (2)	8	Chromoanasythesis
20	15	105	14 (67%)	0	4.23 (15)	85	54	39	5	10	Chromoanasythesis
	Total	682				511	341	151	121	70	

169

170 **Table 1:** Rearrangements' characteristics summary.

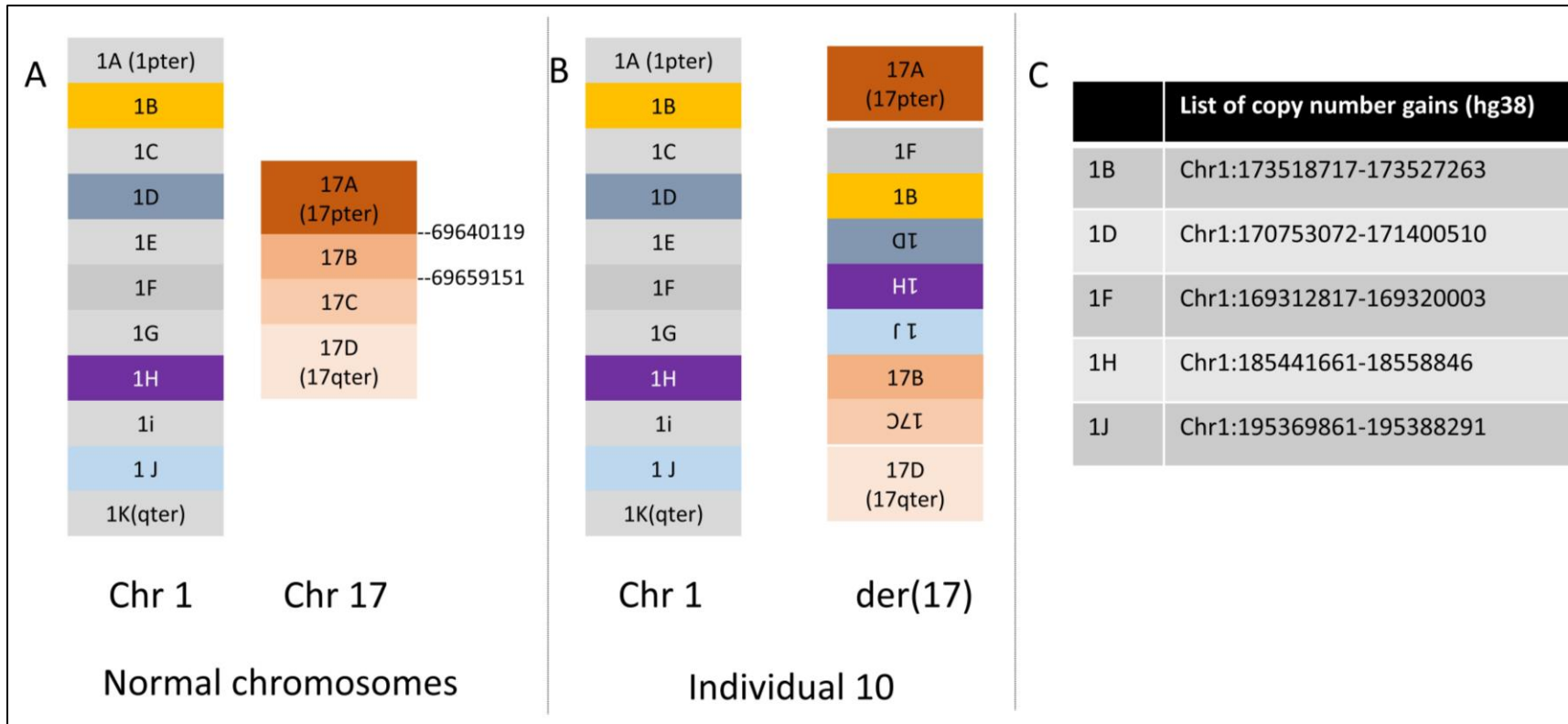
171 We only inferred the junctions for which a same repairing mechanism was considered at both ends. Full details can be found in Supplementary File S1.



172

173

Figure 1: Circos plot visualization of all 20 rearrangements characterized herein. Foci of clustered breakpoints are observed in a majority of individuals.



174

175 **Figure 2:** Schematic reconstruction of I-10's rearrangement: 5 sections of chromosome 1, 1F, 1B, 1D, 1H and 1J (their coordinates are indicated in panel C) are sequentially
 176 inserted in chromosome 17 nearby a 17C paracentric inversion of this chromosome.

177

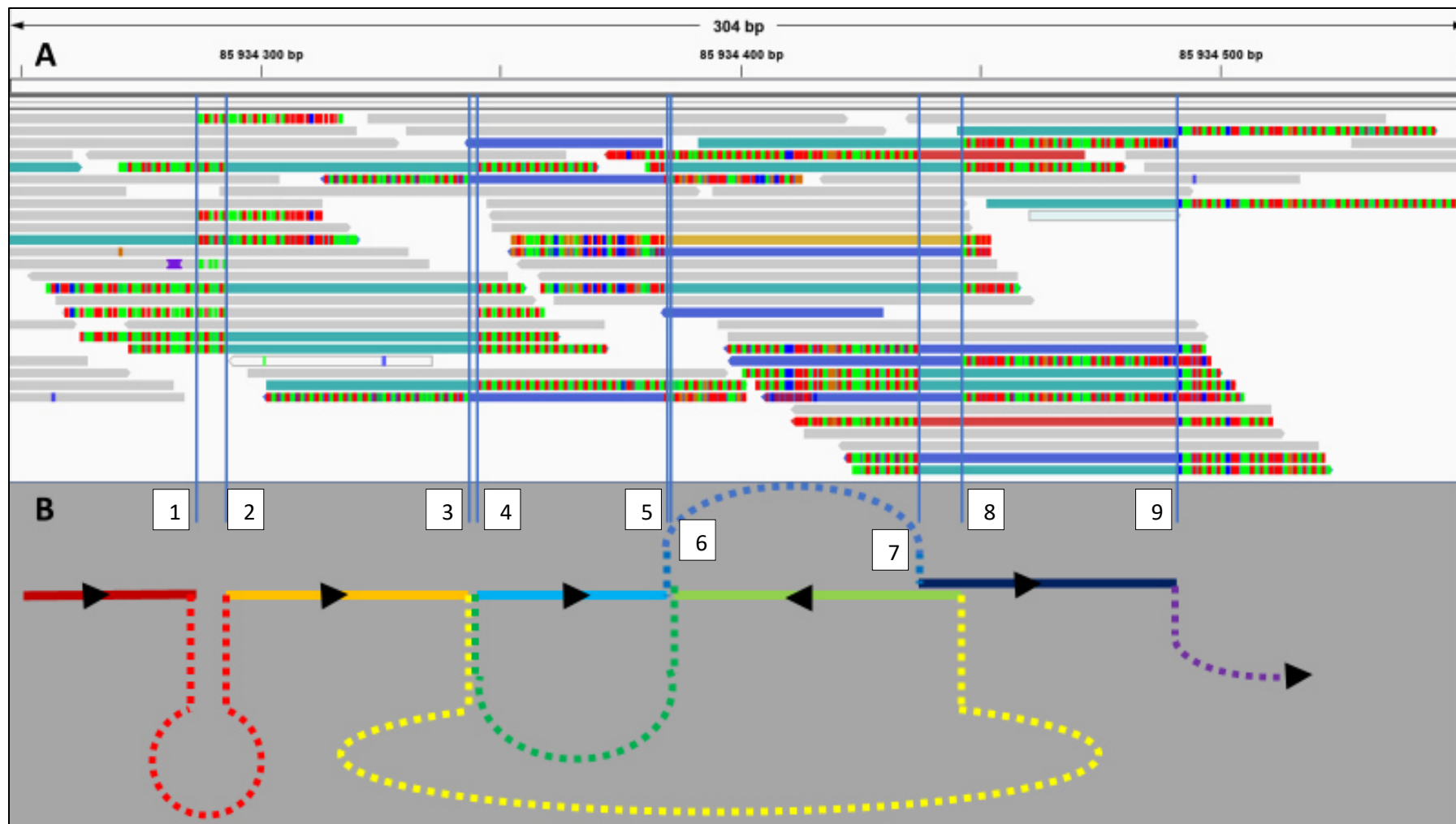
178 *Genotype/phenotype correlation*

179 All individuals, but one, of our cohort had an abnormal phenotype. I-7 was identified after a positive
180 first trimester screening for trisomy 21 that revealed an abnormal chromosome 21. Fifteen individuals
181 presented with different degrees of intellectual disability (ID) from mild (I-1, I-18) to severe (I-14).
182 When present ID was associated with different congenital malformations in all individuals but I-6 and
183 I-17. Of the four individuals without ID, two were not assessed formally as they were identified
184 prenatally (I-13) or in the first days of life (I-12). We did not observe any correlation between the
185 phenotype severity and the rearrangement complexity regarding either number of breakpoints, number
186 of chromosomes involved or genomic imbalance. For several individuals the phenotype could be
187 partially explained by the identification of disrupted genes or a position effect. *FOXP1* (Mental
188 retardation with language impairment (MIM 613670)) (I-1), *MEF2C* (Mental retardation, autosomal
189 dominant 20 (MIM 613443)) (I-5 and I-14), *ADNP* (Helsmoortel-van der Aa syndrome (MIM 615873))
190 (I-6), *ZEB2* (Mowat-Wilson syndrome (MIM 235730)) and *COL3A1* (vascular Ehler-Danlos syndrome
191 (MIM 130050)) (I-11), *YY1* (Gabriele-de Vries syndrome (MIM 617557)) (I-20) could partly drive the
192 respective phenotypes. For I-2, 3, 4, 8, 9, 10, 12, 13, 17 and 19 no major gene could be convincingly
193 identified as driving the phenotype but the genomic imbalance was considered large enough to be
194 causative. The striking absence of phenotype for I-7 while she carries a total genomic imbalance of 4.85
195 Mb on chromosome 21, not affecting the Down Syndrome critical region, could be explained by the
196 insertion of all copy number gains together on the short arm of chromosome 21. Placed in a
197 heterochromatin context the copy number gains could be silenced through a protective position effect.

198

199 *Repairing mechanisms are mixed within rearrangements and create “hubs”*

200 We reached nucleotide resolution in both forward and reverse directions for 511 junctions allowing
201 inferring the repair mechanisms through analysis of junction sequence (Supplementary File S1). Using
202 stringent inferring criteria (Online Methods), 162 out of 511 junctions could not be classified. Non-
203 homologous end-joining (NHEJ) was inferred for 151 junctions (30%), while 121 junctions (24%) fitted
204 with our criteria for MMBIR / FoSTeS. Three of these 121 replicative junctions could also be considered
205 as Microhomology Mediated End Joining (MMEJ) as they had no distinctive feature for either
206 mechanism. For 70/341 junctions (14%) we could not distinguish between NHEJ and MMBIR/FoSTeS.
207 Using our criteria to infer subclasses of CCR we classified 13 cases as *chromoanasythesis* and 4 as
208 *chromothripsis*, leaving 3 rearrangements without clear denomination, mostly due to a limited number
209 of inferable junctions. A single rearrangement in I-10 could be fully inferred with a unique repairing
210 mechanism (NHEJ), whereas all others presented different levels of ascertained mixed signatures and
211 breakpoint distribution was not shown to be significantly different between the two subclasses (data not
212 shown thus questioning the relevance of criteria used for classification and existing definitions.



213

214 **Figure 3:** IGV browser (v2.4) view of read alignment of the largest “rearrangement hub” observed in our cohort (I-12) consisting of 9 breakpoints within a
 215 segment of 211 nucleotides (A). Reads paired with a mate aligned to a different locus are coloured. Partial schematic reconstruction of the rearrangement with
 216 plain lines being sequences of the “hub” and dashed-lines depicting other fragments of the rearrangement. The direction is indicated by black arrowheads (B).

217 At sequence resolution, we describe for the first time features that we name “rearrangement hubs”. They
218 consist of short genomic intervals (<1 kb) contacted by more than four junctions each coming from a
219 different locus and using a few dozens of nucleotides of the hub before restarting another rearrangement
220 loop. We identified 5 of these in I-9 (n=2), 12 (n=2) and I-19 (n=1). The most complex “hub” has 12
221 junctions within a 211 nucleotide span (I-12; Figure 3).

222

223 *Chromoanagenesis breakpoint distribution is not random*

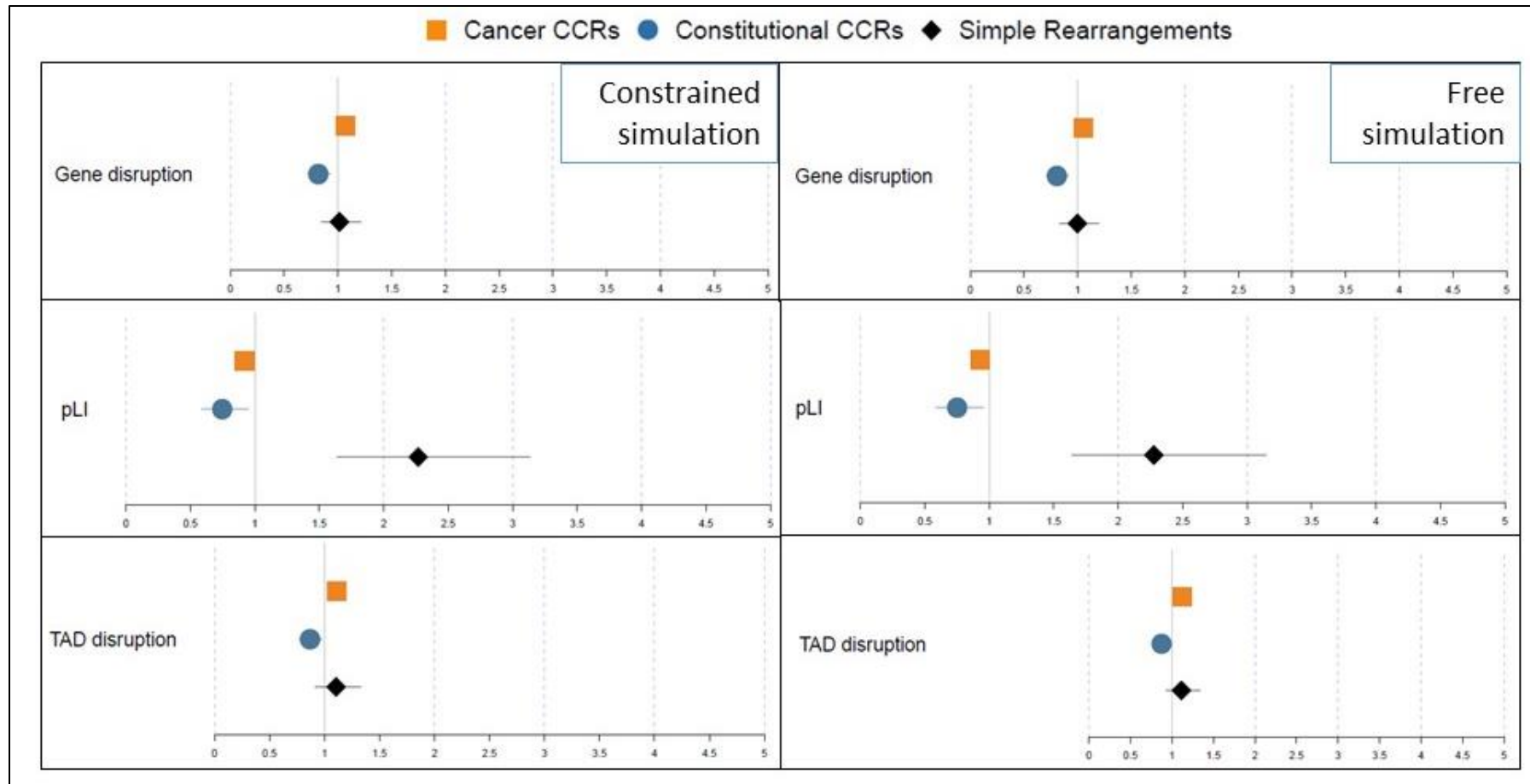
224 We then investigated CCR breakpoint distribution to gain insights on potential biological mechanisms.
225 To increase statistical power, we combined 682 breakpoints of our 20 constitutional CCRs with 350
226 breakpoints of 20 published cases (Yang et al. 2015; Redin et al. 2017). To investigate if the mechanisms
227 involved in constitutional and cancer CCRs are similar we compared the distribution of these 1032
228 breakpoints to that of 13,310 breakpoints of cancer CCRs deposited in ChromothripsisDB (Yang et al.
229 2015). We also compared their distribution with that of 468 breakpoints of 234 simple rearrangements
230 described in (Redin et al. 2017) and (Schluth-Bolard et al. 2019) to test the possible continuum from
231 simple to complex rearrangements. We assessed if the breakpoint distribution of these three different
232 groups of rearrangements were enriched within diverse genomic features, such as haplosensitive genes,
233 repeated elements, G-band staining, chromosomal fragile sites, A/B chromatin compartments, lamina
234 associated domains (LADs), topologically associating domains (TADs), replication origins and
235 replication timing. We are aware that these variables are not independent, LADs, for example, are known
236 to be late-replicating gene poor regions (Supplementary Figure S1). Direction of the distribution
237 (enrichment/depletion) were gauged by comparing to two distinct breakpoint simulations (see Online
238 Methods). All annotations are presented in Supplementary Table S1.

239 First, through univariate analysis, we observed a depletion of both gene-disrupting (Relative Risk Ratio
240 (RRR) = 0.818, Confidence Interval 95% [CI95%] : [0.722-0.927]) and TAD-disrupting (RRR = 0.867
241 [0.767-0.980]) breakpoints in constitutional CCRs (Figure 4 and Supplementary Table S2). In addition
242 to a massive genomic imbalance, we surmise that the accumulation of disrupted or dysregulated genes
243 would reduce fitness. For simple rearrangements gene-disrupting breakpoints tend to affect
244 haplosensitive genes (higher loss-of-function intolerance (pLI); RRR= 2.267 [1.637-3.137]), probably
245 mirroring the ascertainment bias in recruiting these affected individuals. On the contrary we observed a
246 depletion of high pLI genes in CCRs (RRR= 0.746 [0.585-0.952]), suggesting that their pathologies
247 have an oligogenic basis with multiple breakpoint and CNVs contributing and/or that affecting multiple
248 high pLI genes is incompatible with life. Finally, we observed a significant enrichment in genes and
249 TADs in cancer CCRs (RRR = 1.070 [1.034-1.108] and 1.113 [1.075-1.153] respectively) consistent
250 with a distribution driven by oncogenic forces rather than selective pressure.

251 Aside from selective pressure consequences, we identified several genomic features affecting breakpoint
252 distribution. To jointly assess the covariables and confounders (including selective pressure elements)
253 we built a multinomial logistic regression to model the type of rearrangement that originates the
254 breakpoint with the presence/absence of repeated elements, genes, lamina associated domain, and
255 topologically associating domain, the type of chromatin compartment, G-band staining, replication
256 timing status and distance to replication origin (Table 3). In this model, replication timing appears as
257 the unique significant element in the three groups, indeed chromosomal breakpoints are enriched in late-
258 replicating chromatin (RRR = 3.713 [2.583-5.338] for constitutional CCRs; RRR = 1.431 [1.303-1.572]
259 for cancer CCRs; RRR = 2.149 [1.259-3.668] for simple rearrangements) (Figure 5 and Supplementary
260 Table S3).

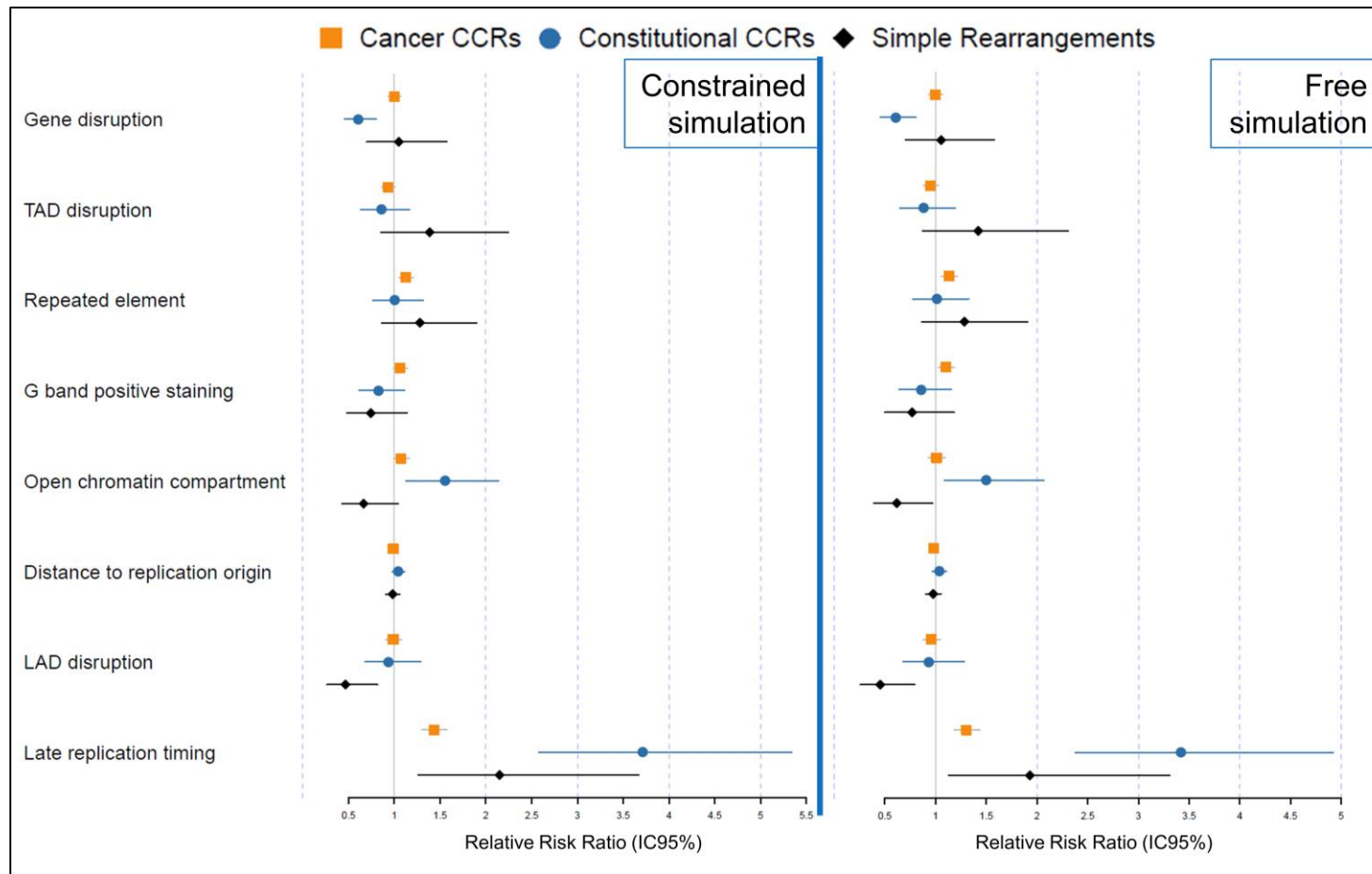
261 For constitutional *chromoanagenesis*, beyond replication timing, depletion in genes (RRR = 0.605
262 [0.454-0.806]) and enrichment in open chromatin regions (RRR = 1.553 [1.129-2.137]) appear as other
263 significant elements. For cancer rearrangements, none of the analyzed covariables but repeated element
264 disruption was significantly different from simulations so that breakpoint distribution is close to random
265 once adjusted for replication timing.

266 Finally, while late-replication is known to be a characteristic of LADs, simple rearrangements'
267 breakpoints appeared depleted within LADs (RRR= 0.466 [0.266-0.816]). Conversely, complex
268 rearrangements' breakpoints were independent of LADs as their distribution is comparable to breakpoint
269 simulation (RRR= 0.987 [0.905-1.077] and 0.935 [0.679-1.289] for cancer and constitutional cases
270 respectively). Our results suggest that preference towards non-peripheral, i.e. non lamina-associated,
271 chromatin appears as the unique criterion distinguishing the distribution of simple rearrangement
272 breakpoints from that of complex cases that are naive to this feature.



273

274 **Figure 4:** Univariate logistic regression model of breakpoint distribution of simple rearrangements, cancer and constitutional CCRs. These distributions are
 275 compared to constrained (left panel) and free simulated distributions (right). Enrichment in genes (top), haplosensitive genes (middle) and TADs (bottom) are
 276 shown. Detailed results are presented in table S2. Constitutional complex rearrangements are depleted in gene-disrupting and TAD-disrupting breakpoints as a
 277 possible consequence of purifying selection. Simple rearrangements' breakpoints are enriched in haploinsufficiency intolerant genes possibly reflecting a
 278 selection bias towards phenotypically abnormal individuals.



279

280 **Figure 5:** Result of the multinomial logistic regression model of breakpoint distribution of simple rearrangement, cancer and constitutional CCRs. These
 281 distributions are compared to constrained (left panel) and free simulated distributions (right) (relative risk ratios and 95% confidence intervals). Distance to
 282 replication origin was log-transformed. Late replication timing is the only significant variable for all three groups. The logistic regression model presented is the
 283 addition of each covariable that has the lowest Akaike information criterion tested (AIC = 36663,63 and 36558.53 for constrained and free simulation as reference
 284 group respectively

285 Discussion

286 With a total of 20 cases, this is the biggest cohort of constitutional CCRs to date offering new insights
287 on the origins and mechanisms of such rearrangements *in vivo*. Interestingly, of the 13 rearrangements
288 we included based on their abnormal array-CGH result, they all presented additional breakpoints and
289 complexity. The coincidence of three non-recurrent non-polymorphic CNVs identified during array-
290 CGH on a single chromosome is a good indication for an unbalanced CCR requiring genome sequencing
291 for detailed characterization. Of note, paired-end short-read sequencing identifies only 91% of
292 karyotype-visible breakpoints (Redin et al. 2017) suggesting that an extra loop of junctions cannot be
293 ruled out in our “solved” cases.

294

295 Our characterization effort challenged several literature definitions. First, we could not disentangle
296 situations where gains are inserted in a single block (I-3, 7 and 10) from replicative repair of a high
297 number of breakpoints in a single chromosome “using” a variety of other chromosomes. This situation
298 questions the definition of *chromoanagenesis* subgroups as it implies a grey zone between
299 *chromothripsis* and *chromoanasythesis*. In *chromoanasythesis*, the complexity rather than resulting
300 from chromosome shattering is the consequence of a DNA break. Alternatively, it could stem from a
301 simple replication fork stalling resolved “using” other sequences of the genome through a replicative
302 process (MMBIR/FoSTeS) (Holland and Cleveland 2012; Liu et al. 2011). Thus, what we observe
303 through genome sequencing can either be a DNA breakpoint or a DNA join point. The fact that none of
304 the eight junctions resolved for individual 10 could be inferred as replicative questions our ability to
305 identify a repairing mechanism and differentiate breakpoints from join points. In this “grey zone”, a
306 shattering event seems likely but junction sequences present distinct signs from the classical non-
307 homologous end joining of *chromothripsis* (Kloosterman et al. 2011; Chiang et al. 2012).

308 Second, we describe a new feature of complex rearrangements that we name “repairing hubs”, where a
309 single locus is “used” several times within the derivative chromosome. We identified five hubs in 3 out
310 of 20 cases; they have no common feature except for G-band positive staining. It will be of particular
311 interest to collect additional observations and further investigate the genomic features shared by these
312 loci. Slamova *et al.* pointed that extremely short sequences can be handled, and here even spread
313 throughout a chromosomal rearrangement, challenging the 50 bp definition of a structural variant
314 (Slamova et al. 2018; Sudmant et al. 2015).

315 Third, whereas *chromothripsis* has often been used to designate all kinds of *chromoanagenesis* (Redin
316 et al. 2017; Collins et al. 2019; Pellestor 2014; Fukami et al. 2017) our identification of only a few clear
317 cut *chromothripsis* cases in our constitutional cohort suggests that such oversimplification should be
318 avoided.

319 Apart from variables affected by selective pressure we could not distinguish constitutional
320 *chromoanagenesis* from cancer cases. This suggests that the shattering scenario is comparable for
321 constitutional and tumoral cases. However, while the influence of each variable goes in the same
322 direction for both cancer and constitutional cases, the observed driving forces are still of different
323 magnitude. We hypothesize that this could result from additional complexity delineating possible
324 subgroups. Drier *et al.* have previously shown that cancer rearrangements' breakpoints have a bimodal
325 distribution, some cancers having breakpoints in early-replicating, highly expressed, GC-rich chromatin
326 while others had their breakpoints preferentially positioned in late-replicating, low transcribed, GC-poor
327 chromatin (Drier et al. 2013). Focusing on CCRs, breakpoints of sarcoma samples are enriched in early-
328 replicating chromatin (Anderson et al. 2018) in contrast to our finding from pooling all cancers together.

329 By opposition to simple rearrangements, complex rearrangements' breakpoints appear independent from
330 LADs. First, this is in favor of the micronucleus hypothesis as it was shown that LADs are lost in a
331 majority of micronuclei (Hatch et al. 2013). Second, this draws a demarcation line in the supposed
332 continuum between simple and complex rearrangements emphasizing differences in underlying
333 mechanisms. This LAD depletion is counter-intuitive as, first, the peripheral chromatin was used to be
334 considered as a "body-guard", absorbing mutagens to the inner chromatin (Hsu 1975) and second,
335 selective pressure should contribute to keep structural variants towards the nuclear periphery as it is both
336 gene-poor and less transcriptionally active.

337 Simple rearrangements are formed after the mobilization and incorrect pairing of two DNA double
338 strand breaks. This mobilization is restricted to a limited volume so that the pairs formed belong to
339 neighboring chromosomal territories (Soutoglou et al. 2007). Being peripheral a DNA break in a LAD
340 has, by definition, a reduced number of potential partners. DNA breaks in LADs are not mobilized
341 towards nuclear positions more favorable to homologous recombination (Lemaître et al. 2014). Their
342 mobilization capacity might even be reduced by their interaction with lamina, once again reducing their
343 ability to pair with another genomic locus. Overall, we suspect that DNA breaks occurring in LADs are
344 as frequent as elsewhere in the genome but have more chance to be repaired to their native partner than
345 to create a rearrangement compared to "central chromatin". These events could only be identified at
346 sequence resolution with repairing sequence scars.

347 In our analysis, late replication timing appears as a key risk factor for *chromoanagenesis* and
348 chromosomal breakage more broadly as it is for other types of genomic variation. It was first associated
349 with important variations in SNP-density across evolution and cancer development with 2-fold and 6-
350 fold enrichment for transitions and transversions mutations respectively (Koren et al. 2012), but profiles
351 are similar in germline and cancer cells (Liu et al. 2013). One possible explanation is that error-prone
352 repairing mechanisms are more active in late replication (Polymerase θ) while early-replicating gene-
353 dense chromatin benefits from high fidelity transcription-coupled nucleotide excision repair (Waters

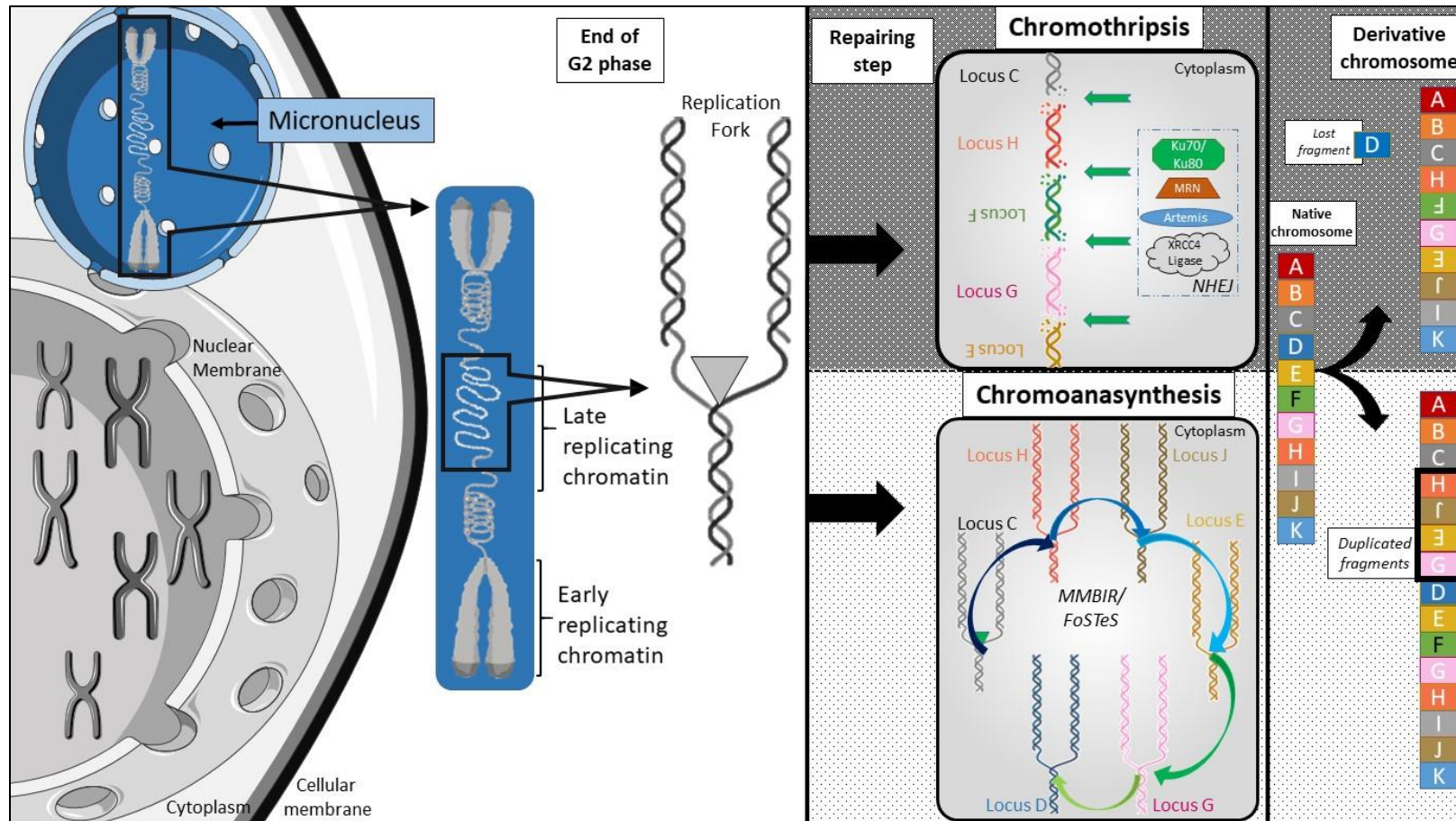
354 and Walker 2006). Interestingly, polymerase θ had previously been considered to explain the features
355 observed in a *chromoanagenesis* case (Masset et al. 2016). CNV distribution was also reported to be
356 influenced by replication timing, gains and losses being enriched in early and late regions respectively
357 in cancer cells, with breakpoints being part of the same replicating domain. However, to date, there is
358 no robust explanation for such findings (De and Michor 2011). Finally, if partners of balanced
359 rearrangements preferentially replicate at the same time (Ryba et al. 2010) preference towards late
360 replicating regions had never been reported before.

361 Observation of a breakpoint enrichment in late replicating chromatin supports the “Prematurely
362 Condensed Chromosome” hypothesis for chromoanagenesis. Indeed, our scenario for
363 *chromoanagenesis*, presented in Figure 6, is that after one (or a few) chromosome(s) gets trapped in a
364 micronucleus it can only replicate more slowly than the main nucleus. While entering the next cell
365 division, replication is not completed in the micronucleus and the entrapped material has to start
366 condensation prematurely. This generates a stress on replication forks still present in the physiologically
367 late replicating regions of this delayed chromosome. The forks would either collapse and provoke DNA
368 breakage or stall and start FoSTeS cycles. This would explain the co-occurrence of different repairing
369 mechanisms observed in a same rearrangement. The rearrangement puzzle will be reassembled after
370 reintegration within the main nucleus, opening the possibility for FoSTeS loops to use other
371 chromosomes.

372 Finally, the low proportion of rearrangements affecting two chromosomes and the importance of
373 replication-timing is not in favor of a telomere crisis scenario that would involve a dicentric
374 chromosome. Indeed, this second hypothesis can explain clustered breakpoints on a maximum of two
375 chromosomes but requires deficiencies of important cell cycle regulators (*RB*, *TP53*) that are relatively
376 unlikely during gametogenesis considering constitutional cases. Instead, several rearrangements present
377 a single junction between a shattered chromosome and another one suggesting the preexistence of a
378 balanced translocation. While no constitutional CCR has been described in offspring of simple balanced
379 rearrangement carrier such event could happen in the same gametogenesis and favor entrapment of one
380 chromosome of the tetravalent in a micronucleus.

381 Overall, we highlight the importance of precise characterization of CCRs as we identified (i)
382 rearrangement “hubs” that had not been described previously and (ii) a “grey zone” between
383 *chromothripsis* and *chromoanasyntesis* challenging actual definitions. By extensive annotation of the
384 largest cohort of constitutional CCRs and use of publicly available data, we identify replication timing
385 as a new element driving chromosomal breakpoint probability. Our *in vivo* dataset supports the
386 implication of premature condensation of genetic material segregated in a micronucleus in
387 *chromoanagenesis*. Finally, we identify a statistically significant depletion of simple rearrangements’
388 breakpoints within LADs. This strongly suggests that *chromoanagenesis* and simple rearrangements are

389 not part of a continuous spectrum and opens new perspectives towards chromosomal rearrangement
390 understanding.



391

392 **Figure 6:** Schematic summary of our scenario to explain chromoanagenesis breakpoint distribution. After missegregation within a micronucleus, DNA
 393 replication is delayed. While other chromosomes are ready for cell division, still active replication forks in constitutively late replicating regions are subject to
 394 premature condensation and will either collapse (and create DNA double strand breaks) or stall (with possible FoSTeS events) so that both mechanisms can be
 395 observed in a single rearrangement.

396 **Online Methods**

397 *Individuals*

398 To recruit CCRs we first initiated a French-national call for collaboration to gather all cases analyzed
399 using cytogenetic microarray and having a minimum of 3 non-polymorphic, non-recurrent copy number
400 variants on a single chromosome (potential pair of chromosomes). Individual 10 was recruited to the
401 ethically approved study *Genetic basis of craniofacial malformations* (London - Riverside REC
402 09/H0706/20) and included after genome sequencing results fitting the entry criterion. In addition, we
403 included all balanced rearrangements characterized in our laboratory as part of our routine activity
404 having more than 10 breakpoints.

405 All individuals or their parents gave written informed consent for this study, which was conducted with
406 respect to the recommendations of the Helsinki Declaration. Noteworthy, balanced rearrangements had
407 previously been published as part of our clinical validation study (Schluth-Bolard et al. 2019). Individual
408 20's description has been recently reported (Ader et al. 2019).

409

410 *Genome sequencing*

411 All cases, but case 10, were genome sequenced using our validated approach (Schluth-Bolard et al.
412 2019). After PCR-free library preparation (Nano, Illumina Inc., San Diego, CA, USA), 2x101 bp paired-
413 end sequencing was performed on a 300 cycles High Output FlowCell for NextSeq500 instrument
414 (Illumina, Inc.). Additionally, samples from individuals I-19, 2, 3, 6, 9, 10 and 19 were 2x151 bp paired-
415 end sequenced on a HiSeq X5 or a HiSeq4000 instrument (Illumina Inc.). Breakpoints were detected
416 using BreakDancer v1.4.5 (Chen et al. 2009) and ERDS v1.1 software was used for CNV calling (Zhu
417 et al. 2012). For all calls, supporting alignment data was systematically inspected using the IGV v2.4
418 software with soft-clipped reads set as visible (Robinson et al. 2011). At each breakpoint, split-reads
419 were extracted manually to align the soft-clipped sequence using BLAT and obtain fragment junction
420 sequence directly (Kent 2002). For all junctions, the process was repeated in both directions of the
421 junction.

422

423 *Breakpoint and junction annotation*

424 Considering extreme complexity, all breakpoints within a kilobase were considered as one for further
425 analyses. We annotated each breakpoint with publicly available datasets using Svagga
426 (<https://gitlab.inria.fr/NGS/svagga>). RefSeq genes coordinates were extracted from the UCSC Genome
427 Browser; pLi scores were extracted from the gnomAD dataset (Lek et al. 2016). Annotation was also
428 made across lamina associated domains (LADs) (Guelen et al. 2008); topologically associated domains

429 (TADs) (Rao et al. 2014); repeated elements (RepeatMasker); chromosomal cytoband and Giemsa
430 staining (Furey and Haussler 2003); chromosomal common fragile sites (Fungtammasan et al. 2012);
431 replication origins (Massip et al. 2019); A/B compartments (Fortin and Hansen 2015) and replication
432 timing (Dixon et al. 2018). For the latter, only constitutive early and late domains were considered. Both
433 strands of a junction had to be inferred with the same mechanism following the same criteria as in
434 (Schluth-Bolard et al. 2019) to be considered.

435 Based on existing literature, we established a decisional algorithm to infer the *chromoanagenesis*
436 subgroup for each rearrangement as a whole (Korbel and Campbell 2013; Liu et al. 2011; Baca et al.
437 2013; Holland and Cleveland 2012). First, rearrangements presenting several copy number gains were
438 classified as *chromoanasyntesis* as the concept of *chromothripsis* does not foresee genomic gains.
439 Second, the rearrangement was considered as *chromothripsis* or *chromoanasyntesis* depending on the
440 dominant repairing mechanism (NHEJ for *chromothripsis*; MMBIR/FoSTeS or clear FoSTeS events
441 with several shifts for *chromoanasyntesis*) with a minimum of 8 inferred junctions to qualify the
442 rearrangement.

443

444 *External cases*

445 To increase statistical power we extended our analysis to other complex rearrangements and used
446 ChromothripsisDB (Yang et al. 2015). We downloaded the complete database and kept for analysis
447 samples that had been genome sequenced using paired-end sequencing and for which breakpoint
448 positions were available. We filtered out every call that was (i) not on chromosomes involved in the
449 rearrangement as described by initial authors; (ii) of uncertain position as they had ‘start’ and ‘stop’
450 positions distant more than 1 kb. We finally extended our cohort of constitutional CCRs with a few more
451 recent constitutional cases (Redin et al. 2017) absent from ChromothripsisDB for a total of 20 external
452 cases (350 breakpoints). Before pooling all constitutional CCRs, we confirmed that our dataset was
453 similar to published cases when considering breakpoint distribution. Compared to simulation data, both
454 distributions were either not significant or in the same direction considering individual genomic features
455 (Supplementary Figure S2).

456 Constitutional simple rearrangements published in (Schluth-Bolard et al. 2019) and (Redin et al. 2017)
457 were pooled in a third group distinct from (a) constitutional and (b) cancer CCRs.

458

459 *Breakpoint simulation*

460 We performed *in silico* breakpoint simulation to create a null hypothesis of breakpoint distribution. We
461 used R software for two different Monte Carlo simulations. In both cases, a random number of

462 breakpoints between 10 (minimal number of breakpoints we used to consider a *chromoanagenesis*) and
463 1000 were picked for 1000 simulations. In the first “free” simulation, each breakpoint could occur in
464 every possible position of the genome. In the “constrained” model, to mimic the observed “geographical
465 clustering” of breakpoints, a random percentage between 50 and 100% of breakpoints had to occur in
466 the same chromosome while the remaining ones were free. As we know that some genomic regions are
467 poorly analyzed for structural variants, we used the “DangerTrack” bed file (Dolgalev et al. 2017) to
468 exclude all breakpoints simulated in these regions and be comparable with the observed dataset. A single
469 breakpoint of our in-house dataset overlaps the DangerTrack. Overall, 399,346 and 399,915 breakpoints
470 were simulated and used for analyses for the “free” and “constrained” models respectively.

471

472 *Statistical analysis*

473 All statistical analyses were performed using R software v.3.5.2. Univariate analysis of each predictor
474 and the logistic regression model building was performed using the “multinom” function of the “nnet”
475 R package (Ripley and Venables 2016) as the Y outcome is multinomial. The Wald test was used for p-
476 value calculation. Considering the number of variables and hypotheses tested we considered $p < 0.01$
477 for statistical significance.

478

479 **Data access**

480 All sequencing data generated were obtained from patients in a diagnostic setting and can be shared
481 individually upon reasonable request for patients that had consented data sharing.

482

483 **Acknowledgments**

484 The authors would like to thank members of the AchroPuce network for their thorough collaboration.
485 NC acknowledges « Fondation Thérèse et René Planiol », Réseau Régional de Rééducation et de
486 Réadaptation Pédiatrique en Rhône Alpes, IDEX Lyon and Hospices Civils de Lyon. GG is recipient of
487 a Pro-Women Scholarship from the Faculty of Biology and Medicine, University of Lausanne. This
488 work was supported by grants from the Swiss National Science Foundation (31003A_182632) and
489 Horizon2020 Twinning project ePerMed (692145) to AR, NHR Oxford Biomedical Research Centre
490 based at Oxford University Hospitals NHS Trust and University of Oxford to AOMW, the Fondation
491 Maladies Rares (WGS20140103) to DS, from the French Ministry of Health (DGOS) and the French
492 National Agency for Research (ANR) (PRTS 2013: PRTSN1300001N) to CSB. Patient 10’s whole
493 genome sequencing was funded by independent research commissioned by the Health Innovation
494 Challenge Fund (JCT), a parallel funding partnership between Wellcome (WT 100127) and the UK

495 Department of Health (R6-388). The views expressed in this publication are those of the authors and not
496 necessarily those of Wellcome, NHS, the NIHR or the Department of Health.

497

498 **Author Contributions**

499 Nicolas Chatron, Pierre-Antoine Rollat Farnier, Flavie Diguët, Tony Yammine, Kevin Uguen, Zohra-
500 Lydia Bellil, Julia Lauer Zillhardt, Arthur Sorlin performed the experiments and analyzed the data.
501 Nicolas Chatron, Giuliana Giannuzzi, Claire Bardel, and Eleonora Porcu performed statistical analyses.
502 Flavie Ader, Alexandra Afenjar, Joris Andrieux, Eduardo Calpena, Sandra-Chantot-Bastaraud, Patrick
503 Callier, Nora Chelloug, Emilie Chopin, Marie-Pierre Cordier, Christèle Dubourg, Laurence Faivre,
504 Françoise Girard, Solveig Heide, Yvan Herenger, Sylvie Jaillard, Boris Keren, Samantha J. L. Knight,
505 James Lespinasse, Laurence Lohmann, Nathalie Marle, Reza Maroofian, Alice Masurel-Paulet, Michèle
506 Mathieu-Dramard, Corinne Metay, Alistair T. Pagnamenta, Marie-France Portnoï, Fabienne Prieur,
507 Marlène Rio, Jean-Pierre Siffroi, Stéphanie Valence, Jenny C. Taylor, Andrew O. M. Wilkie, Patrick
508 Ederly enrolled patients in the study, collected the clinical data, and analyzed the data. N. Chatron,
509 Giuliana Giannuzzi, Caroline Schluth-Bolard and Alexandre Reymond drafted the manuscript. Damien
510 Sanlaville, Andrew O. M. Wilkie, helped in revising and editing the manuscript. All the authors
511 approved the manuscript submission.

512

513 **Disclosure declaration**

514 The authors have no conflict of interest to declare.

515 **References**

516 Ader F, Heide S, Marzin P, Afenjar A, Diguët F, Chantot Bastaraud S, Rollat-Farnier P-A, Sanlaville
517 D, Portnoï M-F, Siffroi J-P, et al. 2019. A 14q distal chromoanagenesis elucidated by whole
518 genome sequencing. *Eur J Med Genet* 103776.

519 <https://linkinghub.elsevier.com/retrieve/pii/S1769721218305949> (Accessed September 30,
520 2019).

521 Anderson ND, Borja R de, Young MD, Fuligni F, Rosic A, Roberts ND, Hajjar S, Layeghifard M,
522 Novokmet A, Kowalski PE, et al. 2018. Rearrangement bursts generate canonical gene fusions in
523 bone and soft tissue tumors. *Science* **361**.

524 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6176908/> (Accessed September 5, 2019).

525 Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N,
526 MacDonald TY, Ghandi M, et al. 2013. Punctuated evolution of prostate cancer genomes. *Cell*
527 **153**: 666–77. <https://linkinghub.elsevier.com/retrieve/pii/S0092867413003437> (Accessed

- 528 December 15, 2015).
- 529 Caspersson T, Zech L, Johansson C. 1970. Differential binding of alkylating fluorochromes in human
530 chromosomes. *Exp Cell Res* **60**: 315–9. <http://www.ncbi.nlm.nih.gov/pubmed/5422961>
531 (Accessed June 22, 2019).
- 532 Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC,
533 Zhang Q, Locke DP, et al. 2009. BreakDancer: an algorithm for high-resolution mapping of
534 genomic structural variation. *Nat Methods* **6**: 677–81. <http://dx.doi.org/10.1038/nmeth.1363>
535 (Accessed December 15, 2015).
- 536 Chiang C, Jacobsen JC, Ernst C, Hanscom C, Heilbut A, Blumenthal I, Mills RE, Kirby A, Lindgren
537 AM, Rudiger SR, et al. 2012. Complex reorganization and predominant non-homologous repair
538 following chromosomal breakage in karyotypically balanced germline rearrangements and
539 transgenic integration. *Nat Genet* **44**: 390–7, S1.
540 <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3340016&tool=pmcentrez&renderty>
541 [pe=abstract](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3340016&tool=pmcentrez&renderty) (Accessed January 2, 2016).
- 542 Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Khera A V., Francioli LC, Gauthier LD,
543 Wang H, Watts NA, et al. 2019. An open resource of structural variation for medical and
544 population genetics. *bioRxiv* 578674. <https://www.biorxiv.org/content/10.1101/578674v1>
545 (Accessed June 28, 2019).
- 546 Cooper GM, Coe BP, Girirajan S, Rosenfeld J a, Vu TH, Baker C, Williams C, Stalker H, Hamid R,
547 Hannig V, et al. 2011. A copy number variation morbidity map of developmental delay. *Nat*
548 *Genet* **43**: 838–46.
549 <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3171215&tool=pmcentrez&renderty>
550 [pe=abstract](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3171215&tool=pmcentrez&renderty) (Accessed July 23, 2014).
- 551 Crasta K, Ganem NJ, Dagher R, Lantermann AB, Ivanova E V, Pan Y, Nezi L, Protopopov A,
552 Chowdhury D, Pellman D. 2012. DNA breaks and chromosome pulverization from errors in
553 mitosis. *Nature* **482**: 53–8. <http://www.nature.com/doi/10.1038/nature10802> (Accessed
554 June 24, 2015).
- 555 De S, Michor F. 2011. DNA replication timing and long-range DNA interactions predict mutational
556 landscapes of cancer genomes. *Nat Biotechnol* **29**: 1103–1108.
557 <http://www.ncbi.nlm.nih.gov/pubmed/22101487> (Accessed July 30, 2019).
- 558 Dixon JR, Xu J, Dileep V, Zhan Y, Song F, Le VT, Yardımcı GG, Chakraborty A, Bann D V., Wang
559 Y, et al. 2018. Integrative detection and analysis of structural variation in cancer genomes. *Nat*
560 *Genet* **50**: 1388–1398. <http://www.ncbi.nlm.nih.gov/pubmed/30202056> (Accessed July 17,

- 561 2019).
- 562 Dolgalev I, Sedlazeck F, Busby B. 2017. DangerTrack: A scoring system to detect difficult-to-assess
563 regions. *F1000Research* **6**: 443. <http://www.ncbi.nlm.nih.gov/pubmed/28503299> (Accessed July
564 16, 2019).
- 565 Drier Y, Lawrence MS, Carter SL, Stewart C, Gabriel SB, Lander ES, Meyerson M, Beroukheim R,
566 Getz G. 2013. Somatic rearrangements across cancer reveal classes of samples with distinct
567 patterns of DNA breakage and rearrangement-induced hypermutability. *Genome Res* **23**: 228–35.
568 <http://www.ncbi.nlm.nih.gov/pubmed/23124520> (Accessed September 5, 2019).
- 569 Fortin J-P, Hansen KD. 2015. Reconstructing A/B compartments as revealed by Hi-C using long-
570 range correlations in epigenetic data. *Genome Biol* **16**: 180.
571 <http://www.ncbi.nlm.nih.gov/pubmed/26316348> (Accessed October 17, 2019).
- 572 Fukami M, Shima H, Suzuki E, Ogata T, Matsubara K, Kamimaki T. 2017. Catastrophic cellular
573 events leading to complex chromosomal rearrangements in the germline. *Clin Genet* **91**: 653–
574 660. <http://doi.wiley.com/10.1111/cge.12928> (Accessed October 2, 2019).
- 575 Fungtammasan A, Walsh E, Chiaromonte F, Eckert KA, Makova KD. 2012. A genome-wide analysis
576 of common fragile sites: What features determine chromosomal instability in the human
577 genome? *Genome Res* **22**: 993–1005. <http://www.ncbi.nlm.nih.gov/pubmed/22456607> (Accessed
578 July 17, 2019).
- 579 Furey TS, Haussler D. 2003. Integration of the cytogenetic map with the draft human genome
580 sequence. *Hum Mol Genet* **12**: 1037–1044. [https://academic.oup.com/hmg/article-
581 lookup/doi/10.1093/hmg/ddg113](https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddg113) (Accessed July 17, 2019).
- 582 Ganem NJ, Godinho SA, Pellman D. 2009. A mechanism linking extra centrosomes to chromosomal
583 instability. *Nature* **460**: 278–282. <http://www.ncbi.nlm.nih.gov/pubmed/19506557> (Accessed
584 July 22, 2019).
- 585 Guelen L, Pagie L, Brassat E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels
586 L, de Laat W, et al. 2008. Domain organization of human chromosomes revealed by mapping of
587 nuclear lamina interactions. *Nature* **453**: 948–951. <http://www.nature.com/articles/nature06947>
588 (Accessed July 17, 2019).
- 589 Hatch EM, Fischer AH, Deerinck TJ, Hetzer MW. 2013. Catastrophic Nuclear Envelope Collapse in
590 Cancer Cell Micronuclei. *Cell* **154**: 47–60. <http://www.ncbi.nlm.nih.gov/pubmed/23827674>
591 (Accessed September 4, 2019).
- 592 Holland AJ, Cleveland DW. 2012. Chromoanagenesis and cancer: mechanisms and consequences of
593 localized, complex chromosomal rearrangements. *Nat Med* **18**: 1630–8.

- 594 <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3616639&tool=pmcentrez&rendertype=abstract> (Accessed January 7, 2016).
- 595
- 596 Hsu TC. 1975. A possible function of constitutive heterochromatin: the bodyguard hypothesis.
- 597 *Genetics* **79 Suppl**: 137–50. <http://www.ncbi.nlm.nih.gov/pubmed/1150080> (Accessed July 30,
- 598 2019).
- 599 Kent WJ. 2002. BLAT---The BLAST-Like Alignment Tool. *Genome Res* **12**: 656–664.
- 600 <http://www.ncbi.nlm.nih.gov/pubmed/11932250> (Accessed July 12, 2018).
- 601 Kloosterman WP, Guryev V, van Roosmalen M, Duran KJ, de Bruijn E, Bakker SCM, Letteboer T,
- 602 van Nesselrooij B, Hochstenbach R, Poot M, et al. 2011. Chromothripsis as a mechanism driving
- 603 complex de novo structural rearrangements in the germline. *Hum Mol Genet* **20**: 1916–24.
- 604 <http://www.ncbi.nlm.nih.gov/pubmed/21349919> (Accessed January 7, 2016).
- 605 Korbel JO, Campbell PJ. 2013. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**:
- 606 1226–36. <http://www.ncbi.nlm.nih.gov/pubmed/23498933> (Accessed December 30, 2015).
- 607 Koren A, Polak P, Nemesh J, Michaelson JJ, Sebat J, Sunyaev SR, McCarroll SA. 2012. Differential
- 608 Relationship of DNA Replication Timing to Different Forms of Human Mutation and Variation.
- 609 *Am J Hum Genet* **91**: 1033–1040. <http://www.ncbi.nlm.nih.gov/pubmed/23176822> (Accessed
- 610 July 30, 2019).
- 611 Kürten S, Obe G. 1975. Premature chromosome condensation in the bone marrow of Chinese hamsters
- 612 after application of bleomycin in vivo. *Mutat Res Mol Mech Mutagen* **27**: 285–294.
- 613 <https://www.sciencedirect.com/science/article/pii/0027510775900895> (Accessed July 26, 2019).
- 614 Lejeune J, Gautier M, Turpin R. 1959. Etude des chromosomes somatiques de 9 enfants mongoliens.
- 615 *C R Hebd Seances Acad Sci* **248**: 1721–2. <http://www.ncbi.nlm.nih.gov/pubmed/13639368>
- 616 (Accessed May 21, 2015).
- 617 Lek M, Karczewski KJ, Minikel E V., Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware
- 618 JS, Hill AJ, Cummings BB, et al. 2016. Analysis of protein-coding genetic variation in 60,706
- 619 humans. *Nature* **536**: 285–291. <http://www.ncbi.nlm.nih.gov/pubmed/27535533> (Accessed July
- 620 24, 2018).
- 621 Lemaître C, Grabarz A, Tsouroula K, Andronov L, Furst A, Pankotai T, Heyer V, Rogier M, Attwood
- 622 KM, Kessler P, et al. 2014. Nuclear position dictates DNA repair pathway choice. *Genes Dev* **28**:
- 623 2450–2463. <http://www.ncbi.nlm.nih.gov/pubmed/25366693> (Accessed July 18, 2019).
- 624 Liu L, De S, Michor F. 2013. DNA replication timing and higher-order nuclear organization determine
- 625 single-nucleotide substitution patterns in cancer genomes. *Nat Commun* **4**: 1502.
- 626 <http://www.ncbi.nlm.nih.gov/pubmed/23422670> (Accessed July 30, 2019).

- 627 Liu P, Erez A, Nagamani SCS, Dhar SU, Kołodziejaska KE, Dharmadhikari A V, Cooper ML,
628 Wiszniewska J, Zhang F, Withers MA, et al. 2011. Chromosome catastrophes involve replication
629 mechanisms generating complex genomic rearrangements. *Cell* **146**: 889–903.
630 [http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3242451&tool=pmcentrez&renderty](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3242451&tool=pmcentrez&rendertype=abstract)
631 [pe=abstract](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3242451&tool=pmcentrez&rendertype=abstract) (Accessed January 7, 2016).
- 632 Maciejowski J, Li Y, Bosco N, Campbell PJ, de Lange T. 2015. Chromothripsis and Kataegis Induced
633 by Telomere Crisis. *Cell* **163**: 1641–1654. <http://www.ncbi.nlm.nih.gov/pubmed/26687355>
634 (Accessed December 18, 2015).
- 635 Mardin BR, Drainas AP, Waszak SM, Weischenfeldt J, Isokane M, Stütz AM, Raeder B,
636 Efthymiopoulos T, Buccitelli C, Segura-Wang M, et al. 2015. A cell-based model system links
637 chromothripsis with hyperploidy. *Mol Syst Biol* **11**: 828.
638 [http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4592670&tool=pmcentrez&renderty](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4592670&tool=pmcentrez&rendertype=abstract)
639 [pe=abstract](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4592670&tool=pmcentrez&rendertype=abstract) (Accessed October 6, 2015).
- 640 Masset H, Hestand MS, Van Esch H, Kleinfinger P, Plaisancié J, Afenjar A, Mollignier R, Schluth-
641 Bolard C, Sanlaville D, Vermeesch JR. 2016. A Distinct Class of Chromoanagenesis Events
642 Characterized by Focal Copy Number Gains. *Hum Mutat*.
- 643 Massip F, Laurent M, Brossas C, Fernández-Justel JM, Gómez M, Prioleau M-N, Duret L, Picard F.
644 2019. Evolution of replication origins in vertebrate genomes: rapid turnover despite selective
645 constraints. *Nucleic Acids Res* **47**: 5114–5125.
646 <https://academic.oup.com/nar/article/47/10/5114/5420529> (Accessed September 4, 2019).
- 647 Obe G, Beek B. 1975. The human leucocyte test system. VII. Further investigations concerning
648 micronucleus-derived premature chromosome condensation. *Humangenetik* **30**: 143–54.
649 <http://www.ncbi.nlm.nih.gov/pubmed/1193601> (Accessed July 26, 2019).
- 650 Pampalona J, Roscioli E, Silkworth WT, Bowden B, Genescà A, Tusell L, Cimini D. 2016.
651 Chromosome Bridges Maintain Kinetochores-Microtubule Attachment throughout Mitosis and
652 Rarely Break during Anaphase. ed. Y. Wang. *PLoS One* **11**: e0147420.
653 <https://dx.plos.org/10.1371/journal.pone.0147420> (Accessed July 15, 2019).
- 654 Pellestor F. 2014. Chromothripsis: how does such a catastrophic event impact human reproduction?
655 *Hum Reprod* **29**: 388–393. <http://www.ncbi.nlm.nih.gov/pubmed/24452388> (Accessed October
656 2, 2019).
- 657 Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol
658 I, Omer AD, Lander ES, et al. 2014. A 3D Map of the Human Genome at Kilobase Resolution
659 Reveals Principles of Chromatin Looping. *Cell* **159**: 1665–1680.

- 660 <http://www.ncbi.nlm.nih.gov/pubmed/25497547> (Accessed July 17, 2019).
- 661 Redin C, Brand H, Collins RL, Kammin T, Mitchell E, Hodge JC, Hanscom C, Pillalamarri V, Seabra
662 CM, Abbott M-A, et al. 2017. The genomic landscape of balanced cytogenetic abnormalities
663 associated with human congenital anomalies. *Nat Genet* **49**: 36–45.
664 <http://www.ncbi.nlm.nih.gov/pubmed/27841880> (Accessed August 6, 2018).
- 665 Ripley B, Venables W. 2016. R Package ‘nnet.’ <https://cran.r-project.org/web/packages/nnet/>.
- 666 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011.
667 Integrative genomics viewer. *Nat Biotechnol* **29**: 24–6. <http://dx.doi.org/10.1038/nbt.1754>
668 (Accessed November 19, 2014).
- 669 Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, Schulz TC, Robins AJ, Dalton S, Gilbert DM.
670 2010. Evolutionarily conserved replication timing profiles predict long-range chromatin
671 interactions and distinguish closely related cell types. *Genome Res* **20**: 761–770.
672 <http://www.ncbi.nlm.nih.gov/pubmed/20430782> (Accessed July 31, 2019).
- 673 Schluth-Bolard C, Diguët F, Chatron N, Rollat-Farnier P-A, Bardel C, Afenjar A, Amblard F, Amiel J,
674 Blesson S, Callier P, et al. 2019. Whole genome paired-end sequencing elucidates functional and
675 phenotypic consequences of balanced chromosomal rearrangement in patients with
676 developmental disorders. *J Med Genet* jmedgenet-2018-105778.
677 <http://www.ncbi.nlm.nih.gov/pubmed/30923172> (Accessed May 10, 2019).
- 678 Slamova Z, Nazaryan-Petersen L, Mehrjouy MM, Drabova J, Hancarova M, Marikova T, Novotna D,
679 Vlckova M, Vlckova Z, Bak M, et al. 2018. Very short DNA segments can be detected and
680 handled by the repair machinery during germline chromothriptic chromosome reassembly. *Hum*
681 *Mutat* **39**: 709–716. <http://www.ncbi.nlm.nih.gov/pubmed/29405539> (Accessed November 1,
682 2019).
- 683 Soutoglou E, Dorn JF, Sengupta K, Jasin M, Nussenzweig A, Ried T, Danuser G, Misteli T. 2007.
684 Positional stability of single double-strand breaks in mammalian cells. *Nat Cell Biol* **9**: 675–682.
685 <http://www.ncbi.nlm.nih.gov/pubmed/17486118> (Accessed October 10, 2019).
- 686 Stankiewicz P, Lupski JR. 2002. Genome architecture, rearrangements and genomic disorders. *Trends*
687 *Genet* **18**: 74–82. <http://www.ncbi.nlm.nih.gov/pubmed/11818139> (Accessed November 1,
688 2019).
- 689 Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D,
690 Stebbings LA, et al. 2011. Massive Genomic Rearrangement Acquired in a Single Catastrophic
691 Event during Cancer Development. *Cell* **144**: 27–40.
692 <http://www.ncbi.nlm.nih.gov/pubmed/21215367> (Accessed December 13, 2018).

- 693 Storchová Z, Kloosterman WP. 2016. The genomic characteristics and cellular origin of
694 chromothripsis. *Curr Opin Cell Biol* **40**: 106–113.
695 <http://www.ncbi.nlm.nih.gov/pubmed/27023493> (Accessed May 22, 2016).
- 696 Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G,
697 Hsi-Yang Fritz M, et al. 2015. An integrated map of structural variation in 2,504 human
698 genomes. *Nature* **526**: 75–81. <http://www.nature.com/articles/nature15394> (Accessed July 4,
699 2019).
- 700 Terzoudi GI, Karakosta M, Pantelias A, Hatzi VI, Karachristou I, Pantelias G. 2015. Stress induced by
701 premature chromatin condensation triggers chromosome shattering and chromothripsis at DNA
702 sites still replicating in micronuclei or multinucleate cells when primary nuclei enter mitosis.
703 *Mutat Res Genet Toxicol Environ Mutagen* **793**: 185–98.
704 <https://linkinghub.elsevier.com/retrieve/pii/S1383571815002041> (Accessed December 18, 2018).
- 705 Titen SWA, Golic KG. 2008. Telomere loss provokes multiple pathways to apoptosis and produces
706 genomic instability in *Drosophila melanogaster*. *Genetics* **180**: 1821–32.
707 <http://www.genetics.org/lookup/doi/10.1534/genetics.108.093625> (Accessed July 15, 2019).
- 708 Tubio JMC, Estivill X. 2011. When catastrophe strikes a cell. *Nature* **470**: 476–477.
709 <http://www.ncbi.nlm.nih.gov/pubmed/21350479> (Accessed September 11, 2019).
- 710 Waters LS, Walker GC. 2006. The critical mutagenic translesion DNA polymerase Rev1 is highly
711 expressed during G(2)/M phase rather than S phase. *Proc Natl Acad Sci U S A* **103**: 8971–6.
712 <http://www.ncbi.nlm.nih.gov/pubmed/16751278> (Accessed July 30, 2019).
- 713 Yang J, Deng G, Cai H. 2015. ChromothripsisDB: a curated database of chromothripsis.
714 *Bioinformatics*. <http://www.ncbi.nlm.nih.gov/pubmed/26722116> (Accessed January 2, 2016).
- 715 Zhu M, Need AC, Han Y, Ge D, Maia JM, Zhu Q, Heinzen EL, Cirulli ET, Pelak K, He M, et al.
716 2012. Using ERDS to infer copy-number variants in high-coverage genomes. *Am J Hum Genet*
717 **91**: 408–21.
718 [http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3511991&tool=pmcentrez&renderty
719 pe=abstract](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3511991&tool=pmcentrez&rendertype=abstract) (Accessed January 8, 2016).

720

721 **Supplementary Table S1:** List of annotated breakpoints detected in our cohort of complex
722 chromosomal rearrangements (CCRs), in external cases of CCRs (cancer and constitutional), simple
723 rearrangements and simulations. For replication timing, CE, CL and S stand for constitutive early,
724 constitutive late and switching respectively. Only the two first ones were used for statistical analyses.

725 **Supplementary File S1:** Individual phenotypic and cytogenetic results before genome sequencing. List
726 of junctions resolved in both forward and reverse orientations with the same repairing mechanism
727 obtained from genome sequencing data with inferred repairing mechanism.

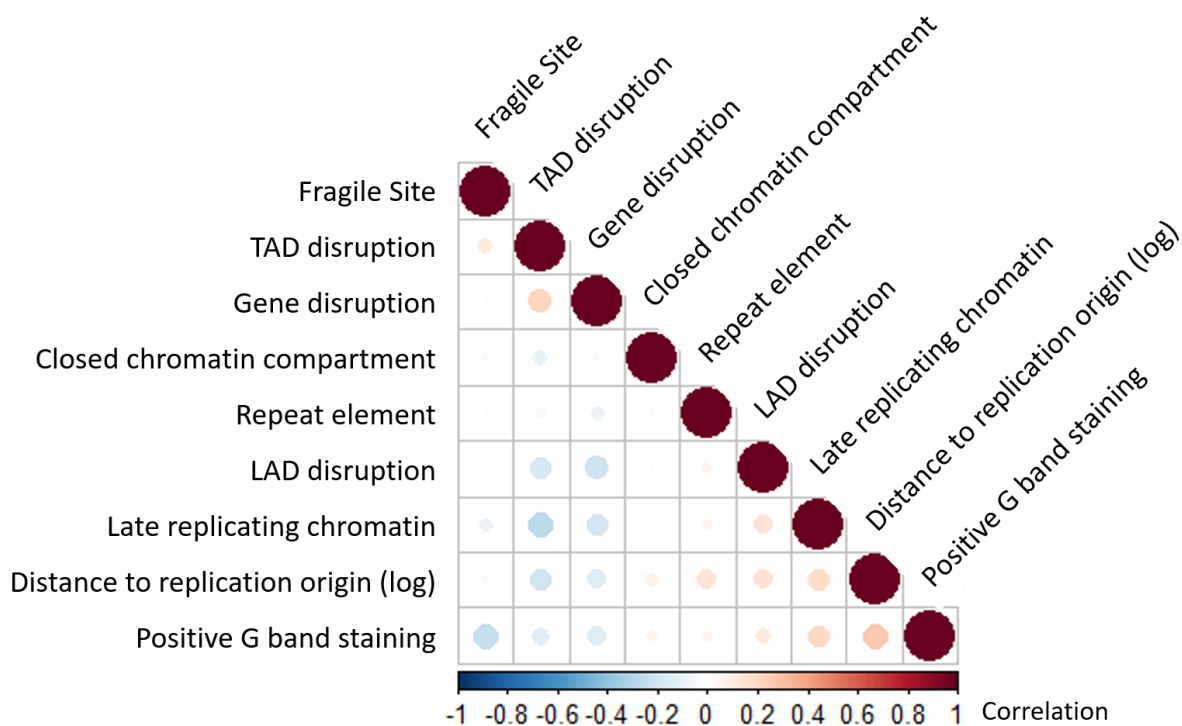
Variable	Group	Free simulation			Constrained simulation		
		Relative Risk Ratio	CI 95%	p value	Relative Risk Ratio	CI 95%	p value
Gene disruption	Cancer CCRs	1,057	[1,021-1,094]	0,002	1,07	[1,034-1,108]	1x10 ⁻⁴
	Constitutional CCRs	0,807	[0,712-0,915]	0,001	0,818	[0,722-0,927]	0,002
	Simple Rearrangements	1	[0,833-1,2]	1	1,013	[0,844-1,216]	0,891
pLI	Cancer CCRs	0,929	[0,873-0,988]	0,02	0,92	[0,864-0,979]	0,008
	Constitutional CCRs	0,751	[0,589-0,958]	0,021	0,746	[0,585-0,952]	0,019
	Simple Rearrangements	2,276	[1,645-3,151]	7x10 ⁻⁷	2,267	[1,637-3,137]	8x10 ⁻⁷
TAD disruption	Cancer CCRs	1,12	[1,081-1,16]	3x10 ⁻¹⁰	1,113	[1,075-1,153]	2x10 ⁻⁹
	Constitutional CCRs	0,872	[0,771-0,985]	0,028	0,867	[0,767-0,98]	0,022
	Simple Rearrangements	1,111	[0,924-1,336]	0,261	1,105	[0,919-1,328]	0,29

728

729 **Supplementary Table S2:** Univariate analysis of breakpoint distribution across genomic features related to gene dysregulation. While constitutional CCRs
730 affect a reduced proportion of genes and TADs compared to simulations, simple rearrangements are more likely to affect haplosensitive genes.

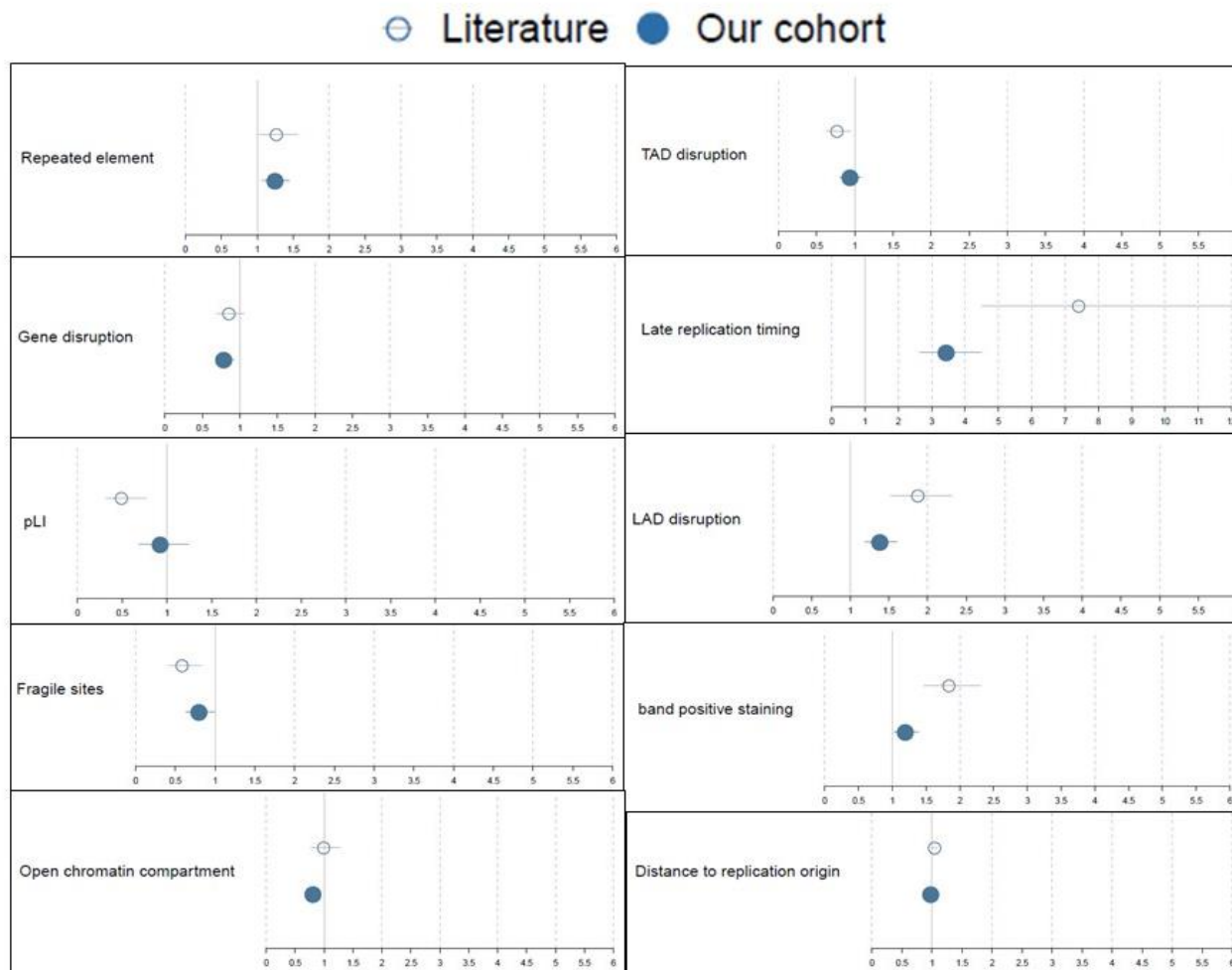
Variable	Group	Free simulation			Constrained simulation		
		Relative Risk Ratio	CI 95%	p,value	Relative Risk Ratio	CI 95%	p,value
Gene disruption	Cancer CCRs	0,998	[0,931-1,069]	0,955	0,998	[0,931-1,069]	0,949
	Constitutional CCRs	0,606	[0,455-0,808]	0,001	0,605	[0,454-0,806]	0,001
	Simple Rearrangements	1,055	[0,704-1,581]	0,795	1,049	[0,7-1,572]	0,818
TAD disruption	Cancer CCRs	0,950	[0,878-1,027]	0,195	0,933	[0,863-1,009]	0,082
	Constitutional CCRs	0,882	[0,652-1,193]	0,417	0,860	[0,636-1,162]	0,326
	Simple Rearrangements	1,421	[0,877-2,303]	0,154	1,386	[0,856-2,242]	0,184
Repeated element	Cancer CCRs	1,135	[1,061-1,214]	2×10^{-4}	1,126	[1,052-1,204]	0,001
	Constitutional CCRs	1,014	[0,775-1,327]	0,919	1,004	[0,767-1,314]	0,977
	Simple Rearrangements	1,285	[0,865-1,908]	0,214	1,280	[0,862-1,901]	0,222
G band positive staining	Cancer CCRs	1,103	[1,026-1,186]	0,008	1,064	[0,99-1,143]	0,094
	Constitutional CCRs	0,860	[0,639-1,156]	0,317	0,826	[0,614-1,111]	0,207
	Simple Rearrangements	0,769	[0,5-1,183]	0,232	0,743	[0,483-1,142]	0,176
Open chromatin compartment	Cancer CCRs	1,009	[0,931-1,094]	0,824	1,074	[0,992-1,164]	0,078
	Constitutional CCRs	1,499	[1,087-2,066]	0,014	1,553	[1,129-2,137]	0,007
	Simple Rearrangements	0,618	[0,394-0,969]	0,036	0,665	[0,426-1,037]	0,072
Distance to replication origin	Cancer CCRs	0,982	[0,969-0,995]	0,008	0,988	[0,975-1,001]	0,075
	Constitutional CCRs	1,036	[0,971-1,105]	0,280	1,042	[0,976-1,111]	0,218
	Simple Rearrangements	0,977	[0,904-1,055]	0,547	0,982	[0,909-1,061]	0,647
LAD disruption	Cancer CCRs	0,960	[0,88-1,046]	0,352	0,987	[0,905-1,077]	0,773
	Constitutional CCRs	0,932	[0,678-1,281]	0,666	0,935	[0,679-1,289]	0,683
	Simple Rearrangements	0,454	[0,26-0,793]	0,006	0,466	[0,266-0,816]	0,008
Late replication timing	Cancer CCRs	1,305	[1,187-1,434]	3×10^{-8}	1,431	[1,303-1,572]	7×10^{-14}
	Constitutional CCRs	3,419	[2,376-4,92]	3×10^{-11}	3,713	[2,583-5,338]	1×10^{-12}
	Simple Rearrangements	1,930	[1,127-3,306]	0,017	2,149	[1,259-3,668]	0,005

731 **Supplementary Table S3:** Multivariate analysis of breakpoint distribution across different types of chromosomal rearrangements. The logistic regression model
732 presented is the addition of each covariable that has the lowest Akaike information criterion tested (AIC = 36663,63 and 36558.53 for constrained and free
733 simulation as reference group respectively. Late replication-timing appears as the only significant variable for both cancer and constitutional CCRs. Simple
734 rearrangements diverge from simulation with a depletion in LADs.



735

736 **Supplementary Figure S1:** Correlation matrix (Pearson) of the analyzed variables based on our dataset
737 of constitutional CCRs.



738

739 **Supplementary Figure S2:** Univariate analysis of breakpoint distribution for our cohort (n=682 breakpoints) and constitutional cases from the literature (n=350
 740 breakpoints) using our “free simulation as a reference group. Relative risk ratios and 95 % confidence intervals systematically overlap and/or are shifted in the
 741 same direction so that the two groups can be merged for further analyses.

8. Annexe 2 :

Next generation cytogenetics: genome-imaging enables comprehensive structural variant detection for 100 constitutional chromosomal aberrations in 85 samples.

1 **Next generation cytogenetics: genome-imaging enables comprehensive**
2 **structural variant detection for 100 constitutional chromosomal**
3 **aberrations in 85 samples**

4

5 Tuomo Mantere,^{1,2,3,11} Kornelia Neveling,^{1,4,11} Céline Pebrel-Richard,⁵ Marion Benoist,⁶ Guillaume
6 van der Zande,¹ Ellen Kater-Baats,¹ Imane Baatout,⁶ Ronald van Beek,¹ Tony Yammine,⁷ Michiel
7 Oorsprong,¹ Daniel Olde-Weghuis,¹ Wed Majdali,⁶ Susan Vermeulen,¹ Marc Pauper,¹ Aziza
8 Lebbar,⁶ Marian Stevens-Kroef,¹ Damien Sanlaville,⁸ Dominique Smeets,¹ Jean Michel Dupont,^{6,9}
9 Alexander Hoischen,^{1,2,10,11} Caroline Schluth-Bolard,^{8,11} Laïla El Khattabi^{6,9,11}

10

11 ¹Department of Human Genetics, Radboud University Medical Center, Nijmegen, The
12 Netherlands,

13 ²Radboud Institute of Medical Life Sciences, Radboud University Medical Center, Nijmegen, The
14 Netherlands,

15 ³Laboratory of Cancer Genetics and Tumor Biology, Cancer and Translational Medicine Research
16 Unit and Biocenter Oulu, University of Oulu, Oulu, Finland

17 ⁴Radboud Institute of Health Sciences, Radboud University Medical Center, Nijmegen, The
18 Netherlands,

19 ⁵Department of Cytogenetics, University hospital of Clermont-Ferrand, France,

20 ⁶Department of Medical Genetics, Cochin Hospital, APHP.Centre, University of Paris, Paris,
21 France,

22 ⁷Institut Neuromyogène, CNRS UMR 5310, INSERM U1217, Lyon 1 University, Lyon, France; Unit
23 of Medical Genetics, Saint-Joseph university, Beyrouth, Lebanon,

24 ⁸Department of Genetics, Hospices Civils de Lyon, Bron, France, Institut Neuromyogène, CNRS
25 UMR 5310, INSERM U1217, Lyon 1 university, Lyon, France,

26 ⁹Cochin Institute, INSERM U1016, Paris, France,

27 ¹⁰Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud

28 University Medical Center, Nijmegen, The Netherlands,

29 ¹¹These authors contributed equally to this work

30

31 **Corresponding Authors**

32 Laïla El Khattabi, PharmD PhD

33 Department of Genomic Medicine

34 Cochin Hospital, APHP.centre - University of Paris

35 75014 Paris, France

36 Email: laila.el-khattabi@aphp.fr

37 AND

38 Alexander Hoischen, PhD

39 Department of Human Genetics & Department of Internal Medicine

40 Radboud university medical center

41 6500HB Nijmegen, The Netherlands

42 Email: alexander.hoischen@radboudumc.nl

43

44 **Abstract**

45 Chromosomal aberrations and structural variations are a major cause of human genetic diseases.
46 Their detection in clinical routine still relies on standard cytogenetics, karyotyping and CNV-
47 microarrays, in spite of the low resolution of the first one and the inability to detect neither
48 balanced SVs nor to provide the genomic localization or the orientation of duplicated segments,
49 of the latter. We here investigated the clinical utility of high resolution optical mapping by
50 genome imaging for patients carrying known chromosomal aberrations in a context of
51 constitutional conditions.

52 For 85 samples, ultra-high molecular weight gDNA was isolated either from blood or cultured
53 cells. After labeling, DNA was processed and imaged on the Saphyr instrument (Bionano
54 Genomics). A *de novo* genome assembly was performed followed by SV and CNV calling and
55 annotation. Results were compared to known aberrations from standard-of-care tests
56 (karyotype, FISH and/or CNV-microarray).

57 In total, we analyzed 100 chromosomal aberrations including 7 aneuploidies, 35 translocations, 6
58 inversions, 2 insertions, 39 copy number variations (20 deletions and 19 duplications), 6
59 isochromosomes, 1 ring chromosome and 4 complex rearrangements. High resolution optical
60 mapping reached 100% concordance compared to standard assays for all aberrations with non-
61 centromeric breakpoints.

62 Our study demonstrates the ability of high resolution optical mapping to detect almost all types
63 of chromosomal aberrations within the spectrum of karyotype, FISH and CNV-microarray. These
64 results highlight its potential to replace these techniques, and provide a cost-effective and easy-
65 to-use technique that would allow for comprehensive detection of chromosomal aberrations.

66 Introduction

67 Structural variants (SV) play an important role in human diversity and diseases. The emergence
68 of cytogenetic tools, starting with karyotyping followed by fluorescence *in situ* hybridization
69 (FISH) and CNV-microarrays, allowed for their detection and thereby significantly contributed to
70 the discovery of disease causing genes.¹⁻³ However, these tools show significant limitations as
71 karyotyping has a very low resolution, estimated at 5-10 Mb on average. Additionally, CNV-
72 microarrays are not able to detect mosaicism lower than 5-20% or balanced chromosomal
73 aberrations, and do not provide information on the location of the structural variation, e.g.
74 mapping of insertions is impossible.

75 Despite their drawbacks, karyotyping and CNV-microarrays still represent major tools in the
76 routine genetic investigation of constitutional and somatic diseases, since chromosomal
77 aberrations are major causes of e.g. reproductive disorders, recurrent miscarriages, congenital
78 malformations or (neuro-)developmental disorders. Karyotyping is thereby indicated for diseases
79 where numerical and structural balanced aberrations are highly represented, such as in
80 reproductive disorders where (sex)chromosomal aneuploidies and large structural aberration
81 including balanced rearrangements are frequently present.⁴⁻⁷ CNV-microarray is recommended as
82 first-tier test for developmental disorders (DD) with or without multiple congenital anomalies
83 (MCA),⁸ as it enables the diagnosis of sub-chromosomal copy number variations (CNV) including
84 clinically relevant microdeletions/microduplications.^{1; 2; 9} In DD/MCA, the diagnostic rate rose
85 from less than 5% with karyotyping^{10; 11} to 15 to 20% with CNV-microarray leading to the
86 replacement of the former analysis by the later as a first-tier test.^{8; 12}

87 The recent breakthrough in sequencing technologies raised great interest in complementing or
88 replacing cytogenetic tools for an all-in-one genetic test allowing for the detection of both
89 nucleotide variants and structural variants.¹³⁻¹⁵ Moreover, short read sequencing became

90 reasonably inexpensive and is versatile in terms of protocols (gene panel, whole exome
91 sequencing (WES) and whole genome sequencing (WGS)). Yet, the detection of structural
92 variants remains challenging because of (i) the relatively limited read length and (ii) the repetitive
93 nature of sequences at some structural variation breakpoints. Although many improvements
94 regarding technical aspects and data analysis pipelines have been achieved, genome sequencing
95 is still not able to comprehensively and cost-effectively detect balanced structural anomalies,
96 impeding its wide implementation in clinical cytogenetic laboratories. Moreover, the most
97 comprehensive analysis of SVs in WGS data requires the use of multiple tools, as established e.g.
98 by the 1000 genomes project SV consortium.¹⁶ Hence, a real-time analysis with fast turnaround
99 time is not yet feasible for each and every laboratory. It is expected that long-read whole
100 genome sequencing (LR-WGS) will dramatically improve the ability to identify SVs in individual
101 genomes,¹⁷ and examples have shown this utility for individual research cases.^{18; 19} However, the
102 routine use of long-read sequencing as a diagnostic tool requires several improvements.

103 To this end, a tool complementary to sequencing that may truly replace standard cytogenetics
104 may offer great additional value. Optical mapping by genome imaging consists of imaging very
105 long linear single DNA molecules (median size >250 kb) that have been labeled at specific sites.
106 Since its first description,²⁰ this formerly tedious technique has been updated by Bionano
107 Genomics. They combined microfluidics, high-resolution microscopy and automated image
108 analysis to allow for high-throughput whole genome imaging and its *de novo* assembly.^{21; 22}
109 Historically, such maps have been used as a scaffold to guide the assembly of NGS contigs to
110 build reference genomes of several plant and animal species.²³⁻²⁵ More recently, methods
111 dedicated to the detection of SVs in humans have been developed. Data analysis thereby
112 includes two distinct pipelines: a CNV pipeline that allows for the detection of large unbalanced
113 aberrations based on normalized molecule coverage, and an SV pipeline that compares the
114 labeling patterns between the constructed genome maps of the studied sample and a given

115 reference. The latter allows for the genome-wide detection of SVs as small as few hundred base
116 pairs, including insertions, deletions, duplications as well as inversions and translocations.

117 Optical mapping using Bionano[®] recently proved to allow for efficient detection of a wide range
118 of chromosomal anomalies in leukemia.²⁶ It has also been used to detect germline SVs in
119 individual research cases^{27; 28} or individuals from the 1000 genomes consortium¹⁶ and to unravel
120 population specific SVs.²⁹

121 The aim of the current study was to benchmark Bionano Genomics' optical mapping technology
122 against standard-of-care cytogenetic tools (karyotype, FISH and/or CNV-microarray). To do so,
123 we analyzed a wide range of simple and challenging chromosomal aberrations, which had been
124 previously characterized by standard approaches, in samples from patients with a broad range of
125 clinical indications.

126 **Subjects and methods**

127 *Patient selection and sample collection*

128 This multicenter study involved a total of 85 samples from four genetic academic centers from
129 the Netherlands (Radboud University Medical Center, RUMC) and France (Cochin hospital in
130 Paris, Hospices Civils in Lyon and the university hospital of Clermont-Ferrand). Patients were
131 referred to one of the inclusion centers for developmental or reproductive diseases.
132 Recommended chromosomal investigations were performed according to the indications.
133 Karyotyping was performed in case of reproductive disorders or family history of balanced
134 chromosomal anomaly. CNV-microarray, and karyotyping for some samples, was performed in
135 case of developmental disorders. In some cases, additional investigations including fluorescence
136 *in situ* hybridization (FISH) were performed to characterize an identified anomaly.

137 Cases for which (i) a chromosomal anomaly was identified by karyotyping, CNV-microarray or
138 FISH, and (ii) for which there was enough residual blood (EDTA or heparin) or cultured cells
139 available after routine testing, were included. Samples were anonymized or informed consent is
140 available, respectively. Blood samples for high molecular weight DNA extraction were stored at -
141 20°C for a maximum of one month and at -80°C for longer term storage. In addition, several
142 cases with known aberrations had material other than blood available as a residual material from
143 routine testing. This included 8 amniotic fluid cell lines, 4 chorionic villi cell lines and 8
144 lymphoblastoid cell lines, which were all generated from primary cultures according to standard
145 diagnostic procedures.

146

147 *Karyotyping*

148 Karyotyping was performed according to previously described standard protocols.³⁰
149 Chromosomal abnormalities were described according to the International System for Human
150 Cytogenetic Nomenclature (ISCN, 2016).

151

152 *Fluorescence in situ hybridization*

153 Fluorescence *in situ* hybridization (FISH) was performed on standard chromosome slides
154 according to the manufacturer's instructions (Vysis, Abbott, USA), or using isolated BAC-clones as
155 FISH-probes following standard procedures.

156

157 *CNV-microarray*

158 CNV-microarray was performed using the Agilent SurePrint G3 ISCA v2 CGH 8x60K or SurePrint
159 G3 Human CGH Microarray 4x180K (Agilent Technologies, Santa Clara, CA, USA), or the
160 Affymetrix Cytoscan™ HD Array (Thermo Fisher Scientific, Waltham, USA). Genome coordinates
161 were provided according to hg19/GRCh37 human reference genome.

162

163 *Ultra-high molecular weight DNA isolation, DNA labeling and data collection for optical mapping*

164 For each patient, ultra-high molecular weight (UHMW) DNA was isolated from 400 μ L of whole
165 peripheral blood (EDTA or Heparin) or 1-1.5 million cultured cells (lymphoblastoid cells, amnion
166 cells or chorionic villi cells), using the SP Blood & Cell Culture DNA Isolation Kit and according to
167 manufacturers' instructions (Bionano Genomics®, San Diego, CA, USA). Briefly, cells were treated
168 with LBB lysis buffer to release genomic DNA (gDNA) which was bound to a nanobind disk,
169 washed and eluted in the provided elution buffer.

170 UHMW DNA molecules were labeled using the DLS (Direct Label and Stain) DNA Labeling Kit
171 (Bionano Genomics®, San Diego, CA, USA). Direct Label Enzyme (DLE-1) and DL-green
172 fluorophores were used to label 750 ng of gDNA. After a wash-out of the DL-Green fluorophores
173 excess, DNA backbone was counterstained overnight before quantitation and visualization on a
174 Saphyr® instrument.

175 Labeled UHMW gDNA was loaded on a Saphyr chip® for linearization and imaging on the Saphyr
176 instrument (Bionano Genomics, San Diego USA) (Supplementary Figure 1).

177 *De novo assembly and structural variant calling*

178 The *de novo* assembly and Variant Annotation Pipeline were executed with Bionano Solve
179 software v3.4 or v.3.5. Results were analyzed through two distinct pipelines: a CNV pipeline that
180 allows for the detection of large unbalanced aberrations based on normalized molecule
181 coverage, and an SV pipeline that compares the labeling patterns between the constructed
182 sample genome maps and a reference genome map. Reporting and direct visualization of
183 structural variants were performed using Bionano Access software v1.4.3 or v.1.5.1. The
184 following filtering thresholds were applied: confidence values for insertion/deletion=0,
185 inversion=0.01, duplications= -1, translocation=0 and CNV=0.99. SV calls were compared to an

186 optical mapping dataset of 204 human control samples (provided by Bionano Genomics) to filter
187 out common SVs and potential artifacts (both technical and reference-genome related)
188 (Supplementary Figure 1).

189 *Data analysis and comparisons*

190 All optical mapping results were analyzed genome-wide (Supplementary Table 1) for all samples
191 irrespective of the patient's chromosomal status. We subsequently compared SVs and CNVs
192 detected by optical mapping to the ones previously identified by standard-of-care techniques
193 (karyotype and/or CNV-microarray).

194

195 **Results**

196 *Population description*

197 All 85 samples included in this study were previously analyzed by karyotyping, FISH and/or CNV-
198 microarray according to the reason for referral and the respective international
199 recommendations (Figure 1A and B, Supplementary Table 1). Reasons for referral included
200 developmental delay including autism spectrum disorders or intellectual disability, associated or
201 not with congenital malformations (49 patients, 57.6%), reproductive disorders (15 patients,
202 17.6%), familial history of chromosomal aberration (12 patients, 14.1%), and abnormal prenatal
203 test results (9 patients, 10.6%). These samples exhibited a total of 100 chromosomal aberrations
204 with 11 different types of aberrations from the previous standard diagnostics tests, summarized
205 in Figure 1C. Additionally, nine known aberrations in this cohort were beyond the scope of this
206 study due to breakpoints in the (peri-)centromeric regions of any chromosome or p-arm of
207 acrocentric chromosomes, and were therefore excluded from further analyses.

208

209 *Results of Optical Mapping with Bionano Genome Imaging*

210 Bionano genome imaging generated on average 655 Gbp of data per sample (853 Gbp for
211 samples processed in Nijmegen, aiming at ~200X genome coverage, and 463 Gbp for samples
212 processed in France, aiming at ≥80X genome coverage per sample, respectively). The average
213 N50 molecule length (> 150 Kbp) was 267 Kbp and label density was 15.1 labels/100 Kbp. This
214 resulted in an average map rate of 76.8% and an effective coverage of 152x (192x for Radboud
215 samples, 114x for French samples) (Supplementary Table 2).

216

217 Structural variant calling identified on average 5,758 (+/- 344) SVs per sample, of which the vast
218 majority corresponded to insertions and deletions (with an average of 4,127 (+/- 239) and 1,549
219 (+/- 108) respectively). Filtering out events which were present in a database comprising of 204
220 population control samples resulted in an average of 80 (+/-65) rare SVs per sample, of which 41
221 (+/- 28) were overlapping with genes (Figure 2, Supplementary Table 3). Besides SV detection,
222 CNV detection was performed using a separate coverage-depth based algorithm that is included
223 in the *de novo* assembly and variant calling pipeline.^{31; 32} This analysis resulted in an average of 1
224 gain and 10 losses per sample without applying any size threshold cut-offs (Supplementary Figure
225 2). Of note, CNV calls are often segmented into multiple calls, hence the true number of CNVs is
226 expected to be lower.

227

228 *Detection of diagnostically reported aberrations with genome imaging*

229 All diagnostically reported aberrations in our study cohort were detected correctly either by the
230 SV or the CNV calling, with several aberrations being identified by both algorithms, reaching a
231 100% concordance for optical mapping with the previous diagnostic test results (Supplementary
232 Table 1). For five samples however, filter settings needed to be adapted in order to detect the

233 expected aberrations (see Supplementary Table 1). Adaptation included setting the confidence
234 value for CNVs to 0 (3 samples) and turning off the SV DLE-1 mask filter (2 samples).

235 The 100 identified aberrations included 7 aneuploidies, 20 deletions and 19 duplications, 35
236 translocations, 6 inversions, 2 insertions, 6 isochromosomes and 1 ring chromosome (Figure 1C).

237 In addition, four of our patients showed complex chromosomal rearrangements, defined as cases
238 where aberrations involve three or more chromosomes or when at least four SVs are detected
239 on the same chromosome. For graphical representation of different types of chromosomal
240 aberrations see Figures 3 and 4.

241

242 *Aneuploidies, partial aneuploidies and large CNVs*

243 Our study cohort included 7 full aneuploidy samples, including 3x XXY, 2x monosomy X, 1x
244 trisomy 14 and 1x trisomy 21 (the two latter ones were detected in prenatal samples and were
245 mediated by Robertsonian translocations). In addition, 4 mosaic monosomy X samples were
246 included (Supplementary Table 1). All aneuploidies of the autosomes were called correctly with
247 the used algorithms, whereas the aneuploidies of the sex chromosomes had to be manually
248 inferred from the visualized data of the CNV plot (Figure 3). This manual inference is no longer
249 required with the recent Bionano Solve v3.5. In addition to whole chromosome aneuploidies, five
250 large CNVs ranging in size between 6.6 and 14 Mb, and 7 large aberrations corresponding to
251 derivative chromosomes from unbalanced translocations detected by karyotyping, were included
252 and detected correctly.

253

254 *Isochromosome*

255 Six of our samples contained isochromosomes. Four of those were iso-dicentric Y-chromosomes,
256 one sample contained an isodicentric chromosome 15, and another had an isodicentric
257 chromosome X. The four isodicentric Y-chromosomes all showed a similar genome map pattern

258 (Figure 5). Whereas all four have normal coverage at the p-arm and a small part of the q-arm
259 (until q11.221), there is no coverage at q11.222 to q11.23 and at the end of q12. The largest part
260 of q12 had no coverage in none of the samples (including controls), as this part of the
261 chromosome represents a gap in the reference genome (hg19, N-base gap). Interestingly,
262 whereas samples 27, 57, and 79 have a nearly identical coverage pattern, only sample 55 shows a
263 slightly different breakpoint, with a part of q11.222 still being covered. While the CNV or
264 coverage pattern undoubtedly allows to decide about the presence of isochromosomes in all
265 samples, it should be noted that centromeres itself cannot be detected, hence the distinction
266 between dicentric vs. monocentric status may remain uncertain in some cases. Of note, the
267 isochromosomes of chromosomes 15 and X presents are remarkable: For chromosome 15, the
268 fractional copy numbers of the affected regions differed, and were called as 3 and 4. The
269 isodicentric chromosome X was present in low mosaic state (see Supplementary Figure 3).

270

271 *Ring chromosome*

272 One of the samples analyzed contained a mosaic ring chromosome X, as previously detected by
273 karyotyping (Figure 6). The karyotype reported was 45,X[14]/46,X,r(X)(p11.21;q21.1)[21]. The
274 patient presented with growth retardation and development delay. Following genome imaging,
275 an intrachromosomal translocation on chromosome X was detected, connecting positions chrX:
276 g.57,009,891 (p11.21) and chrX:g.78,599,384 (q21.1), confirming and refining the positions
277 previously detected by karyotyping. The fractional copy number of 1.6 for this region, compared
278 to 1.0 for the rest of this chromosome confirmed the mosaic state of this ring chromosome.

279

280 *Translocations and inversions*

281 Thirty-five of the investigated samples carried previously identified balanced (n=28) and
282 unbalanced (n=7) translocations, which were all detected by genome imaging. As expected,
283 unbalanced translocations were detectable by both structural variant calling and CNV calling,
284 whereas balanced translocations and inversions were only detected by SV calling (Figure 4B, C).
285 Traditionally, balanced translocations can be detected via karyotyping but not via CNV-
286 microarray. Genome imaging is able to refine translocation breakpoints for such cases.
287 Accordingly, several balanced translocations and inversions were shown to likely disrupt protein-
288 coding genes, including the well described *SETBP1* (MIM: 611060), *KANSL1* (MIM: 612452),
289 *DYRK1A* (MIM: 600855), and *PIGU* (MIM: 608528) genes, with the latter two being disrupted by
290 the same translocation (Figure 7). The breakpoints for *KANSL1* (sample 49) had previously been
291 validated using FISH and whole-genome sequencing (WGS),³³ whereas the others are newly
292 uncovered and still need to be confirmed. In all cases, the patient's phenotype matches the
293 expected phenotype for the dominant diseases associated with the respective genes. Detection
294 of breakpoints with optical mapping is much more accurate than karyotyping. For the few
295 breakpoints for which WGS data were available for comparison,³³ the breakpoint accuracy was
296 within five kb (Supplementary Figure 4).

297

298 *Microdeletions and -duplication*

299 In addition to large chromosomal aberrations (aneuploidies, large CNVs and translocations), our
300 cohort included 34 microdeletion/-duplications (<5 Mb). These microdeletion/duplications
301 ranged in size from 34 Kbp (sample 84) to 4.2 Mbp (sample 44), and included some of the well-
302 known microdeletion/duplication syndromes such as DiGeorge syndrome (22q11.2 deletion
303 syndrome, OMIM: 188400), Williams-Beuren syndrome (deletion 7q11.23, OMIM: 194050),
304 Charcot-Marie-Tooth syndrome type 1A (CMT1A, duplication 17p12, OMIM: 118220) and 1q21.1
305 susceptibility locus for Thrombocytopenia-Absent Radius (TAR) syndrome (OMIM: 274000).
306 Although the presence of segmental duplications (SegDups) for several of these

307 microdeletions/duplications often leads to breaking of the genome maps, all
308 microdeletions/duplications were correctly called by either the SV or CNV algorithms or both,
309 although most events were called by the CNV tool. SeqDups are often mediating recurrent CNVs,
310 for example the 22q11.2 microdeletion-causing DiGeorge syndrome (Figure 8). Depending on
311 size and structure of these SeqDups, the SVs were occasionally disrupted, or falsely filtered out
312 due to high population frequencies of partially overlapping SVs (Supplementary Table 1). We
313 expect that the analysis of individual molecules of sizes up to 2Mb shall allow full assembly maps
314 even for those regions but additional software improvements may be required.

315

316

317 *Complex cases*

318 Finally, four of the samples included in this study presented with complex rearrangements (28,
319 52, 55, 66), four of which are samples of patients with developmental delay and/or intellectual
320 disability (Supplementary Table 1). For example, karyotype of Sample 28 (Figure 9) showed a
321 translocation $t(3;6)(q1?2;p2?2)$, a derivative chromosome 4 ($der(4)(:p1?2->q1?2:)$) and a
322 derivative chromosome 5 ($der(5)(4pter->4p1?2::4q1?2->4q34.2::5p14.2->5qter)$) in different
323 clones. CNV-microarray showed losses on 4q34 ($4q34.2q34.3(176587929_190957474)x1$ dn) and
324 5p15 ($5p15.33p14.2(113577_24449849)x1$ dn). Following optical mapping, the translocation
325 $t(3;6)(q1?2;p2?2)$ was identified as $t(3;6)(q13.12;p24.3)$. In addition, a translocation
326 $t(4;5)(q34.2;p14.2)$, a loss of 4q34.2q34.3 and a loss of 5p15.33p14.2 were detected, concordant
327 with previous results. In the same sample, genome imaging also revealed putative additional
328 translocations $t(3;4)(q13.11;q12)$, $t(3;4)(q13.11;p11)$ $t(4;6)(q12;p22.3)$ and an inversion
329 $inv(13)(q31.2;q33.3)$ (Figure 9).

330

331 Another sample (66) showed a 3-way translocation $t(3,13,5)(p11.1;p12;p14)$ after karyotyping
332 and four losses on chromosome 3
333 ($3p14.1(65238298_68667113)\times 1,3p13(70127345_73724765)\times 1$
334 $,3p12.1(83784489_85467284)\times 1,3q11.2(97180779_97270083)\times 1$) following CNV-microarray
335 (Figure 10). Optical mapping confirmed these aberrations, but unraveled additional complex
336 rearrangements on chromosome 3, leading to the identification of a chromoanagenesis. For all
337 residual samples with complex rearrangements, see Supplementary Table 1 and Supplementary
338 Figures 5 and 6.

339

340 Taken together, optical mapping allowed the correct unravelling of complex karyotypes, which
341 previously required the combination of karyotyping, FISH and CNV-microarrays, by combining the
342 detection of translocations and imbalances (CNVs, gains and loss of genetic material) including
343 balanced and unbalanced events in one assay and at an unprecedented resolution.

344

345

346 Discussion

347 Chromosomal aberrations and SVs are frequently involved in many genetic diseases including
348 developmental disorders, congenital malformations, intellectual deficiency, reproductive
349 disorders as well as cancer. Hence, their accurate detection is critical to achieving a complete
350 genetic investigation but limitations of the current standard-of-care genetic analyses
351 (karyotyping, FISH, CNV-microarray and NGS) preclude any comprehensive characterization
352 without combining multiple approaches.^{16, 34} Indeed, to date, not a single technology offers a full
353 resolution of chromosomal aberrations in all samples. The traditional karyotyping is still
354 performed as the first-tier test in case of reproductive disorders in spite of its poor diagnostic
355 rate (overall less than 10%), likely due to its very low resolution. Moreover, its quality is

356 unpredictable since it varies between samples and laboratories, it depends on the availability of
357 viable cells and relies on the expertise of the technician and the cytogeneticist which is
358 decreasing over the years because of lack of training. Hence, there is need for a more robust,
359 high-resolution and automatable method. On the other hand, CNV-microarray represents one
360 such robust routine tool, that has allowed an improved diagnostic yield e.g. to approximately
361 15% for neurodevelopmental disorders,⁸ but it lacks the ability to detect balanced aberrations
362 such as translocations or inversions or to decipher the orientation of duplicated or inserted
363 segments, and resolution remains restricted to a few kilobases. Sequencing based assays for SV
364 detection are constantly improving.^{35; 36} This includes improved CNV calling from exome or
365 genome sequencing, however most comprehensive detection requires a combination of analysis
366 tools.³⁷⁻⁴⁰ Moreover, it requires the local implementation of bioinformatic pipelines that have not
367 yet undergone a large scale clinical validation. In addition to technical and computational
368 hurdles, SV detection by sequencing based technologies becomes difficult when breakpoints
369 localize within repetitive sequences which is frequently observed since many SVs are caused by
370 the non-allelic homologous recombination of repeats in the first place. It is expected that long-
371 read sequencing may enable near perfect genomes one day, but so far technologies, analyses as
372 well as throughput and prices do not allow its routine clinical use.¹⁷

373 In this manuscript, we have shown that genome imaging is capable of comprehensively and
374 easily detecting all classes of chromosomal aberrations and may complement or replace current
375 cytogenetic technologies. Our cohort was composed of a large panel of different tissues,
376 aberrations and indications representative of what can be encountered in clinical routine. We
377 demonstrated that optical mapping allows for the detection of balanced as well as unbalanced
378 rearrangements at sizes ranging from few kilobases to several megabases or even entire
379 chromosome aneuploidies. The method allows for detection of SVs down to 500 bp, but none of
380 our clinically reported SVs were that small. Copy number variations and aneuploidies were all

381 detected by either the coverage analysis pipeline and/or structural variation analysis. Combining
382 two analysis pipelines, one based on coverage depth and the other one based on the comparison
383 of a *de novo* assembled genome map with a reference map, allows for the most complete
384 detection of all balanced and unbalanced aberrations, as shown by our results. In fact, the first
385 pipeline performs better for large deletions and duplications, and is currently the only tool to
386 detect terminal chromosomal gains/losses or other events that do not create the fusion of
387 unique novel molecules. The second pipeline is more sensitive to small CNVs down to few
388 hundred base pairs, and allows for best breakpoint resolution.

389 Regarding translocation-, inversion- or insertion- events, we demonstrate that they can all be
390 detected as long as the breakpoints do not lie within large repetitive, unmappable regions such
391 as centromeres, p-arm of acrocentric chromosomes, or heterochromatin stretches. Challenges to
392 map such breakpoints along with the inability of the current software to detect loss of
393 heterozygosity were known prior to this study and such results were expected. In fact, these
394 regions are likely not well represented in the human reference genome,⁴¹ they cannot be
395 specifically labeled and they are several megabases long, far larger than the longest single
396 molecules that can be obtained with any current technology. Interestingly, in some cases we
397 were able to detect translocations with breakpoints that lie in pericentromeric regions and which
398 were not detected by paired-end whole-genome sequencing (samples 50, 51 and 54; NGS data
399 not shown, manuscript in preparation). This additional detection did solve the molecular
400 diagnosis for patient 54 whose karyotype is 46,XY,t(20;21)(q11.2;q21). Optical mapping showed
401 that this balanced translocation disrupts the *DYRK1A* gene and refined the breakpoint to
402 21q22.13. This patient displays autism spectrum disorder and microcephaly consistent with a
403 *DYRK1A* haploinsufficiency, which has been shown to be associated with autism spectrum
404 disorder, intellectual disability and microcephaly.⁴²⁻⁴⁴ Similarly, other cases of gene disruption
405 provided hints into the molecular diagnosis of intellectual disability. For example, patient 47 had

406 an inversion that we showed to likely disrupt the *SETBP1* gene and patient 49 had a t(9;17) that
407 potentially disrupts *KANSL1* as previously identified by WGS.³³ In both cases, haploinsufficiency of
408 the respective genes is known to lead to clinical syndromes including intellectual disability^{45, 46}
409 consistent with our patient's phenotypes.

410 Optical mapping was also able to detect complex rearrangements including multiple
411 translocations, or even chromoanagenesis. In some cases, optical mapping results suggested a
412 more complex event than expected (28, 47, 52, 66, 70, 74) where the additional SV calls need to
413 be further validated. Furthermore, in four cases of isochromosome Y, the CNV pipeline detected
414 a coverage profile that is very suggestive of an isochromosome (samples 27, 55, 57, 79) similarly
415 to or better than CNV-microarray results, although the Bionano SV pipeline did not call the
416 isochromosome Y. Such isochromosomes with breakpoints in the long arm of chromosome Y are
417 not detectable by sequencing technologies.

418 As a non-sequencing based technology, single molecule optical mapping overcomes issues due to
419 repetitive regions inaccessible to sequencing. For some aberrations, it even enabled the
420 detection of breakpoints mapping at segmental duplications (Figure 8).⁴⁷

421 It is not unlikely that at some stage (long read) sequencing approaches may allow fully
422 comprehensive assessment of all SVs and chromosomal aberrations in each personal genome,
423 possibly after *de novo* genome assembly instead of re-sequencing.^{41, 48} Some benefits of optical
424 mapping may prevail: 1.) relative ease of analysis, 2.) relative low costs, 3.) optical mapping can
425 produce 300-1600X genome coverage allowing the reliable detection of rare somatic events,
426 with additional improvements in development.²⁶

427 From a technical point of view, optical mapping using Bionano can best be compared to an ultra-
428 high resolution karyotype (~10,000 times higher resolution than the conventional karyotype) that
429 offers a fast (3-4 days from sample to variant calls) and cost-effective (~\$450 list price per

430 genome) alternative to both karyotyping and CNV-microarray. Neither significant data storage
431 capacities nor bioinformatics processing are required. Turnaround time has also been
432 significantly improved in the recent years, as six genomes can be processed in a single run, and
433 instrument price has been reduced (\$150k list price). It is also worth noting that the filter settings
434 suggested here result in a small number of events per case (n= 41 rare SVs on average), while
435 detecting all previously known events. This is suggestive for a low false positive rate, although
436 orthogonal validations e.g. by sequencing were beyond the scope of this study. This may be in
437 contrast to NGS-based SV calling: several reports point out the high number of false calls with
438 sequencing based technologies.^{49, 50} Additional clinical analysis filter may include overlap-analysis
439 of SVs with known disease genes or loci. This is a crucial point since in the context of clinical
440 routine, genetic investigation should be time-efficient since the longer the turn-around time the
441 lower the quality of disease management especially in case of reproductive disorders. In
442 addition, our results support the robustness of the technology as our samples were processed in
443 three different facilities. Results were highly similar in terms of quality metrics, number of variant
444 calls and performance stated by the rate of concordance with conventional cytogenetic analyses.
445 Some differences in the number of calls are most likely due to different versions of analysis
446 software used. Clinically relevant results were unchanged.

447 The technology has the potential to keep improving at both technical and analytical levels.
448 Indeed, a closer examination of the maps or loosening the filters for few samples (34, 42, 50, 76,
449 and 81) led to the identification of initially missed structural variants, supporting the potential of
450 improvement of the software or analysis settings. Other aspects that are being improved to meet
451 cytogeneticist expectations include loss of heterozygosity analysis, ideogram style representation
452 of chromosomal aberrations, e.g. translocations, ISCN nomenclature outputs, and hyperlinks to
453 genome databases. As with comparative genomic hybridization, polyploidies cannot be detected
454 with the current analysis pipeline but haplotype analysis should make this detection possible.

455 The main focus here was to investigate the concordance i.e. true positive rate for known
456 aberrations as a first step to explore the possibility to replace standard cytogenetic assays by
457 optical mapping. In addition, it is also attractive that this can complement NGS to achieve a
458 better and possibly nearly complete genome analysis. In developmental disorders, optical
459 mapping could complement sequencing approaches to allow for a comprehensive genomic
460 investigation. In reproductive disorders, it could replace karyotyping as the main method
461 complemented by a count of few metaphase spreads by karyotyping to prevent that balanced
462 Robertsonian or whole arm translocations are missed in few respective cases.

463 Our results pave the road to a second phase that would aim at evaluating the clinical utility of the
464 technology for all patients referred for cytogenetic investigation and assess the added-value in
465 terms of diagnostic yield (detection of novel SVs) and genetic counselling. In fact, samples
466 currently investigated with CNV-microarray could have undetected balanced structural
467 variations, or other pathogenic SVs in complex regions of the genome that remain inaccessible to
468 CNV-microarray and NGS detection, as suggested by most recent findings in singleton research
469 cases.^{47, 51} Similarly, patients who have normal karyotype could bear variants that were
470 undetected because they are below the karyotype resolution. Furthermore, the absence of
471 sequencing could be preferred in some cases to avoid undesired incidental findings especially for
472 some patients referred for reproductive disorders. The sensitivity of optical mapping to detect a
473 repeat-contraction related disease such as Facio-Scapulo-Humeral Muscular Dystrophy (FSHD)⁵²
474 ⁵³ opens up new perspectives for the detection of expansion diseases such as Fragile X syndrome
475 or Huntington disease and SVs on the Y chromosome which is rich in repetitive sequences and
476 still challenging for sequencing.

477 To conclude, this is the first clinical study to validate genome-mapping as a solid alternative
478 approach to karyotyping, FISH and CNV-microarray for the detection of chromosomal aberrations
479 in constitutional diseases. We showed that optical mapping is capable of reaching 100%

480 concordance, while detecting all different types of chromosomal anomalies including
481 aneuploidies, CNVs as well as balanced chromosomal abnormalities and complex chromosomal
482 rearrangements.

483

484 **Acknowledgments**

485 We are thankful to the Department of Human Genetics in Nijmegen, especially Helger Yntema,
486 Lisenka Vissers, Marcel Nelen, and Han Brunner for providing support and critical feedback. We
487 are grateful to the Radboudumc Genome Technology Center for infrastructural and
488 computational support. A. Hoischen, Ph.D. was supported by the Solve-RD project. The Solve-RD
489 project has received funding from the European Union's Horizon 2020 research and innovation
490 program under grant agreement No 779257. SOLVE-RD. This research was part of the
491 Netherlands X-omics Initiative and partially funded by NWO, project 184.034.019 and Radboud
492 Institute for Molecular Life Sciences PhD grants (to A. Hoischen). TM was supported by the Sigrid
493 Jusélius Foundation.

494 We are grateful to the French "Agence de la Biomédecine" and the APHP.center Paris university
495 hospitals for their financial support the French part of the project (granted to L. El Khattabi; AOR
496 2018 "ART, prenatal diagnosis and genetic diagnosis" and Merri-SERI 2019, respectively). We
497 would like also to thank Faten Hsoumi (Cytogenetics department, Cochin Hospital, Paris, France)
498 for getting some CNV-microarray images, Emilie Chopin and Isabelle Rouvet (Cellular
499 biotechnology center, Hospices Civils de Lyon, France) for providing lymphoblastoid cell lines, the
500 Gentyane facility staff for providing genome imaging service for samples from Clermont-Ferrand
501 hospital (France), and the clinical geneticists and cytogeneticists from the French ANI project
502 whom patients were included in the present study (Bruno Delobel, Bénédicte Duban-Bedu,
503 Dominique Martin-Coignard, Marc Planes, Céline Freihuber, Jean-Pierre Siffroi, Florence

504 Amblard, Marlène Rio, Laurence Lohman, Véronique Paquis, Françoise Devillard, Bertrand Isidor,
505 James Lespinasse, Gwenaël Nadeau and Laurent Pasquier). The ANI project aims at better
506 characterizing *de novo* apparently balanced chromosomal rearrangements associated with
507 intellectual disability using high throughput sequencing (ANI study is supported by the French
508 Ministry of Health (DGOS) and the French National Agency for Research (ANR) PRTS 2013 grant
509 to C. Schluth-Bolard, n° PRTSN1300001N).

510 We would like to acknowledge support from scientists and staff at Bionano Genomics including
511 Alex Hastie, Andy Pang, Lucia Muraro, Kees-Jan Francoijs, Sven Bocklandt, Yannick Delpu, Mark
512 Oldakowski, Ernest Lam, Thomas Anantharaman, Scott Way, Henry Sadowski, Amy Files, Carly
513 Proskow.

514

515 **Declaration of Interests**

516 Bionano Genomics sponsored part of the reagents used for this manuscript. Other than this, the
517 authors declare no competing interest.

518

519 **Web Resources**

520 Bionano Access™: <https://bionanogenomics.com/support/software-downloads/#bionanoaccess>

521

522 **Figure Titles and Legends**

523 **Figure 1.** Description of the study population (n=85). A) Main reason for referral. B) Distribution
524 of the different cytogenetic and molecular tests used for diagnosis. C) Distribution of
525 chromosomal aberrations as assessed by standard of care genetic investigations.

526 **Figure 2:** SV detection and filtering. Average number of SVs detected per sample, given per type
527 of SV (total, insertion, deletion, inversion, duplication, interchromosomal translocation and
528 intrachromosomal translocation). Dark blue: all variants, median blue: rare variants only (not
529 found in control database including 204 samples), light blue: rare variants that overlap with
530 genes.

531 **Figure 3:** Visual representation of optical mapping data **A)** Genome-wide circos plot showing all
532 24 chromosomes in a circular way. For each chromosome, the ideograms are shown at the
533 outside of the circosplot, with ideogram-style chromosomal banding and the centromeres in red.
534 Different colored dots in the boxes underneath represent different called SVs. The blue line in
535 the box underneath represents the CNV profile, with each peak representing a CNV call. **B)** Part
536 of a circos plot, showing the sex chromosomes. The blue CNV line shows two copies of
537 chromosome X, as for autosomes, and one copy of chromosome Y consistent with a sex
538 chromosome aneuploidy (47,XXY, resulting in Klinefelter syndrome).

539 **Figure 4:** Representation of different chromosomal aberrations. A) Sample 1. Left: CNV
540 microarray data showing an 8p22p21.3(18825888_19364764) deletion. Middle: Genome-wide
541 circos plot showing all chromosomes. The deletion is detected by the CNV and the SV pipeline
542 (blue circle). Right: genome map with reference, showing the deletion and affected genes (hg19).
543 B) Sample 18. Left: karyogram showing a 46,XY,t(5;8)(p13.1;p11.2) karyotype. Middle: circos
544 plots with a pink line connecting chromosomes 5 and 8, representing the translocation. Right:
545 genome map, of which the left part maps to chromosome 8 and the right part to chromosome 5.

546 C) Sample 15. Left: karyogram showing an inversion on chromosome 13 (red arrow). Middle:
547 circos plot showing the inversion as an intrachromosomal translocation. Right: genome map that
548 is partly inverted when compared to the reference. One of the breakpoints is interrupting the
549 gene *KLHL1*.

550 **Figure 5:** Isodicentric Y-chromosomes show specific bionano assembly map patterns. A) GTG-and
551 RHG banding of X- and Y- chromosomes of sample 57. B) FISH for sample 57 using probes
552 TelXp/Yp and RP11-209I11 (Yq11.223). C) Bionano genome maps of Y-chromosomes of samples
553 69 (no isodicentric chrY), 55, 57, 27 and 79. Dotted red boxes indicate where isodicentric Y-
554 chromosomes have no coverage when compared to non-isodicentric Y chromosomes (the top
555 genome map).

556 **Figure 6:** Small X ring chromosome. A) Karyogram of sample 39. The red arrow is pointing
557 towards the small X ring chromosome. B) Circos-plot (of chromosome X only) of sample 39. The
558 pink line in the center of the circosplot is indicating the presence of the ring chromosome (called
559 as an intrachromosomal translocation). C) Different genome maps (dark blue bars on top and
560 below the reference) indicating the presence of the ring chromosome. The individual molecules
561 for the genome map below the reference (highlighted by a red circle) are shown at the bottom of
562 this figure. The left part of these molecules (light green bar) map to a region upstream of the
563 centromere, whereas the right part of the same molecules (light blue bar) map to a region
564 downstream of the centromere.

565 **Figure 7.** Examples of inversions and translocations interrupting well known disease causing
566 genes. A) Inversion $\text{inv}(18\text{q})(\text{q}22.1\text{q}12.3)$, disrupting the *SETBP1* gene in sample 47. B)
567 Translocation $\text{t}(9;17)(\text{p}13.3;\text{q}21.31)$, interrupting the gene *KANSL1* in sample 49. C) Translocation
568 $\text{t}(20;21)(\text{q}11.22;\text{q}22.13)$, interrupting the genes *DYRK1A* and *PIGU* in sample 54.

569 **Figure 8.** Example of a typical 22q11.2 microdeletion syndrome, in sample 2 (VCF, Di-George
570 syndrome). A) Affymetrix CytoScan HD array, showing the 22q11.21(18308819_18519921)x1
571 deletion. B) Bionano circos plot, showing an aberrant CNV profile on 22.q11.21. C) Bionano
572 genome maps of chr22q11.21, showing the CNV calls (on top), a segmental duplication bed file
573 (below), the chromosome 22 reference genome map (again below) and the different genome
574 maps (at the bottom). The deletion is clearly called by the CNV calls, and is surrounded by
575 segmental duplications. *SV not called with standard filters, but recovered when % of the
576 Bionano control sample overlap was set to 80%.

577 **Figure 9:** Complex sample 28. A) Karyogram showing the translocation t(3;6) and the derivative
578 chromosomes 4 and 5. B) CNV microarray data showing two *de novo* deletions on chromosome 4
579 and 5. C) Bionano circos plot of chromosomes 3, 4, 5, 6 and 13. The Bionano data confirm all
580 previous data and show the presence of additional translocations t(3;4) and t(4;6) plus an
581 inversion on chromosome 13.

582 **Figure 10:** Complex sample 66. A) CGH Panes of sample 66, showing the whole chromosome 3
583 (left) and chr3:64105005-74361806 (right). B) Karyogram of sample 66:
584 46,XY,t(3;13;5)(p11.1;p12;p14). C) Genome maps of chromosome 3 of sample 66, showing
585 multiple rearrangements on chromosome 3. D) Circos plot of chromosomes 3 and 5, showing
586 multiple intrachromosomal translocations on chromosome 3 and a translocation between
587 chromosomes 3 and 5.

588

589 **Abbreviations**

590 CNV copy number variant

591 DD developmental disorder

- 592 DLE-1 direct Labeling Enzyme-1
- 593 DLS direct Label and Stain
- 594 EDTA ethylenediaminetetraacetic acid
- 595 FISH fluorescence in situ hybridization
- 596 FSHD facioscapulohumeral dystrophy
- 597 gDNA genomic DNA
- 598 ID intellectual disability
- 599 i.e. id est (that is)
- 600 MCA multiple congenital malformations
- 601 NGS next generation sequencing
- 602 SV structural variant
- 603 UHMW ultra-long high molecular weight
- 604 WES whole exome sequencing
- 605 WGS whole genome sequencing
- 606
- 607 **Supplemental Data Description**
- 608 Supplemental Data includes 6 figures and 3 tables.

609 **Supplementary Figure 1.** Workflow of the Bionano technique. For this study, 85 samples for
610 whom extra material was available were included. Ultra-high molecular weight DNA was
611 extracted using the Bionano solution phase DNA isolation method. Labeling was done using the
612 DLE-1 chemistry. High resolution imaging of DNA molecules was done on Bionano Saphyr
613 instruments. As different centers were included, different amount of data was produced
614 (~800Gbp for Radboud, ~300Gbp for the French centers), and samples were analyzed using
615 different software versions (3.4.1 and 3.5). A *de novo* assembly was performed, and both SVs and
616 CNVs were called.

617 **Supplementary Figure 2.** Representative Bionano CNV profiles for different samples. A) Sample
618 2. Loss of 22q11.21(18645354_21465660). B) Sample 8. Gain of 17p12(14087934_15436895). C)
619 Sample 70. Loss of 6q14.1q14.3(76385698_86884355), and gain of 6q16.1(97661978_98726638).
620 D) Genome-wide CNV view (available in Bionano Solve v1.5) of sample 73 with E) chromosome 8
621 highlighted (showing a deletion) and F) chromosome 17 highlighted (showing a duplication).
622 Blue: gains, Red: losses.

623 **Supplementary Figure 3:** Isochromosomes. A) Sample 77 (ish
624 idic(15)(D15Z1+,SNRPN++,D15Z1+)). Left: Circos plot showing an abnormal CNV profile on
625 chromosome 15. Top right: CNV-microarray data showing a gain on chr15. Bottom right: optical
626 mapping data, showing a CNV profile that is nearly identical to the CNV-microarray profile.
627 Numbers present fractional copy numbers. B) Sample 78 (46,X,idic(X)(p11.21)). Left: Circos plot
628 showing a CNV baseline suggesting one copy of chromosome X (compared to the CNV line of
629 chr22 partially shown on the left side). Additionally, the CNV profile shows a mosaic “gain”
630 (compared to the baseline) on part of the chrX p-arm and the whole q-arm. Top right: CNV-
631 microarray data showing a global loss on chrX (compared to a 46,XX control sample). However,
632 the degree of loss varies within the chromosome consistent with a mosaic
633 45,X/46,X,idic(X)(p11.21) karyotype. Bottom right: optical mapping data showing a CNV profile

634 that is nearly identical to the CNV-microarray profile. Numbers present fractional copy numbers.
635 Red box shows parts of the chromosome 15 and X respectively that make up the iso-chromosomes.
636 Grey box indicates the centromere (15 and X) and/or acrocentric p-arm (15).

637

638 **Supplementary Figure 4.** Genome imaging breakpoint detection for translocation
639 $t(9;17)(p13;q21)$, disrupting the gene *KANSL1* (patient 49). The two green bars represent the
640 references of chromosomes 9 and 17, respectively. The mint bar in between represents the
641 genome map of the translocation. The blue bar underneath represents the *KANSL1* gene. Small
642 vertical black lines represent identified labels, and the red vertical lines indicate the translocation
643 breakpoints, with an uncertain region of 3,828 bp in between shown in purple. The breakpoints
644 are located between basepair-positions 35,771,617 and 35,773,383 on chromosome 9, and
645 between 44,137,912 and 44,141,740 on chromosome 17.

646 **Supplementary Figure 5:** Complex sample 52. A) Karyotype of sample 52, interpreted as
647 $46,XY,der(8)t(8;22)(q12;q12),der(13)t(8;13)(q31;q23),der(14)t(14;15)(q11.2;q25),der(15)t(14;15)($
648 $q21;q24),der(22)t(13;22)(q31.1;p11.2)$. B) FISH of sample 52, using FISH probes wcp8 (red),
649 wcp14 (green). C) FISH of sample 52, using FISH probes wcp8 (green), wcp13 (red). D) FISH of
650 sample 52, using FISH probes wcp15 (green), wcp22 (red). E) Bionano circos plot, showing
651 different translocations $t(8;13)$, $t(8;14)$, $t(14,15)$, and intrachromosomal translocations on chr 8
652 and chr 15.

653 **Supplementary Figure 6:** Complex sample 55. A) Karyotype of sample 55, interpreted as
654 $46,X,idelic(Y)(q11.22),t(5;8)(q23;q24),t(5;11)(p12;p13)[32/50]/45,X,t(5;8)(q23;q24),t(5;11)(p12;p13$
655 $)[10/50]/47,XY,idelic(Y)(q11.22),t(5;8)(q23;q24),t(5;11)(p12;p13)[8/50]$. B) FISH of sample 55,
656 showing the translocations $t(5;8)$ (left) and $t(5;11)$ (right). C) Bionano circos plot of sample 55,
657 showing the translocations $t(5;8)$, $t(5;11)$ and an intrachromosomal translocations 5. D) Bionano

658 genome maps, showing the intrachromosomal translocation on chromosome 5, which is
659 disrupting the gene *GHR*.

660

661 Supplementary Table 1: Comparison of previous diagnostic findings with genome imaging results

662 Supplementary Table 2: Technical performance of genome imaging

663 Supplementary Table 3: Overall numbers of variants per sample

664

665 **References**

- 666 1. Vissers, L.E., Veltman, J.A., van Kessel, A.G., and Brunner, H.G. (2005). Identification
667 of disease genes by whole genome CGH arrays. *Hum Mol Genet* 14 Spec No. 2,
668 R215-223.
- 669 2. Speicher, M.R., and Carter, N.P. (2005). The new cytogenetics: blurring the
670 boundaries with molecular biology. *Nat Rev Genet* 6, 782-792.
- 671 3. Smeets, D.F. (2004). Historical prospective of human cytogenetics: from microscope to
672 microarray. *Clin Biochem* 37, 439-446.
- 673 4. Chantot-Bastaraud, S., Ravel, C., and Siffroi, J.P. (2008). Underlying karyotype
674 abnormalities in IVF/ICSI patients. *Reprod Biomed Online* 16, 514-522.
- 675 5. Hofherr, S.E., Wiktor, A.E., Kipp, B.R., Dawson, D.B., and Van Dyke, D.L. (2011).
676 Clinical diagnostic testing for the cytogenetic and molecular causes of male
677 infertility: the Mayo Clinic experience. *J Assist Reprod Genet* 28, 1091-1098.
- 678 6. De Braekeleer, M., and Dao, T.N. (1990). Cytogenetic studies in couples experiencing
679 repeated pregnancy losses. *Hum Reprod* 5, 519-528.
- 680 7. De Braekeleer, M., and Dao, T.N. (1991). Cytogenetic studies in male infertility: a
681 review. *Hum Reprod* 6, 245-250.
- 682 8. Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P.,
683 Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J., et al. (2010). Consensus
684 statement: chromosomal microarray is a first-tier clinical diagnostic test for
685 individuals with developmental disabilities or congenital anomalies. *Am J Hum*
686 *Genet* 86, 749-764.
- 687 9. Alkan, C., Coe, B.P., and Eichler, E.E. (2011). Genome structural variation discovery
688 and genotyping. *Nat Rev Genet* 12, 363-376.
- 689 10. Schinzel, A. (2001). Catalogue of unbalanced chromosome aberrations in man.
- 690 11. van Karnebeek, C.D., Jansweijer, M.C., Leenders, A.G., Offringa, M., and Hennekam,
691 R.C. (2005). Diagnostic investigations in individuals with mental retardation: a
692 systematic literature review of their usefulness. *Eur J Hum Genet* 13, 6-25.
- 693 12. de Vries, B.B., Pfundt, R., Leisink, M., Koolen, D.A., Vissers, L.E., Janssen, I.M.,
694 Reijmersdal, S., Nillesen, W.M., Huys, E.H., Leeuw, N., et al. (2005). Diagnostic
695 genome profiling in mental retardation. *Am J Hum Genet* 77, 606-616.

- 696 13. Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W.,
697 Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., et al.
698 (2014). Genome sequencing identifies major causes of severe intellectual
699 disability. *Nature* 511, 344-347.
- 700 14. Lionel, A.C., Costain, G., Monfared, N., Walker, S., Reuter, M.S., Hosseini, S.M.,
701 Thiruvahindrapuram, B., Merico, D., Jobling, R., Nalpathamkalam, T., et al.
702 (2018). Improved diagnostic yield compared with targeted gene sequencing
703 panels suggests a role for whole-genome sequencing as a first-tier genetic test.
704 *Genet Med* 20, 435-443.
- 705 15. Stavropoulos, D.J., Merico, D., Jobling, R., Bowdin, S., Monfared, N.,
706 Thiruvahindrapuram, B., Nalpathamkalam, T., Pellecchia, G., Yuen, R.K.C.,
707 Szego, M.J., et al. (2016). Whole Genome Sequencing Expands Diagnostic Utility
708 and Improves Clinical Management in Pediatric Medicine. *NPJ Genom Med* 1.
- 709 16. Chaisson, M.J.P., Sanders, A.D., Zhao, X., Malhotra, A., Porubsky, D., Rausch, T.,
710 Gardner, E.J., Rodriguez, O.L., Guo, L., Collins, R.L., et al. (2019). Multi-platform
711 discovery of haplotype-resolved structural variation in human genomes. *Nat*
712 *Commun* 10, 1784.
- 713 17. Mantere, T., Kersten, S., and Hoischen, A. (2019). Long-Read Sequencing Emerging
714 in Medical Genetics. *Front Genet* 10, 426.
- 715 18. Merker, J.D., Wenger, A.M., Sneddon, T., Grove, M., Zappala, Z., Fresard, L.,
716 Waggott, D., Utiramerur, S., Hou, Y., Smith, K.S., et al. (2018). Long-read
717 genome sequencing identifies causal structural variation in a Mendelian disease.
718 *Genet Med* 20, 159-163.
- 719 19. Mizuguchi, T., Suzuki, T., Abe, C., Umemura, A., Tokunaga, K., Kawai, Y.,
720 Nakamura, M., Nagasaki, M., Kinoshita, K., Okamura, Y., et al. (2019). A 12-kb
721 structural variation in progressive myoclonic epilepsy was newly identified by
722 long-read whole-genome sequencing. *J Hum Genet* 64, 359-368.
- 723 20. Schwartz, D.C., Li, X., Hernandez, L.I., Ramnarain, S.P., Huff, E.J., and Wang, Y.K.
724 (1993). Ordered restriction maps of *Saccharomyces cerevisiae* chromosomes
725 constructed by optical mapping. *Science* 262, 110-114.
- 726 21. Lam, E.T., Hastie, A., Lin, C., Ehrlich, D., Das, S.K., Austin, M.D., Deshpande, P.,
727 Cao, H., Nagarajan, N., Xiao, M., et al. (2012). Genome mapping on nanochannel
728 arrays for structural variation analysis and sequence assembly. *Nat Biotechnol*
729 30, 771-776.
- 730 22. Chan, S., Lam, E., Saghbin, M., Bocklandt, S., Hastie, A., Cao, H., Holmlin, E., and
731 Borodkin, M. (2018). Structural Variation Detection and Analysis Using Bionano
732 Optical Mapping. *Methods Mol Biol* 1833, 193-203.
- 733 23. Wang, M., Tu, L., Yuan, D., Zhu, Shen, C., Li, J., Liu, F., Pei, L., Wang, P., Zhao, G.,
734 et al. (2019). Reference genome sequences of two cultivated allotetraploid
735 cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat Genet* 51, 224-
736 229.
- 737 24. Kronenberg, Z.N., Fiddes, I.T., Gordon, D., Murali, S., Cantsilieris, S., Meyerson,
738 O.S., Underwood, J.G., Nelson, B.J., Chaisson, M.J.P., Dougherty, M.L., et al.
739 (2018). High-resolution comparative analysis of great ape genomes. *Science* 360.
- 740 25. Nowoshilow, S., Schloissnig, S., Fei, J.F., Dahl, A., Pang, A.W.C., Pippel, M.,
741 Winkler, S., Hastie, A.R., Young, G., Roscito, J.G., et al. (2018). The axolotl
742 genome and the evolution of key tissue formation regulators. *Nature* 554, 50-55.
- 743 26. Neveling, K., Mantere, T., Vermeulen, S., Oorsprong, M., van Beek, R., Kater-Baats,
744 E., Pauper, M., van der Zande, G., Smeets, D., Weghuis, D.O., et al. (2020). Next
745 generation cytogenetics: comprehensive assessment of 48 leukemia genomes by
746 genome imaging. *bioRxiv*, 2020.2002.2006.935742.
- 747 27. Barseghyan, H., Delot, E.C., and Vilain, E. (2018). New technologies to uncover the
748 molecular basis of disorders of sex development. *Mol Cell Endocrinol* 468, 60-69.
- 749 28. Du, C., Mark, D., Wappenschmidt, B., Bockmann, B., Pabst, B., Chan, S., Cao, H.,
750 Morlot, S., Scholz, C., Auber, B., et al. (2018). A tandem duplication of BRCA1

- 751 exons 1-19 through DHX8 exon 2 in four families with hereditary breast and
752 ovarian cancer syndrome. *Breast Cancer Res Treat* 172, 561-569.
- 753 29. Levy-Sakin, M., Pastor, S., Mostovoy, Y., Li, L., Leung, A.K.Y., McCaffrey, J., Young,
754 E., Lam, E.T., Hastie, A.R., Wong, K.H.Y., et al. (2019). Genome maps across 26
755 human populations reveal population-specific patterns of structural variation. *Nat*
756 *Commun* 10, 1025.
- 757 30. Bates, S.E. (2011). Classical cytogenetics: karyotyping techniques. *Methods Mol Biol*
758 767, 177-190.
- 759 31. [https://bionanogenomics.com/wp-content/uploads/2018/04/30110-Bionano-Solve-](https://bionanogenomics.com/wp-content/uploads/2018/04/30110-Bionano-Solve-Theory-of-Operation-Structural-Variant-Calling.pdf)
760 [Theory-of-Operation-Structural-Variant-Calling.pdf](https://bionanogenomics.com/wp-content/uploads/2018/04/30110-Bionano-Solve-Theory-of-Operation-Structural-Variant-Calling.pdf).
- 761 32. [https://bionanogenomics.com/wp-content/uploads/2018/04/30210-Introduction-to-](https://bionanogenomics.com/wp-content/uploads/2018/04/30210-Introduction-to-Copy-Number-Analysis.pdf)
762 [Copy-Number-Analysis.pdf](https://bionanogenomics.com/wp-content/uploads/2018/04/30210-Introduction-to-Copy-Number-Analysis.pdf).
- 763 33. Schluth-Bolard, C., Diguët, F., Chatron, N., Rollat-Farnier, P.A., Bardel, C., Afenjar,
764 A., Amblard, F., Amiel, J., Blesson, S., Callier, P., et al. (2019). Whole genome
765 paired-end sequencing elucidates functional and phenotypic consequences of
766 balanced chromosomal rearrangement in patients with developmental disorders. *J*
767 *Med Genet* 56, 526-535.
- 768 34. Weissensteiner, M.H., Pang, A.W.C., Bunikis, I., Hoijer, I., Vinnere-Petterson, O.,
769 Suh, A., and Wolf, J.B.W. (2017). Combination of short-read, long-read, and
770 optical mapping assemblies reveals large-scale tandem repeat arrays with
771 population genetic implications. *Genome Res* 27, 697-708.
- 772 35. Redin, C., Brand, H., Collins, R.L., Kammin, T., Mitchell, E., Hodge, J.C., Hanscom,
773 C., Pillalamarri, V., Seabra, C.M., Abbott, M.A., et al. (2017). The genomic
774 landscape of balanced cytogenetic abnormalities associated with human
775 congenital anomalies. *Nat Genet* 49, 36-45.
- 776 36. Dong, Z., Wang, H., Chen, H., Jiang, H., Yuan, J., Yang, Z., Wang, W.J., Xu, F., Guo,
777 X., Cao, Y., et al. (2018). Identification of balanced chromosomal rearrangements
778 previously unknown among participants in the 1000 Genomes Project:
779 implications for interpretation of structural variation in genomes and the future of
780 clinical cytogenetics. *Genet Med* 20, 697-707.
- 781 37. Kosugi, S., Momozawa, Y., Liu, X., Terao, C., Kubo, M., and Kamatani, Y. (2019).
782 Comprehensive evaluation of structural variation detection algorithms for whole
783 genome sequencing. *Genome Biol* 20, 117.
- 784 38. Monlong, J., Cossette, P., Meloche, C., Rouleau, G., Girard, S.L., and Bourque, G.
785 (2018). Human copy number variants are enriched in regions of low mappability.
786 *Nucleic Acids Res* 46, 7236-7249.
- 787 39. Luo, F. (2019). A systematic evaluation of copy number alterations detection methods
788 on real SNP array and deep sequencing data. *BMC Bioinformatics* 20, 692.
- 789 40. Zhao, L., Liu, H., Yuan, X., Gao, K., and Duan, J. (2020). Comparative study of whole
790 exome sequencing-based copy number variation detection tools. *BMC*
791 *Bioinformatics* 21, 97.
- 792 41. Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks,
793 S., Howe, E., Porubsky, D., Logsdon, G.A., et al. (2020). Telomere-to-telomere
794 assembly of a complete human X chromosome. *Nature*.
- 795 42. Lee, K.S., Choi, M., Kwon, D.W., Kim, D., Choi, J.M., Kim, A.K., Ham, Y., Han, S.B.,
796 Cho, S., and Cheon, C.K. (2020). A novel de novo heterozygous DYRK1A
797 mutation causes complete loss of DYRK1A function and developmental delay. *Sci*
798 *Rep* 10, 9849.
- 799 43. van Bon, B.W., Coe, B.P., Bernier, R., Green, C., Gerdts, J., Witherspoon, K.,
800 Kleefstra, T., Willemsen, M.H., Kumar, R., Bosco, P., et al. (2016). Disruptive de
801 novo mutations of DYRK1A lead to a syndromic form of autism and ID. *Mol*
802 *Psychiatry* 21, 126-132.
- 803 44. Ji, J., Lee, H., Argiropoulos, B., Dorrani, N., Mann, J., Martinez-Agosto, J.A., Gomez-
804 Ospina, N., Gallant, N., Bernstein, J.A., Hudgins, L., et al. (2015). DYRK1A
805 haploinsufficiency causes a new recognizable syndrome with microcephaly,

- 806 intellectual disability, speech impairment, and distinct facies. *Eur J Hum Genet* 23,
807 1473-1481.
- 808 45. Coe, B.P., Witherspoon, K., Rosenfeld, J.A., van Bon, B.W., Vulto-van Silfhout, A.T.,
809 Bosco, P., Friend, K.L., Baker, C., Bueno, S., Vissers, L.E., et al. (2014). Refining
810 analyses of copy number variation identifies specific genes associated with
811 developmental delay. *Nat Genet* 46, 1063-1071.
- 812 46. Koolen, D.A., Kramer, J.M., Neveling, K., Nillesen, W.M., Moore-Barton, H.L.,
813 Elmslie, F.V., Toutain, A., Amiel, J., Malan, V., Tsai, A.C., et al. (2012). Mutations
814 in the chromatin modifier gene *KANSL1* cause the 17q21.31 microdeletion
815 syndrome. *Nat Genet* 44, 639-641.
- 816 47. Demaerel, W., Mostovoy, Y., Yilmaz, F., Vervoort, L., Pastor, S., Hestand, M.S.,
817 Swillen, A., Vergaelen, E., Geiger, E.A., Coughlin, C.R., et al. (2019). The 22q11
818 low copy repeats are characterized by unprecedented size and structural
819 variability. *Genome Res* 29, 1389-1401.
- 820 48. Chaisson, M.J., Wilson, R.K., and Eichler, E.E. (2015). Genetic variation and the de
821 novo assembly of human genomes. *Nat Rev Genet* 16, 627-640.
- 822 49. Cameron, D.L., Di Stefano, L., and Papenfuss, A.T. (2019). Comprehensive
823 evaluation and characterisation of short read general-purpose structural variant
824 calling software. *Nat Commun* 10, 3240.
- 825 50. Mahmoud, M., Gobet, N., Cruz-Davalos, D.I., Mounier, N., Dessimoz, C., and
826 Sedlazeck, F.J. (2019). Structural variant calling: the long and the short of it.
827 *Genome Biol* 20, 246.
- 828 51. Cretu Stancu, M., van Roosmalen, M.J., Renkens, I., Nieboer, M.M., Middelkamp, S.,
829 de Ligt, J., Pregno, G., Giachino, D., Mandrile, G., Espejo Valle-Inclan, J., et al.
830 (2017). Mapping and phasing of structural variation in patient genomes using
831 nanopore sequencing. *Nat Commun* 8, 1326.
- 832 52. Zheng, Y., Kong, L., Xu, H., Lu, Y., Zhao, X., Yang, Y., Yu, G., Li, P., Liang, F., Jin,
833 H., et al. (2020). Rapid prenatal diagnosis of Facioscapulohumeral Muscular
834 Dystrophy 1 by combined Bionano optical mapping and karyomapping. *Prenat*
835 *Diagn* 40, 317-323.
- 836 53. Dai, Y., Li, P., Wang, Z., Liang, F., Yang, F., Fang, L., Huang, Y., Huang, S., Zhou, J.,
837 Wang, D., et al. (2020). Single-molecule optical mapping enables quantitative
838 measurement of D4Z4 repeats in facioscapulohumeral muscular dystrophy
839 (FSHD). *J Med Genet* 57, 109-120.
- 840

Figure 1: Description of the study population (n=85)

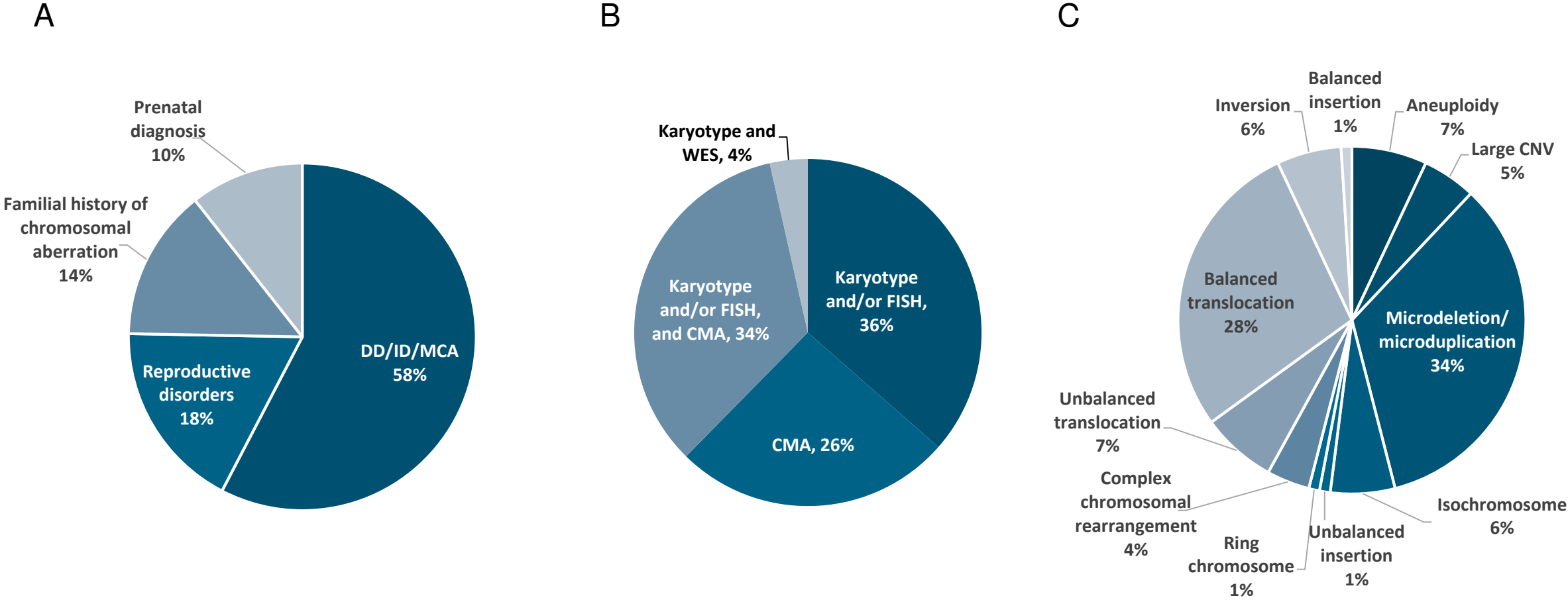


Figure 2: SV detection and filtering

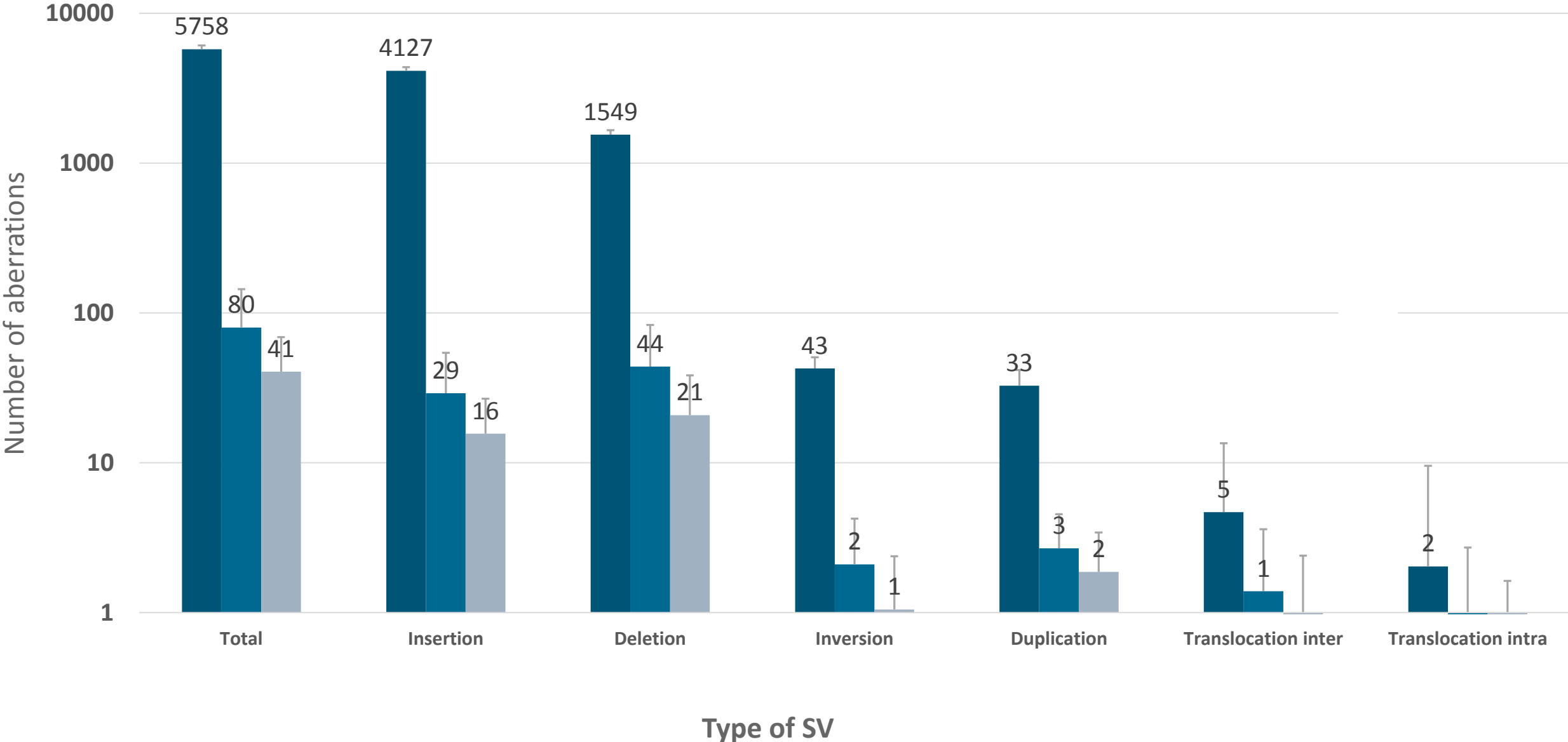


Figure 3: Visual representation of optical mapping data

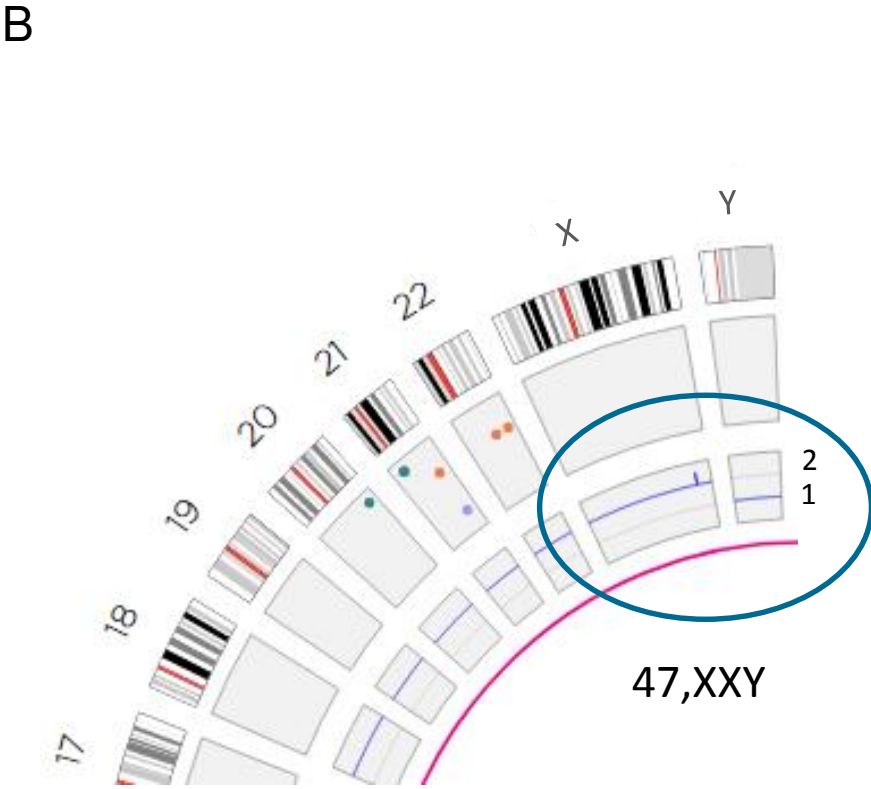
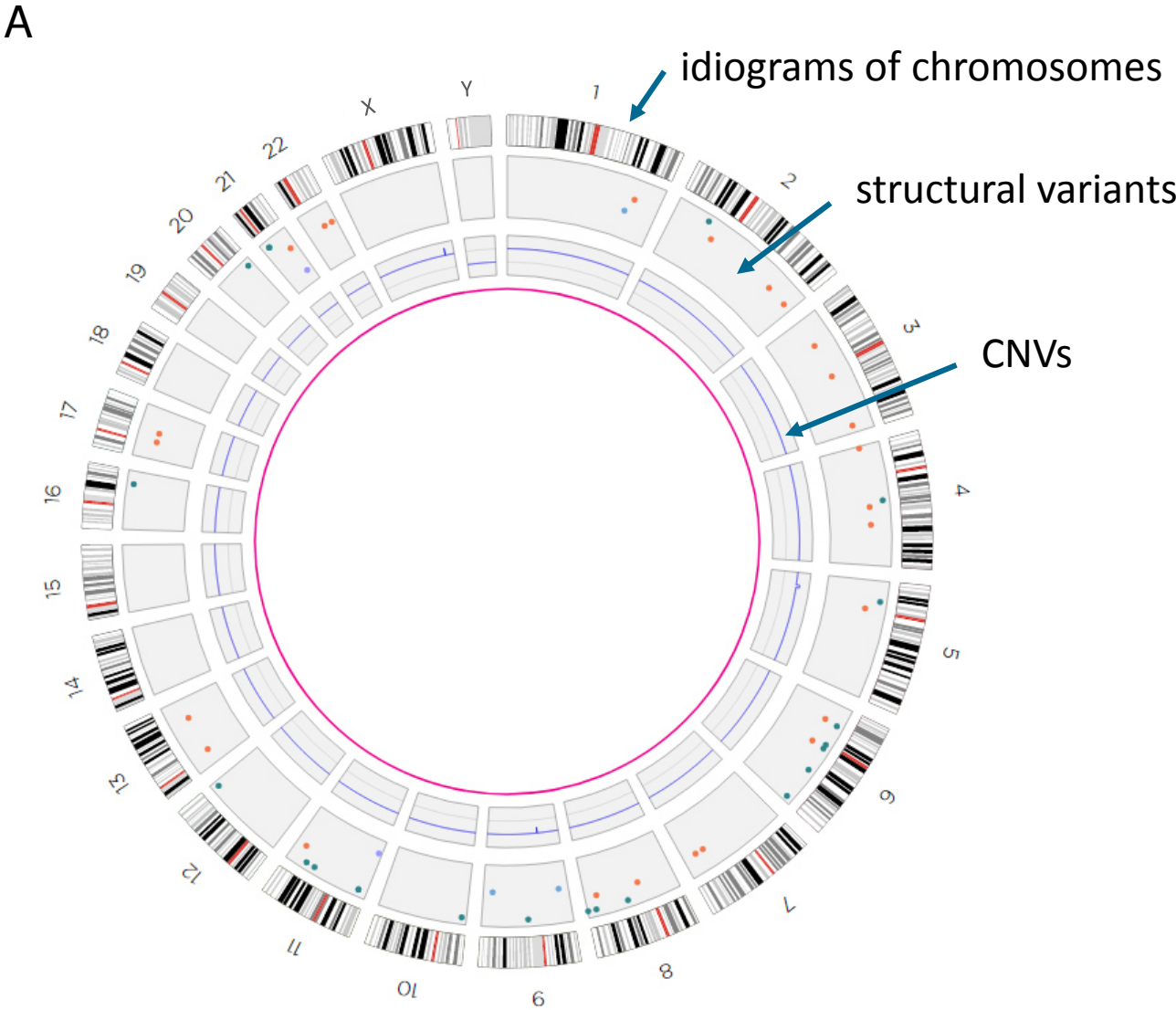
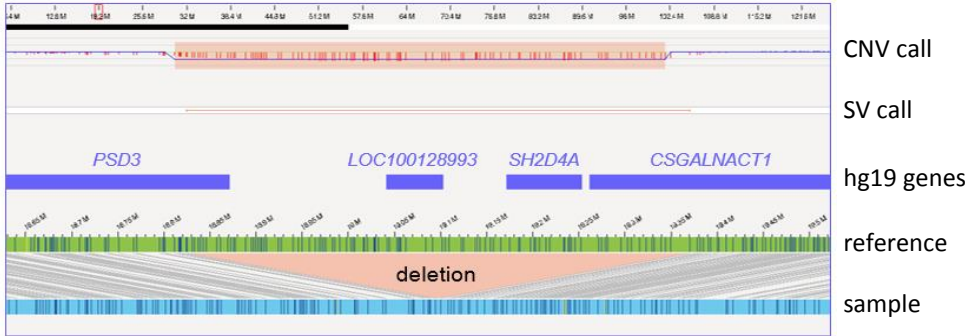
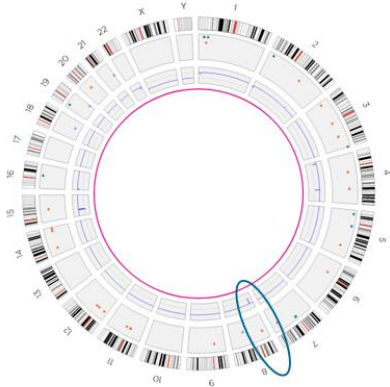
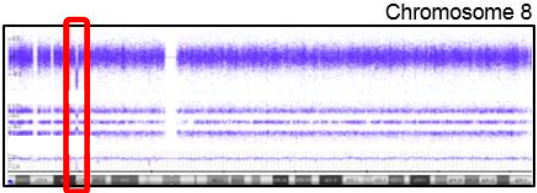
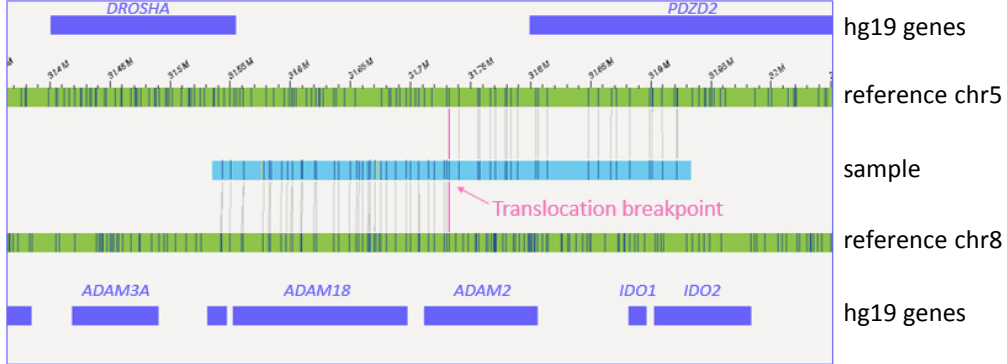
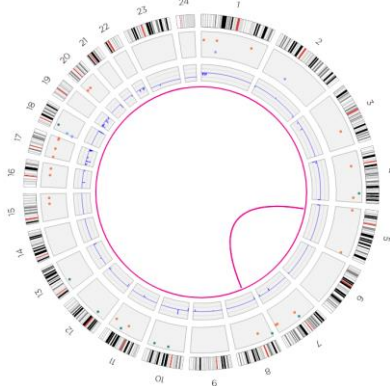


Figure 4: Representation of different chromosomal aberrations

A



B



C

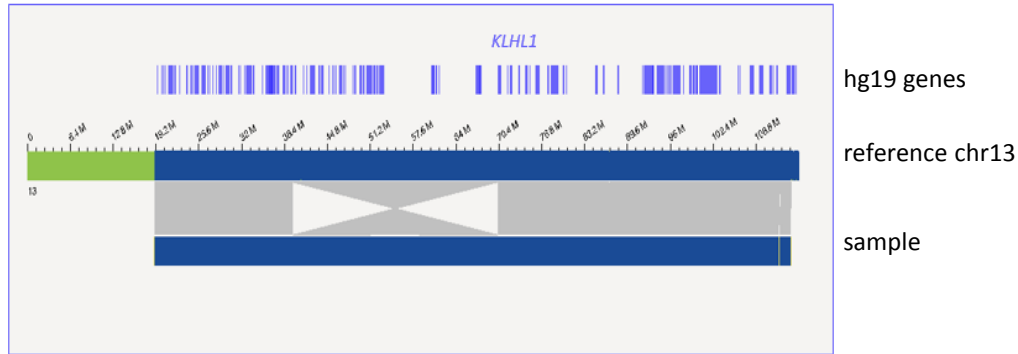
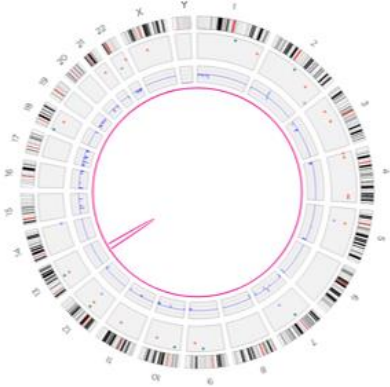
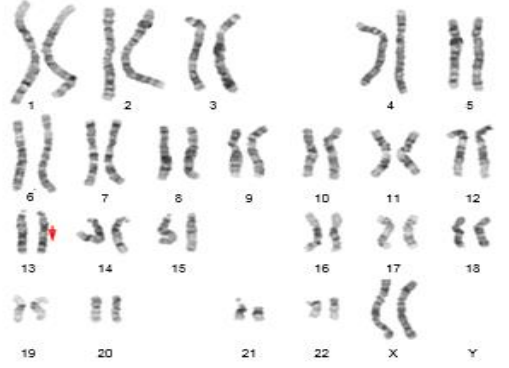


Figure 5: Isodicentric Y-chromosomes show specific bionano assembly map patterns

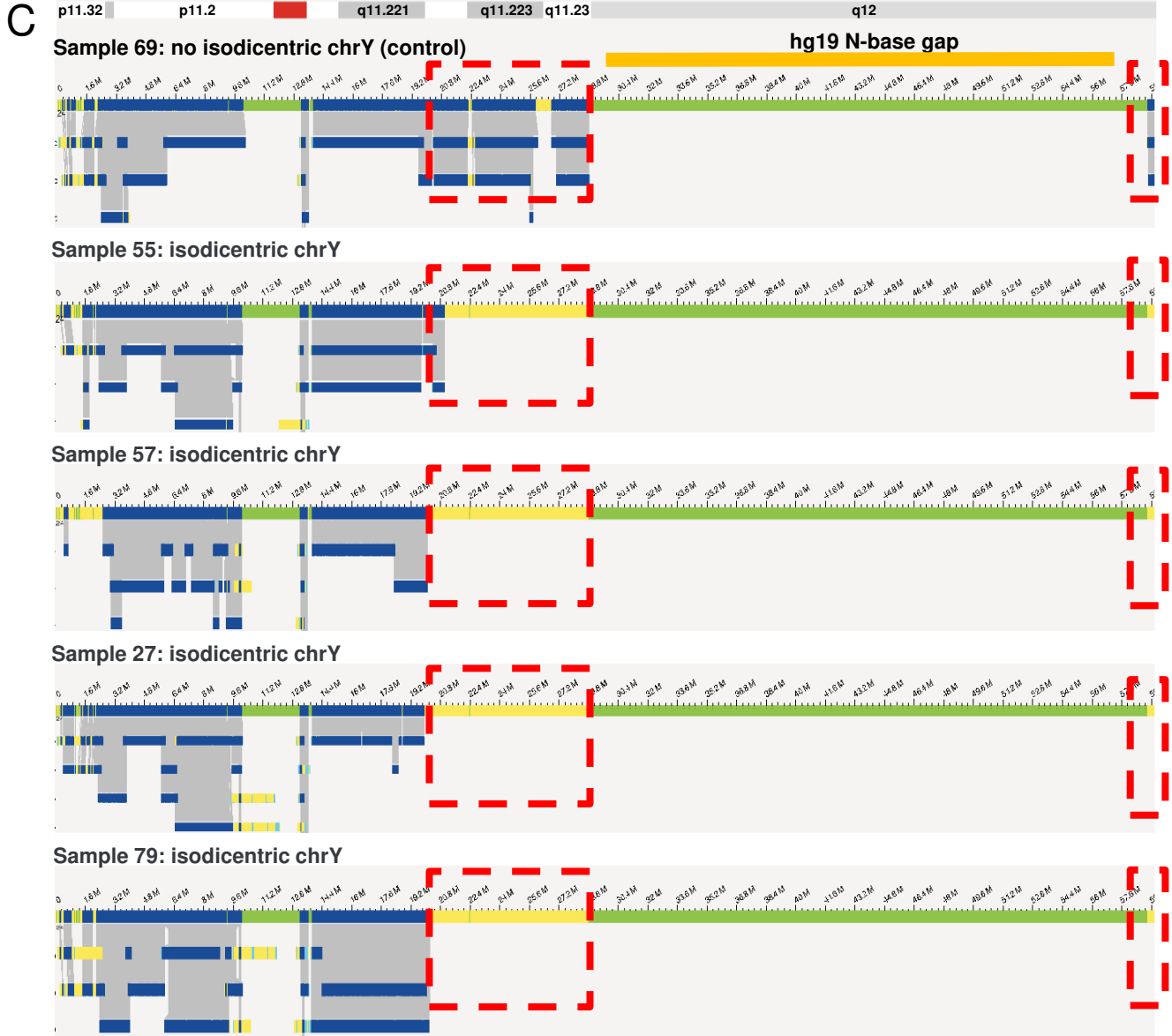
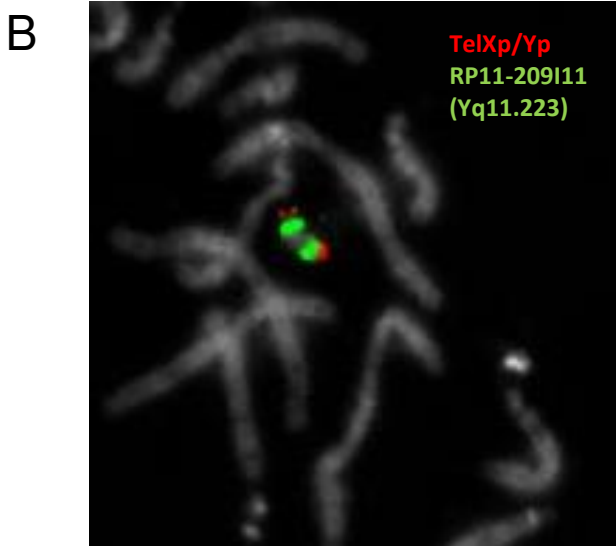
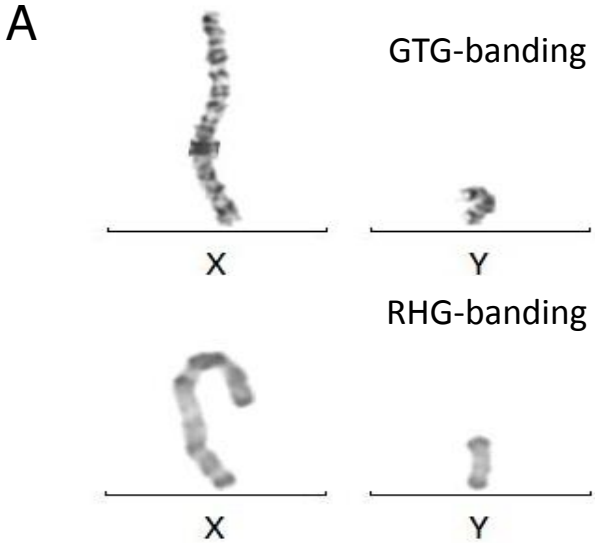


Figure 6: Small X ring chromosome

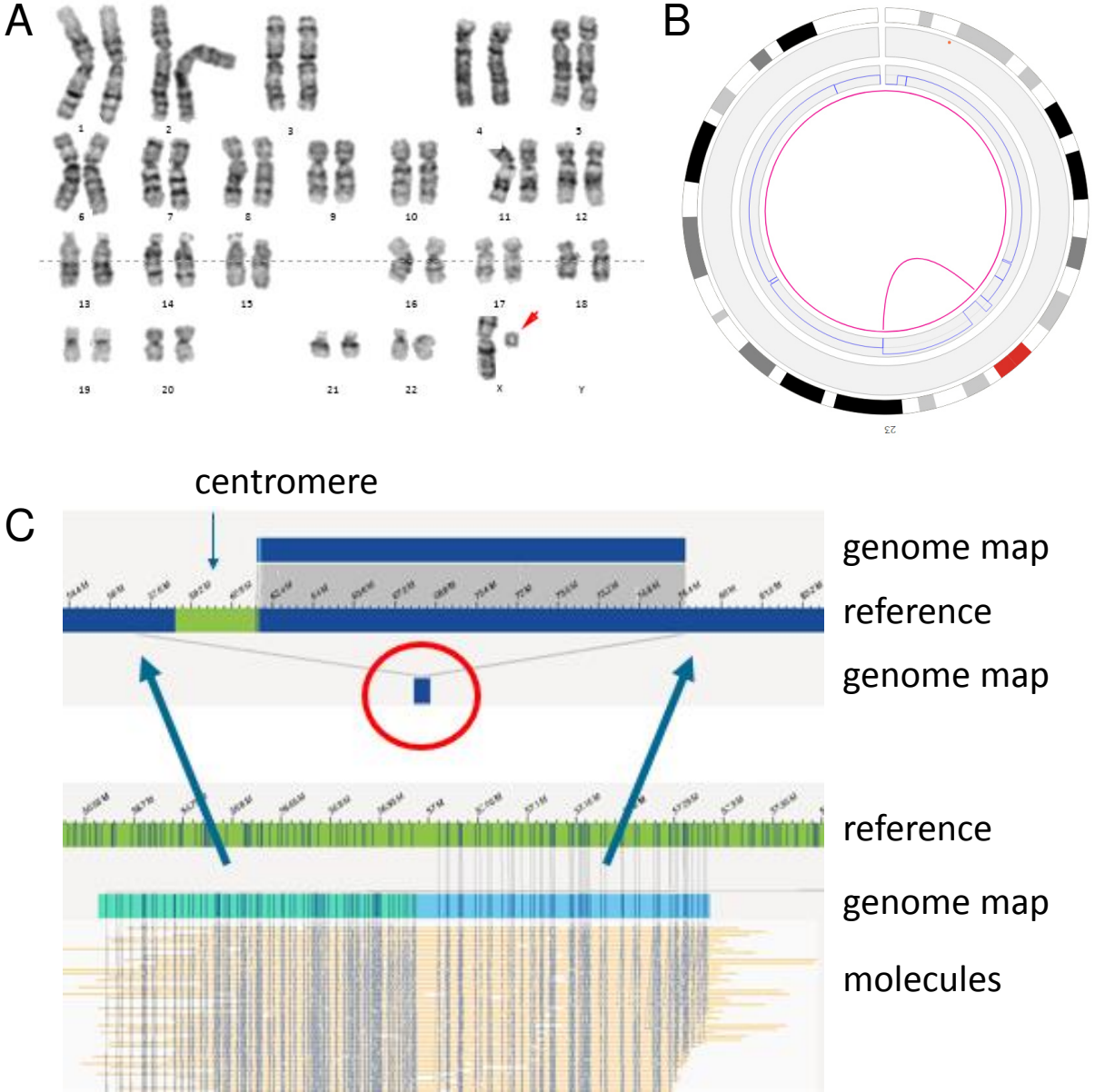


Figure 7: Examples of inversions and translocations interrupting well known disease causing genes

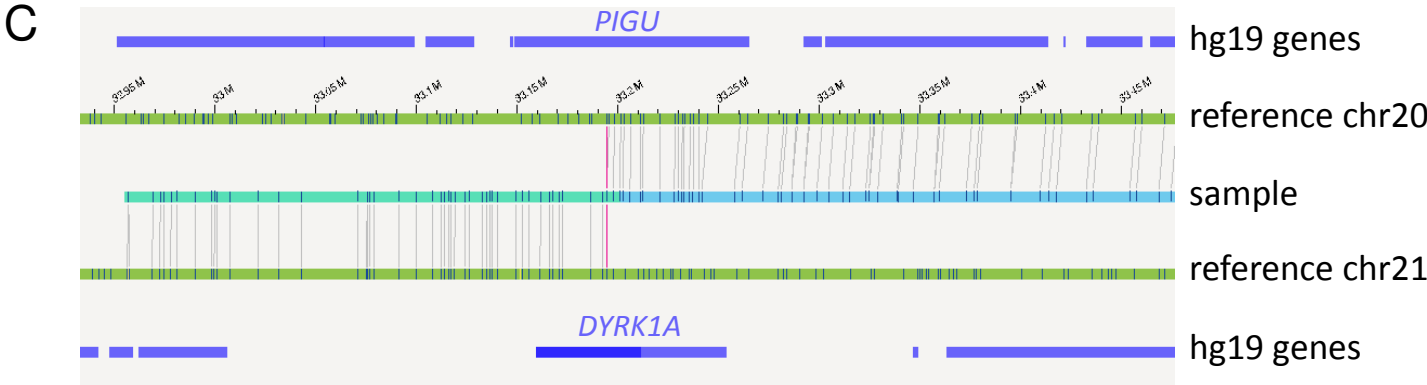
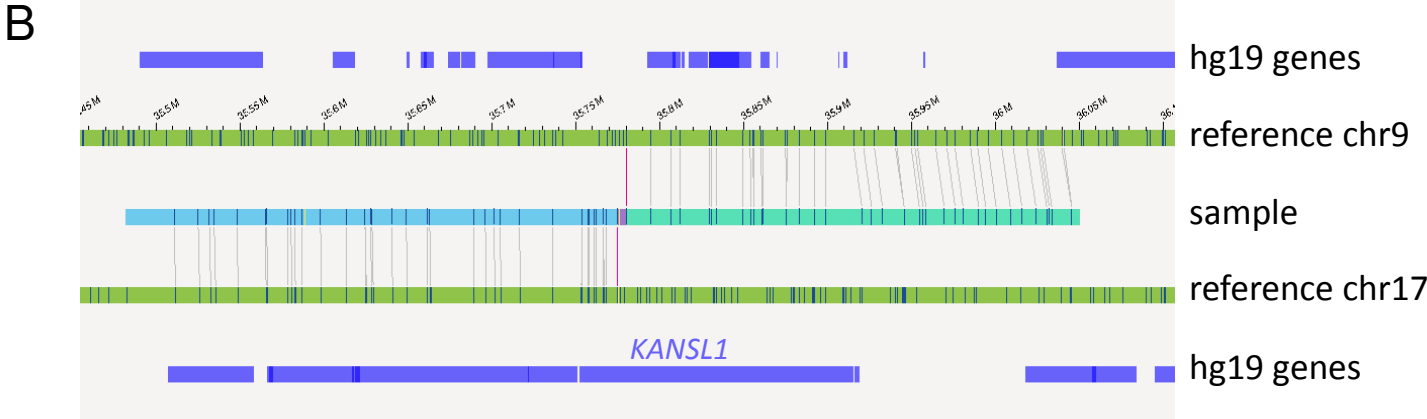
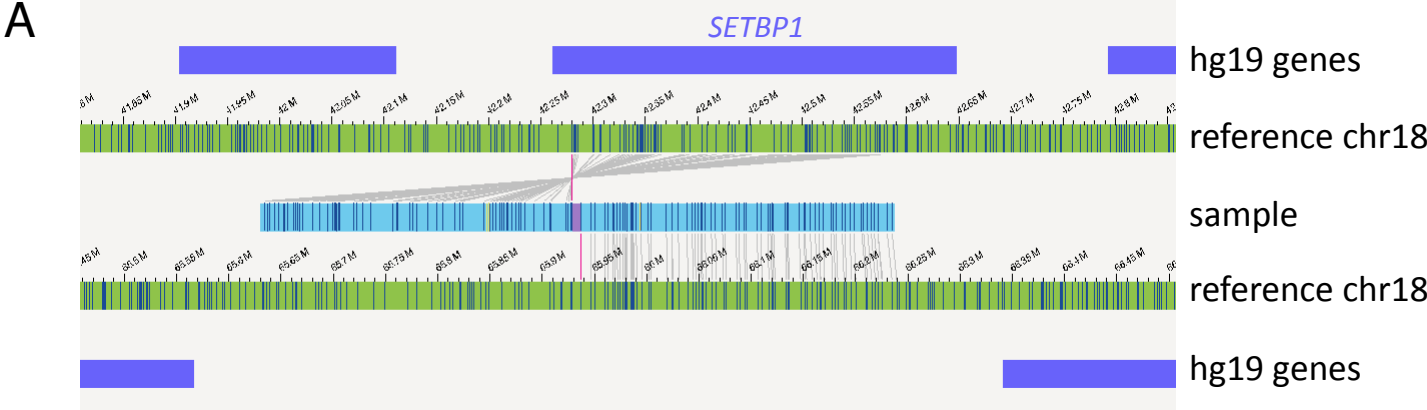


Figure 8: Example of a typical 22q11.2 microdeletion syndrome (VCF, Di-George syndrome)

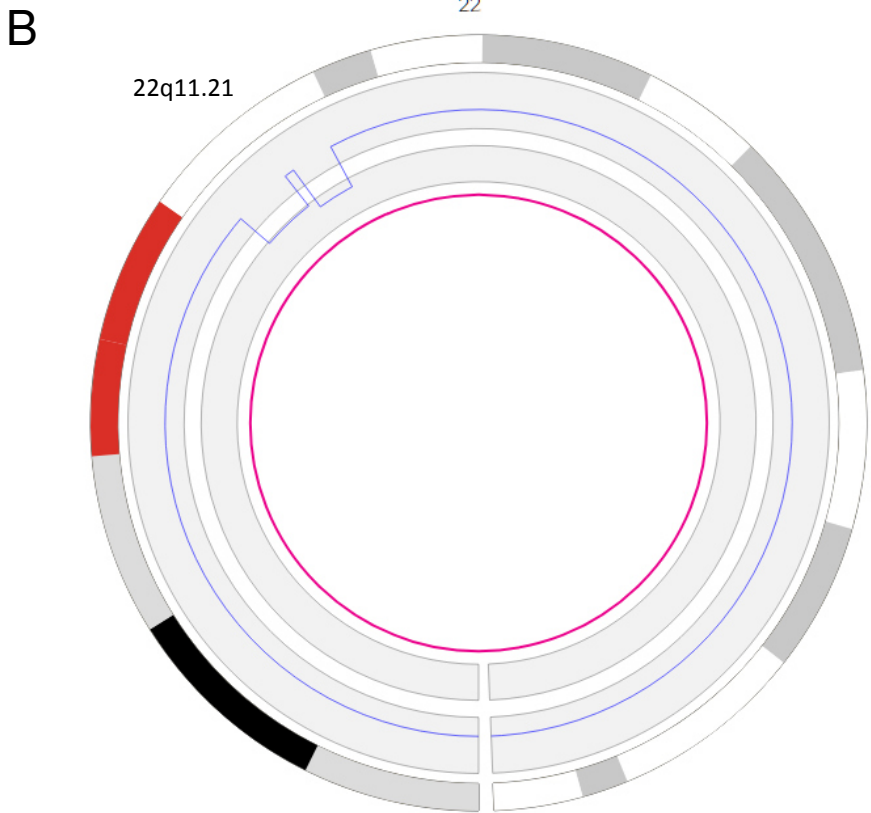
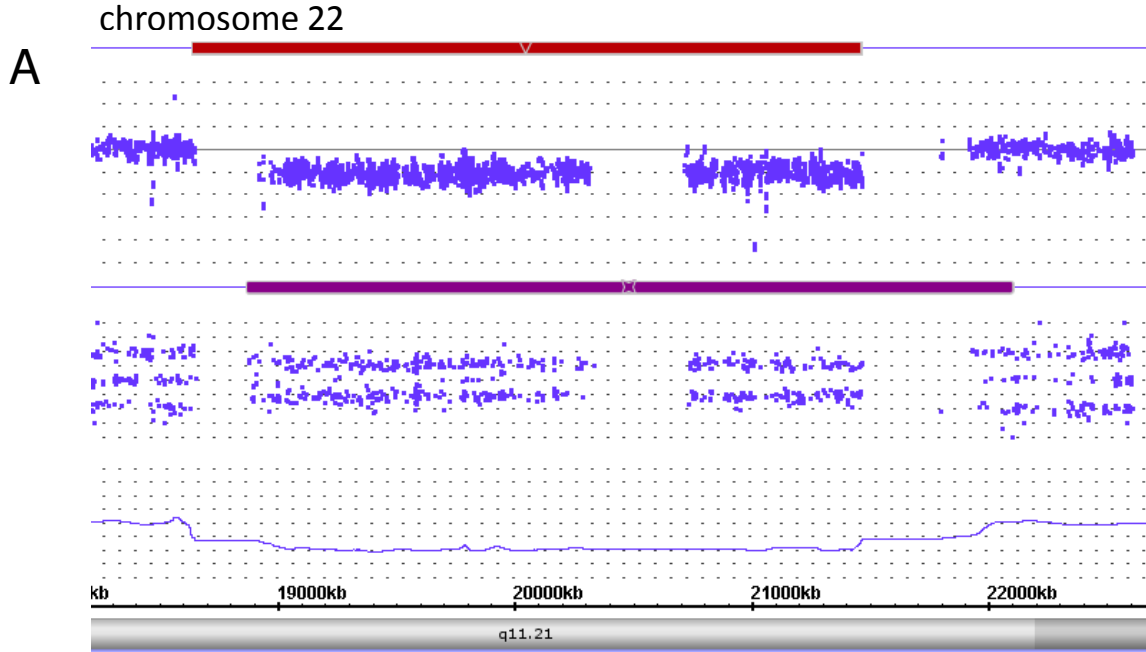


Figure 9: Complex sample 28

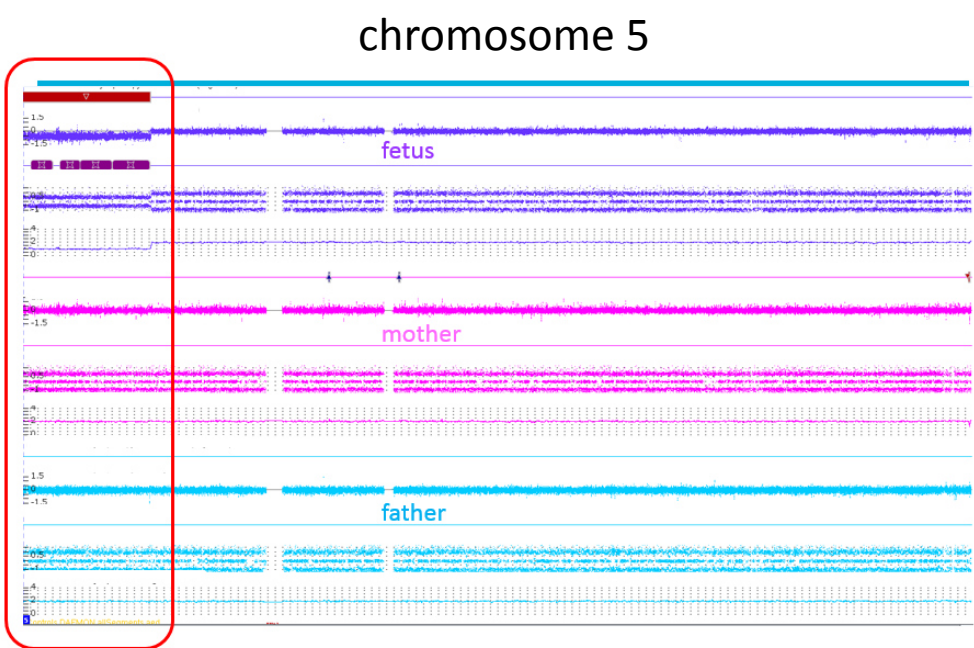
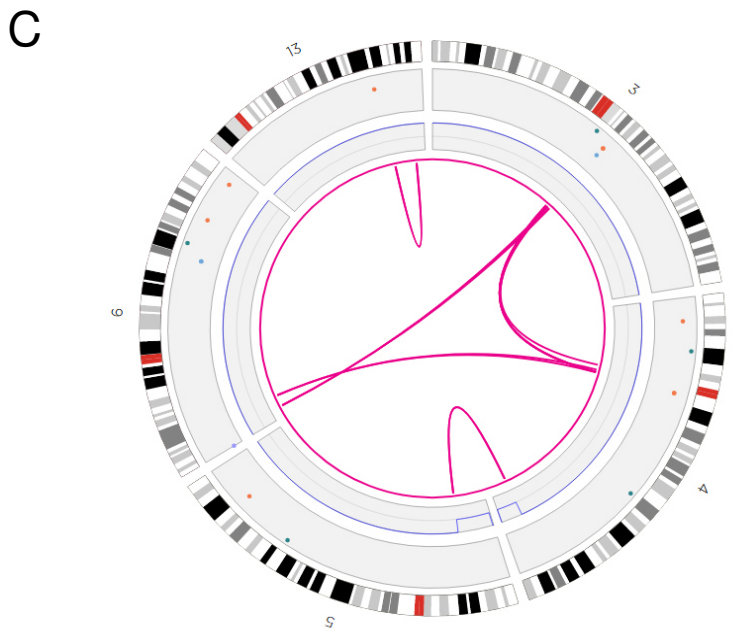
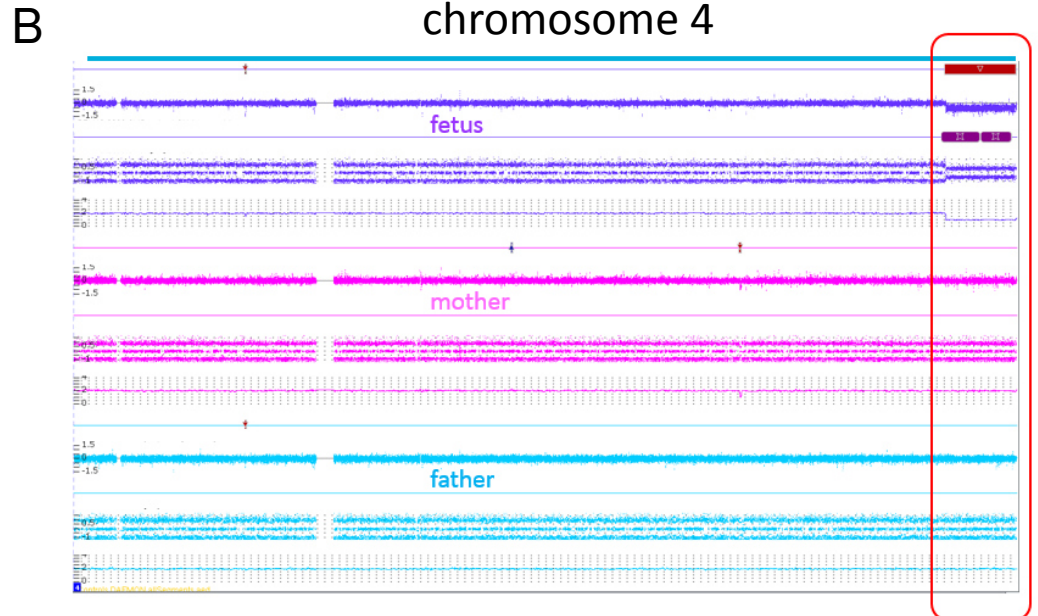
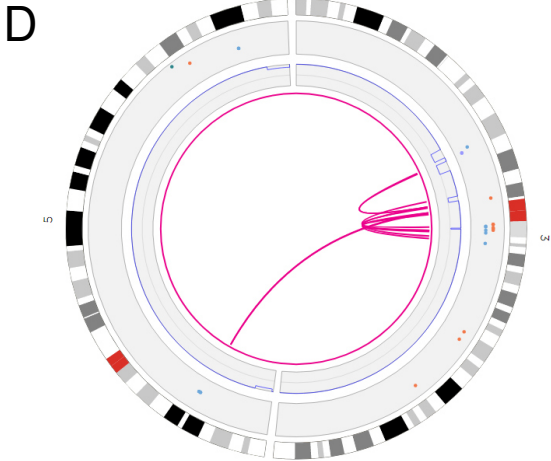
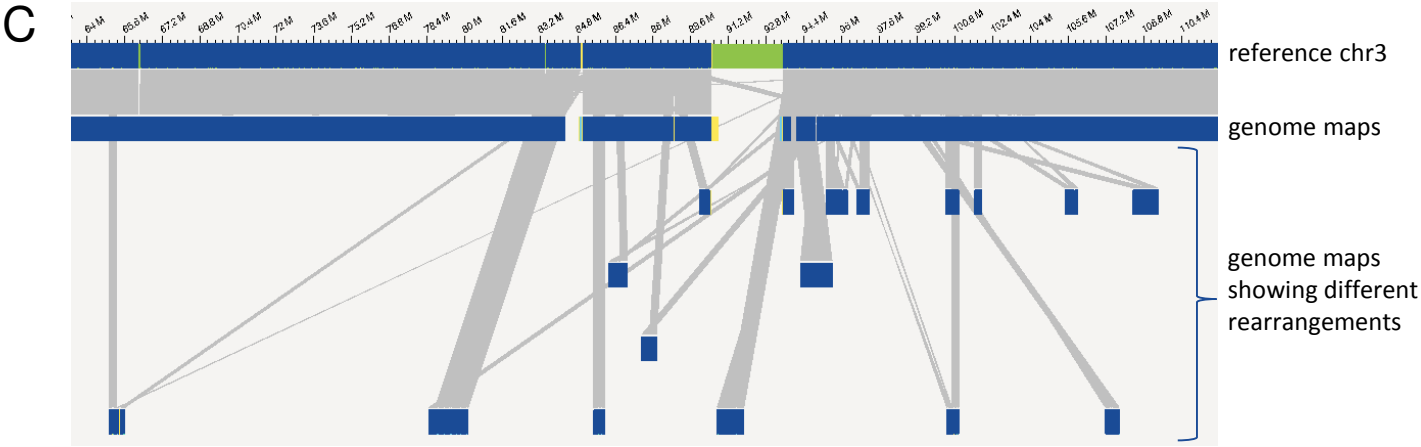
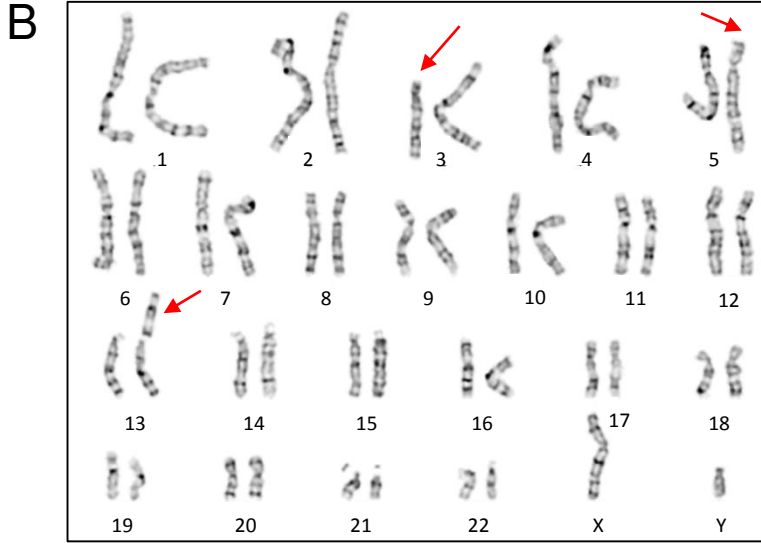
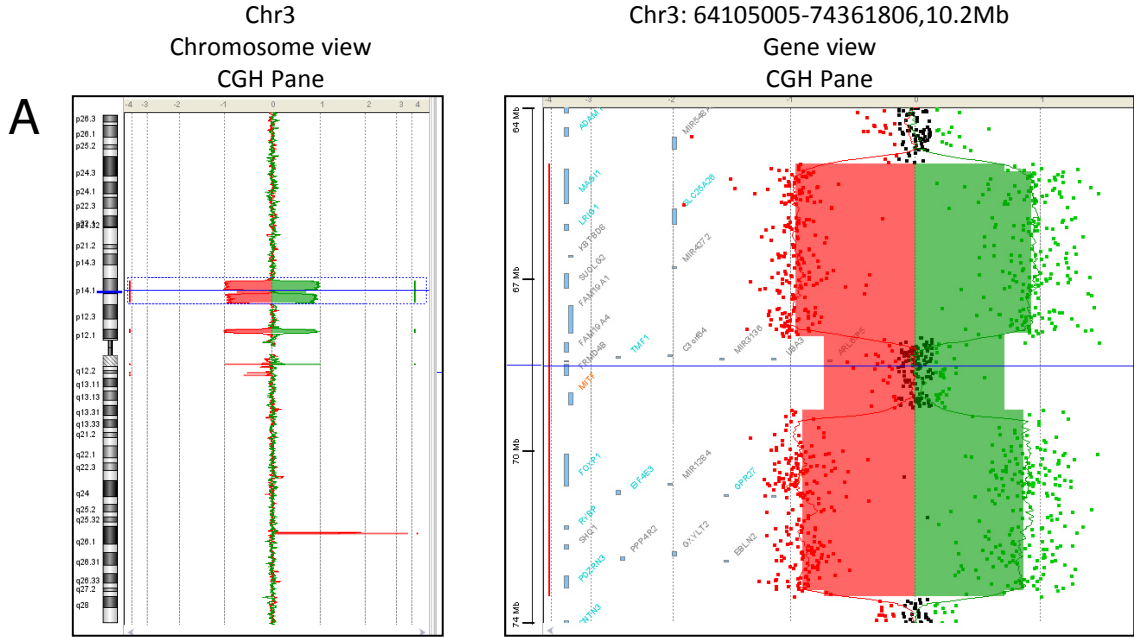


Figure 10: Complex sample 66



Ressources Web :

Integrative Genomics Viewer: <http://software.broadinstitute.org/software/igv/>

Human BLAT Search: <http://www.genome.ucsc.edu/cgi-bin/hgBlat>

GATK: <https://gatk.broadinstitute.org/hc/en-us>

SIFT : <https://sift.bii.a-star.edu.sg/>

PolyPhen-2: <http://genetics.bwh.harvard.edu/pph2/>

MutationTaster: <http://www.mutationtaster.org/>

GTEEx Portal : <https://gtexportal.org/home/>

gnomAD. <https://gnomad.broadinstitute.org/>

UCSC Genome Browser: <https://genome.ucsc.edu/>

ClinVar – NCBI : <https://www.ncbi.nlm.nih.gov/clinvar/variation/454741/>

3D Genome Browser : <http://promoter.bx.psu.edu/hi-c/>

OMIM: <https://www.omim.org/>

Expasy : <https://web.expasy.org/>

Mouse genome informatics: <http://www.informatics.jax.org/>

Orphanet: <https://www.orpha.net/>

9. Bibliographie

1. Dimassi S, Tilla M, Sanlaville D. Anomalies chromosomiques. *Journal de Pédiatrie et de Puériculture*. 1 nov 2017;30(5):249-70.
2. Scheuerlein H, Henschke F, Köckerling F. Wilhelm von Waldeyer-Hartz—A Great Forefather: His Contributions to Anatomy with Particular Attention to “His” Fascia. *Front Surg* [Internet]. 4 déc 2017 [cité 14 sept 2020];4. Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5723023/>
3. Wessel GM. The sub-culture of cell culture. *Molecular Reproduction and Development*. 2011;78(2):Fm i-Fm i.
4. GNN - Genetics and Genomics Timeline [Internet]. [cité 14 sept 2020]. Disponible sur: http://www.genomenewsnetwork.org/resources/timeline/1902_Boveri_Sutton.php
5. Abercrombie M. Ross Granville Harrison. 1870-1959. *Biographical Memoirs of Fellows of the Royal Society*. 1961;7:111-26.
6. Blakeslee AF, Avery AG. METHODS OF INDUCING DOUBLING OF CHROMOSOMES IN PLANTS By Treatment With Colchicine. *J Hered*. 1 déc 1937;28(12):393-411.
7. Hsu TC. The Hypotonic Miracle. In: Hsu TC, éditeur. *Human and Mammalian Cytogenetics: An Historical Perspective* [Internet]. New York, NY: Springer; 1979 [cité 14 sept 2020]. p. 15-24. (Heidelberg Science Library). Disponible sur: https://doi.org/10.1007/978-1-4612-6159-9_3
8. Tjio JH, Levan A. The Chromosome Number of Man. *Hereditas*. 1956;42(1-2):1-6.
9. Caspersson T, Castleman KR, Lomakka G, Modest EJ, Moller A, Nathan R, et al. Automatic karyotyping of quinacrine mustard stained human chromosomes. *Exp Cell Res*. juill 1971;67(1):233-5.
10. Pathak S, Arrighi FE. Loss of DNA following C-banding procedures. *Cytogenet Cell Genet*. 1973;12(6):414-22.
11. Dutrillaux B, Finaz C, de Grouchy J, Lejeune J. Comparison of banding patterns of human chromosomes obtained with heating, fluorescence, and proteolytic digestion. *Cytogenetics*. 1972;11(2):113-6.
12. Dutrillaux B. [New system of chromosome banding: the T bands (author's transl)]. *Chromosoma*. 27 avr 1973;41(4):395-402.
13. Yunis JJ, Chandler ME. High-resolution chromosome analysis in clinical medicine. *Prog Clin Pathol*. 1978;7:267-88.
14. Gall JG, Pardue ML. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc Natl Acad Sci USA*. juin 1969;63(2):378-83.

15. Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Döhner H, et al. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer*. déc 1997;20(4):399-407.
16. Schluth-Bolard C, Diguët F, Chatron N, Rollat-Farnier P-A, Bardel C, Afenjar A, et al. Whole genome paired-end sequencing elucidates functional and phenotypic consequences of balanced chromosomal rearrangement in patients with developmental disorders. *J Med Genet*. 2019;56(8):526-35.
17. Kloosterman WP, Guryev V, van Roosmalen M, Duran KJ, de Bruijn E, Bakker SCM, et al. Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. *Hum Mol Genet*. 15 mai 2011;20(10):1916-24.
18. Chromosomal Abnormalities [Internet]. [cité 18 sept 2020]. Disponible sur: <http://atlasgeneticsoncology.org/Educ/PolyMecaFr.html>
19. Sanlaville D, Turleau C. Types, fréquences et mécanismes de formation des anomalies chromosomiques. :23.
20. Liehr T, Claussen U, Starke H. Small supernumerary marker chromosomes (sSMC) in humans. *Cytogenet Genome Res*. 2004;107(1-2):55-67.
21. Birch J, Zomerdijk J. Structure and function of ribosomal RNA gene chromatin. *Biochemical Society transactions*. 1 août 2008;36:619-24.
22. Pellestor F, Anahory T, Lefort G, Puechberty J, Liehr T, Hédon B, et al. Complex chromosomal rearrangements: origin and meiotic behavior. *Hum Reprod Update*. août 2011;17(4):476-94.
23. Fredga K, Hall B. A complex familial translocation involving chromosomes 5, 9 and 13. *CGR*. 1970;9(4):294-306.
24. Kausch K, Haaf T, Köhler J, Schmid M. Complex chromosomal rearrangement in a woman with multiple miscarriages. *Am J Med Genet*. oct 1988;31(2):415-20.
25. Gatinois V, Puechberty J, Lefort G, Geneviève D, Pellestor F. Les remaniements chromosomiques complexes - Un paradigme pour l'étude de l'instabilité chromosomique. *Med Sci (Paris)*. 1 janv 2014;30(1):55-63.
26. Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development: Cell [Internet]. [cité 31 juill 2020]. Disponible sur: [https://www.cell.com/fulltext/S0092-8674\(10\)01377-2](https://www.cell.com/fulltext/S0092-8674(10)01377-2)
27. Kloosterman WP, Guryev V, van Roosmalen M, Duran KJ, de Bruijn E, Bakker SCM, et al. Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. *Hum Mol Genet*. 15 mai 2011;20(10):1916-24.
28. Constitutional Chromothripsis Rearrangements Involve Clustered Double-Stranded DNA Breaks and Nonhomologous Repair Mechanisms: Cell Reports [Internet]. [cité 4 oct 2020]. Disponible sur: [https://www.cell.com/cell-reports/fulltext/S2211-1247\(12\)00134-9?code=cell-site](https://www.cell.com/cell-reports/fulltext/S2211-1247(12)00134-9?code=cell-site)

29. Pellestor F, Gatinois V. Chromoanasythesis: another way for the formation of complex chromosomal abnormalities in human reproduction. *Hum Reprod.* 01 2018;33(8):1381-7.
30. Chromoanasythesis: another way for the formation of complex chromosomal abnormalities in human reproduction | Human Reproduction | Oxford Academic [Internet]. [cité 4 oct 2020]. Disponible sur: <https://academic.oup.com/humrep/article/33/8/1381/5062506>
31. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated Evolution of Prostate Cancer Genomes. *Cell.* 25 avr 2013;153(3):666-77.
32. Anderson ND, de Borja R, Young MD, Fuligni F, Rosic A, Roberts ND, et al. Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. *Science (New York, NY).* 31 2018;361(6405).
33. Shen MM. Chromoplexy: a new category of complex rearrangements in the cancer genome. *Cancer Cell.* 13 mai 2013;23(5):567-9.
34. (PDF) [Chromothripsis, an unexpected novel form of complexity for chromosomal rearrangements] [Internet]. ResearchGate. [cité 5 oct 2020]. Disponible sur: https://www.researchgate.net/publication/261255133_Chromothripsis_an_unexpected_novel_form_of_complexity_for_chromosomal_rearrangements
35. T R, Dt J, M Z, Am S, T Z, J W, et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations [Internet]. Vol. 148, *Cell.* Cell; 2012 [cité 5 oct 2020]. Disponible sur: <https://pubmed.ncbi.nlm.nih.gov/22265402/>
36. Maher CA, Wilson RK. Chromothripsis and human disease: piecing together the shattering process. *Cell.* 20 janv 2012;148(1-2):29-32.
37. Tubio JMC, Estivill X. Cancer: When catastrophe strikes a cell. *Nature.* 24 févr 2011;470(7335):476-7.
38. Forment JV, Kaidi A, Jackson SP. Chromothripsis and cancer: causes and consequences of chromosome shattering. *Nature Reviews Cancer.* oct 2012;12(10):663-70.
39. Rosin MP, German J. Evidence for chromosome instability in vivo in Bloom syndrome: increased numbers of micronuclei in exfoliated cells. *Human Genetics.* 1985;71(3):187-91.
40. Kort DH, Chia G, Treff NR, Tanaka AJ, Xing T, Vensand LB, et al. Human embryos commonly form abnormal nuclei during development: a mechanism of DNA damage, embryonic aneuploidy, and developmental arrest. *Hum Reprod.* 1 févr 2016;31(2):312-23.
41. Crasta K, Ganem NJ, Dagher R, Lantermann AB, Ivanova EV, Pan Y, et al. DNA breaks and chromosome pulverization from errors in mitosis. *Nature.* 18 janv 2012;482(7383):53-8.

42. Terradas M, Martín M, Genescà A. Detection of Impaired DNA Replication and Repair in Micronuclei as Indicators of Genomic Instability and Chromothripsis. *Methods in Molecular Biology* (Clifton, NJ). 2018;1769:197-208.
43. Chatron N, Giannuzzi G, Rollat-Farnier P-A, Diguët F, Porcu E, Yammine T, et al. The enrichment of breakpoints in late-replicating chromatin provides novel insights into chromoanagenesis mechanisms. *bioRxiv*. 17 juill 2020;2020.07.17.206771.
44. Wu Z-H. The concept and practice of Fanconi Anemia: from the clinical bedside to the laboratory bench. *Transl Pediatr*. juill 2013;2(3):112-9.
45. Spence JE, Perciaccante RG, Greig GM, Willard HF, Ledbetter DH, Hejtmancik JF, et al. Uniparental disomy as a mechanism for human genetic disease. *Am J Hum Genet*. févr 1988;42(2):217-26.
46. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. *Pathogenetics*. 3 nov 2008;1:4.
47. Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. Segmental Duplications: Organization and Impact Within the Current Human Genome Project Assembly. *Genome Res*. juin 2001;11(6):1005-17.
48. Barbouti A, Stankiewicz P, Nusbaum C, Cuomo C, Cook A, Höglund M, et al. The Breakpoint Region of the Most Common Isochromosome, i(17q), in Human Neoplasia Is Characterized by a Complex Genomic Architecture with Large, Palindromic, Low-Copy Repeats. *American Journal of Human Genetics*. janv 2004;74(1):1.
49. Jr L. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits [Internet]. Vol. 14, *Trends in genetics : TIG*. Trends Genet; 1998 [cité 18 sept 2020]. Disponible sur: <https://pubmed.ncbi.nlm.nih.gov/9820031/>
50. Cardoso AR, Oliveira M, Amorim A, Azevedo L. Major influence of repetitive elements on disease-associated copy number variants (CNVs). *Hum Genomics*. 23 sept 2016;10(1):30.
51. Lieber MR. The Mechanism of Human Nonhomologous DNA End Joining. *J Biol Chem*. 4 janv 2008;283(1):1-5.
52. Lee JA, Carvalho CMB, Lupski JR. A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders. *Cell*. 28 déc 2007;131(7):1235-47.
53. Branzei D, Foiani M. Template Switching: From Replication Fork Repair to Genome Rearrangements. *Cell*. 28 déc 2007;131(7):1228-30.
54. Hastings PJ, Ira G, Lupski JR. A Microhomology-Mediated Break-Induced Replication Model for the Origin of Human Copy Number Variation. *Matic I, éditeur. PLoS Genet*. 30 janv 2009;5(1):e1000327.
55. File:FISH (technique).gif. In: *Wikipedia* [Internet]. [cité 18 sept 2020]. Disponible sur: [https://en.wikipedia.org/wiki/File:FISH_\(technique\).gif](https://en.wikipedia.org/wiki/File:FISH_(technique).gif)

56. Kearney L. The impact of the new fish technologies on the cytogenetics of haematological malignancies. *Br J Haematol.* mars 1999;104(4):648-58.
57. Tsukamoto K, Tohma T, Ohta T, Yamakawa K, Fukushima Y, Nakamura Y, et al. Cloning and characterization of the inversion breakpoint at chromosome 2q35 in a patient with Waardenburg syndrome type I. *Hum Mol Genet.* août 1992;1(5):315-7.
58. Romana S, Malan V. *Cytogénétique moléculaire.* :25.
59. Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet.* oct 1998;20(2):207-11.
60. Malan V, Romana S. [Diagnosis of chromosomal abnormalities by array CGH in constitutional pathology: the end of the first-line karyotype]. *Arch Pediatr.* avr 2012;19(4):437-42.
61. Thorisson GA, Smith AV, Krishnan L, Stein LD. The International HapMap Project Web site. *Genome Res.* nov 2005;15(11):1592-3.
62. 10k_datasheet.pdf [Internet]. [cité 20 sept 2020]. Disponible sur: http://www.affymetrix.com/support/technical/datasheets/10k_datasheet.pdf
63. South ST, Lee C, Lamb AN, Higgins AW, Kearney HM, Working Group for the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee. ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. *Genet Med.* nov 2013;15(11):901-9.
64. Réseau AChro-Puce [Internet]. [cité 31 oct 2020]. Disponible sur: <http://acpa-achropuce.com>
65. 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.* 14 mai 2010;86(5):749-64.
66. Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST, Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med.* juill 2011;13(7):680-5.
67. De Braekeleer M, Dao TN. Cytogenetic studies in male infertility: a review. *Hum Reprod.* févr 1991;6(2):245-50.
68. De Gregori M, Ciccone R, Magini P, Pramparo T, Gimelli S, Messa J, et al. Cryptic deletions are a common finding in « balanced » reciprocal and complex chromosome rearrangements: a study of 59 patients. *J Med Genet.* déc 2007;44(12):750-62.
69. Schluth-Bolard C, Delobel B, Sanlaville D, Boute O, Cuisset J-M, Sukno S, et al. Cryptic genomic imbalances in de novo and inherited apparently balanced chromosomal

- rearrangements: Array CGH study of 47 unrelated cases. *European Journal of Medical Genetics*. sept 2009;52(5):291-6.
70. Lossi A, Laugier-Anfossi F, Depetris D, Gecz J, Gedeon A, Kooy F, et al. Abnormal expression of the KLF8 (ZNF741) gene in a female patient with an X;autosome translocation t(X;21)(p11.2;q22.3) and non-syndromic mental retardation. *J Med Genet*. févr 2002;39(2):113-7.
 71. Di Gregorio E, Bianchi FT, Schiavi A, Chiotto AMA, Rolando M, Verdun di Cantogno L, et al. A de novo X;8 translocation creates a PTK2-THOC2 gene fusion with THOC2 expression knockdown in a patient with psychomotor retardation and congenital cerebellar hypoplasia. *J Med Genet*. août 2013;50(8):543-51.
 72. Kleinjan DJ, van Heyningen V. Position effect in human genetic disease. *Hum Mol Genet*. 1998;7(10):1611-8.
 73. Posey JE, Harel T, Liu P, Rosenfeld JA, James RA, Coban Akdemir ZH, et al. Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. *New England Journal of Medicine*. 5 janv 2017;376(1):21-31.
 74. Young J. Infertilité masculine: mécanismes, causes et exploration. 2016;80:8.
 75. Griswold MD. Spermatogenesis: The Commitment to Meiosis. *Physiol Rev*. janv 2016;96(1):1-17.
 76. Allais-Bonnet A, Pailhoux E. Role of the prion protein family in the gonads. *Front Cell Dev Biol*. 2014;2:56.
 77. Clermont Y. The cycle of the seminiferous epithelium in man. *American Journal of Anatomy*. 1963;112(1):35-51.
 78. Wp P, Wz C. Coordinating the events of the meiotic prophase. *Trends Cell Biol*. 27 oct 2005;15(12):674-81.
 79. Bolcun-Filas E, Handel MA. Meiosis: the chromosomal foundation of reproduction. *Biol Reprod*. 01 2018;99(1):112-26.
 80. Inaba K. Molecular architecture of the sperm flagella: molecules for motility and signaling. *Zool Sci*. sept 2003;20(9):1043-56.
 81. Bao J, Bedford MT. Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. *Reproduction*. mai 2016;151(5):R55-70.
 82. Ahmadi A, Ng SC. Destruction of protamine in human sperm inhibits sperm binding and penetration in the zona-free hamster penetration test but increases sperm head decondensation and male pronuclear formation in the hamster-ICSI assay. *J Assist Reprod Genet*. mars 1999;16(3):128-32.
 83. Schagdarsurengin U, Paradowska A, Steger K. Analysing the sperm epigenome: roles in early embryogenesis and assisted reproduction. *Nat Rev Urol*. nov 2012;9(11):609-19.

84. Mao H-T, Yang W-X. Modes of acrosin functioning during fertilization. *Gene*. 10 sept 2013;526(2):75-9.
85. Inaba K. Molecular architecture of the sperm flagella: molecules for motility and signaling. *Zool J Linn Soc*. sept 2003;20(9):1043-56.
86. Ravel, C. 2009. Infertilités masculines [Internet]. Ravel, C., Berthaut, I., & Siffroi, J.-P. (2009). Infertilités masculines. *EMC - Endocrinologie - Nutrition*, 6(1), 1–18. doi:10.1016/s1155-1941(08)50295-2. [cité 26 août 2020]. Disponible sur: <https://www.em-consulte.com/article/198680/infertilites-masculines>
87. Bedford JM. Human spermatozoa and temperature: the elephant in the room. *Biol Reprod*. oct 2015;93(4):97.
88. Roux C, Berdin A, Amiot C. [Medically assisted procreation. Main biological, medical and ethical aspects]. *Rev Prat*. déc 2014;64(10):1435-44.
89. Ezech UI, Taub NA, Moore HD, Cooke ID. Establishment of predictive variables associated with testicular sperm retrieval in men with non-obstructive azoospermia. *Hum Reprod*. avr 1999;14(4):1005-12.
90. MacLEOD J, Gold RZ. The male factor in fertility and infertility. IV. Sperm morphology in fertile and infertile marriage. *Fertil Steril*. oct 1951;2(5):394-414.
91. Wharton IP, Chaudhry AH, French ME. A case of mumps epididymitis. *Lancet*. 25 févr 2006;367(9511):702.
92. Robin G, Marcelli F, Mitchell V, Marchetti C, Lemaitre L, Dewailly D, et al. [Why and how to assess hypospermia?]. *Gynecol Obstet Fertil*. oct 2008;36(10):1035-42.
93. WHO | WHO laboratory manual for the examination and processing of human semen [Internet]. WHO. World Health Organization; [cité 26 août 2020]. Disponible sur: <http://www.who.int/reproductivehealth/publications/infertility/9789241547789/en/>
94. Auger J, Eustache F. Standardisation de la classification morphologique des spermatozoïdes humains selon la méthode de David modifiée. :16.
95. Blanchard M, Haguenoer K, Apert A, Poret H, Barthélémy C, Royère D, et al. Sperm morphology assessment using David's classification: time to switch to strict criteria? Prospective comparative analysis in a selected IVF population. *Int J Androl*. avr 2011;34(2):145-52.
96. Menkveld R, Wong WY, Lombard CJ, Wetzels AM, Thomas CM, Merkus HM, et al. Semen parameters, including WHO and strict criteria morphology, in a fertile and subfertile population: an effort towards standardization of in-vivo thresholds. *Hum Reprod*. juin 2001;16(6):1165-71.
97. Venkatesh T, Suresh PS, Tsutsumi R. New insights into the genetic basis of infertility. *Appl Clin Genet*. 1 déc 2014;7:235-43.
98. Ravel C, Cecos and the GC of the FF of, Berthaut I, Cecos and the GC of the FF of, Bresson JL, Cecos and the GC of the FF of, et al. Prevalence of chromosomal

- abnormalities in phenotypically normal and fertile adult males: large-scale survey of over 10 000 sperm donor karyotypes. *Hum Reprod.* 1 juin 2006;21(6):1484-9.
99. Ferguson-Smith MA, Lennox B, Mack WS, Stewart JS. Klinefelter's syndrome; frequency and testicular morphology in relation to nuclear sex. *Lancet.* 27 juill 1957;273(6987):167-9.
 100. Jacobs PA, Strong JA. A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature.* 31 janv 1959;183(4657):302-3.
 101. May-Panloup P, Malinge MC, Larget-Piet L, Chrétien MF. Infertilité masculine d'origine génétique et assistance médicale à la procréation. *Gynécologie Obstétrique & Fertilité.* 1 sept 2001;29(9):583-93.
 102. Piomboni P, Stendardi A, Gambera L. Chromosomal aberrations and aneuploidies of spermatozoa. *Adv Exp Med Biol.* 2014;791:27-52.
 103. Ferlin A, Garolla A, Foresta C. Chromosome abnormalities in sperm of individuals with constitutional sex chromosomal abnormalities. *Cytogenet Genome Res.* 2005;111(3-4):310-6.
 104. Solari AJ. Synaptonemal complex analysis in human male infertility. *Eur J Histochem.* 1999;43(4):265-76.
 105. Sandberg AA, Koepf GF, Ishihara T, Hauschka TS. An XYY human male. *Lancet.* 26 août 1961;2(7200):488-9.
 106. Siffroi JP, Chantot-Bastaraud S, Ravel C. Origines géniques et chromosomiques des anomalies de la spermatogenèse : aspects cliniques et rapports avec les modèles animaux. *Gynécologie Obstétrique & Fertilité.* juin 2003;31(6):504-15.
 107. Perrin A, Morel F, Douet-Guilbert N, Bris M-JL, Amice J, Amice V, et al. A study of meiotic segregation of chromosomes in spermatozoa of translocation carriers using fluorescent in situ hybridisation. *Andrologia.* 2010;42(1):27-34.
 108. Dupuy O, Palou M, Mayaudon H, Sarret D, Bordier L, Garcin JM, et al. [De La Chapelle syndrome]. *Presse Med.* 3 mars 2001;30(8):369-72.
 109. Croft B, Ohnesorg T, Hewitt J, Bowles J, Quinn A, Tan J, et al. Human sex reversal is caused by duplication or deletion of core enhancers upstream of SOX9. *Nat Commun.* 14 2018;9(1):5319.
 110. Bhagavath B, Layman LC, Ullmann R, Shen Y, Ha K, Rehman K, et al. Familial 46,XY sex reversal without campomelic dysplasia caused by a deletion upstream of the SOX9 gene. *Mol Cell Endocrinol.* 5 août 2014;393(1-2):1-7.
 111. Benko S, Fantes JA, Amiel J, Kleinjan D-J, Thomas S, Ramsay J, et al. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nat Genet.* mars 2009;41(3):359-64.
 112. Krausz C, Giachini C, Lo Giacco D, Daguin F, Chianese C, Ars E, et al. High Resolution X Chromosome-Specific Array-CGH Detects New CNVs in Infertile Males.

PLoS One [Internet]. 9 oct 2012 [cité 16 août 2020];7(10). Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3467283/>

113. Jorgez CJ, Wilken N, Addai JB, Newberg J, Vangapandu HV, Pastuszak AW, et al. Genomic and genetic variation in E2F transcription factor-1 in men with nonobstructive azoospermia. *Fertil Steril*. janv 2015;103(1):44-52.e1.
114. Coutton C, Abada F, Karaouzene T, Sanlaville D, Satre V, Lunardi J, et al. Fine characterisation of a recombination hotspot at the DPY19L2 locus and resolution of the paradoxical excess of duplications over deletions in the general population. *PLoS Genet*. mars 2013;9(3):e1003363.
115. Tiepolo L, Zuffardi O. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Hum Genet*. 28 oct 1976;34(2):119-24.
116. Fan Y, Silber SJ. Y Chromosome Infertility. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJ, Stephens K, et al., éditeurs. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 1993 [cité 27 sept 2020]. Disponible sur: <http://www.ncbi.nlm.nih.gov/books/NBK1339/>
117. Ravel C, Chantot-Bastaraud S, McElreavey K, Siffroi J-P. Polymorphismes du chromosome Y et fertilité masculine. *Gynécologie Obstétrique & Fertilité*. 1 oct 2006;34(10):885-93.
118. Chen H, Ruan YC, Xu WM, Chen J, Chan HC. Regulation of male fertility by CFTR and implications in male infertility. *Hum Reprod Update*. 1 nov 2012;18(6):703-13.
119. Tomaiuolo R, Fausto M, Elce A, Strina I, Ranieri A, Amato F, et al. Enhanced frequency of CFTR gene variants in couples who are candidates for assisted reproductive technology treatment. *Clin Chem Lab Med*. août 2011;49(8):1289-93.
120. Sironen A, Shoemark A, Patel M, Loebinger MR, Mitchison HM. Sperm defects in primary ciliary dyskinesia and related causes of male infertility. *Cell Mol Life Sci*. 28 nov 2019;
121. Oud MS, Volozonoka L, Smits RM, Vissers LELM, Ramos L, Veltman JA. A systematic review and standardized clinical validity assessment of male infertility genes. *Hum Reprod*. 1 mai 2019;34(5):932-41.
122. *Surveillance Des Anomalies Congnïtales Un Manuel Pour Les Administrateurs De Programme*. World Health Organization; 2016.
123. Spranger J, Benirschke K, Hall JG, Lenz W, Lowry RB, Opitz JM, et al. Errors of morphogenesis: concepts and terms. Recommendations of an international working group. *J Pediatr*. janv 1982;100(1):160-5.
124. Orphanet: Syndrome de Pierre Robin isolé [Internet]. [cité 4 oct 2020]. Disponible sur: [https://www.orpha.net/consor/cgi-bin/Disease_Search.php?lng=FR&data_id=562&Disease_Disease_Search_diseaseGroup=pierre-robin&Disease_Disease_Search_diseaseType=Pat&Maladie\(s\)/groupes%20de%20mala](https://www.orpha.net/consor/cgi-bin/Disease_Search.php?lng=FR&data_id=562&Disease_Disease_Search_diseaseGroup=pierre-robin&Disease_Disease_Search_diseaseType=Pat&Maladie(s)/groupes%20de%20mala)

dies=Syndrome-de-Pierre-Robin-
isole&title=Syndrome%20de%20Pierre%20Robin%20isol%E9&search=Disease_Searc
h_Simple

125. Francine R, Pascale S, Aline la H. Congenital Anomalies: Prevalence and Risk Factors. *Universal Journal of Public Health*. févr 2014;2(2):58-63.
126. Oliveira CIF, Fett-Conte AC. Birth defects: Risk factors and consequences. *J Pediatr Genet*. juin 2013;2(2):85-90.
127. E A, Dm C, T S, Mg B. The 7q11.23 Microduplication Syndrome: A Clinical Report with Review of Literature. *J Pediatr Genet*. 15 juin 2016;5(3):129-40.
128. Weise A, Mrasek K, Klein E, Mulatinho M, Llerena JC, Hardekopf D, et al. Microdeletion and Microduplication Syndromes. *J Histochem Cytochem*. mai 2012;60(5):346-58.
129. Orphanet [Internet]. [cité 13 sept 2020]. Disponible sur: <https://www.orpha.net/consor/cgi-bin/index.php>
130. Mwaniki MK, Atieno M, Lawn JE, Newton CRJC. Long-term neurodevelopmental outcomes after intrauterine and neonatal insults: a systematic review. *Lancet*. 4 févr 2012;379(9814):445-52.
131. Benoist G, Jacquemard F, Leruez-Ville M, Ville Y. [Cytomegalovirus (CMV) congenital infection]. *Gynecol Obstet Fertil*. mars 2008;36(3):248-60.
132. Liu Y, McDermott S, Lawson A, Aelion CM. The relationship between mental retardation and developmental delays in children and the levels of arsenic, mercury and lead in soil samples taken near their mother's residence during pregnancy. *Int J Hyg Environ Health*. mars 2010;213(2):116-23.
133. Mendola P, Selevan SG, Gutter S, Rice D. Environmental factors associated with a spectrum of neurodevelopmental deficits. *Ment Retard Dev Disabil Res Rev*. 2002;8(3):188-97.
134. Lamoril J, Ameziane N, Deybach J-C, Bouizegarène P, Bogard M. Les techniques de séquençage de l'ADN : une révolution en marche. Première partie. *Immuno-analyse & Biologie Spécialisée*. oct 2008;23(5):260-79.
135. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bembem LA, et al. Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. *Nature*. 15 sept 2005;437(7057):376-80.
136. Metzker ML. Sequencing technologies — the next generation. *Nat Rev Genet*. janv 2010;11(1):31-46.
137. Loman NJ, Constantinidou C, Chan JZM, Halachev M, Sergeant M, Penn CW, et al. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol*. sept 2012;10(9):599-606.

138. Bras J, Guerreiro R, Hardy J. Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. *Nat Rev Neurosci.* 20 juin 2012;13(7):453-64.
139. tony,abm. NGS - Experimental Design [Internet]. [cité 6 sept 2020]. Disponible sur: [/marketing/knowledge_base/next_generation_sequencing_experimental_design](#)
140. Yoon S, Xuan Z, Makarov V, Ye K, Sebat J. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res.* sept 2009;19(9):1586-92.
141. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, et al. Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. *Science.* 19 oct 2007;318(5849):420-6.
142. Xi R, Kim T-M, Park PJ. Detecting structural variations in the human genome using next generation sequencing. *Brief Funct Genomics.* déc 2010;9(5-6):405-15.
143. Figure 1: Four main methods for detecting CNVs with NGS data: (1)... [Internet]. ResearchGate. [cité 31 oct 2020]. Disponible sur: https://www.researchgate.net/figure/Four-main-methods-for-detecting-CNVs-with-NGS-data-1-Read-pair-RP-2-Split-read_fig1_275248075
144. Nielsen J, Wohler M. Chromosome abnormalities found among 34,910 newborn children: results from a 13-year incidence study in Aarhus, Denmark. *Hum Genet.* mai 1991;87(1):81-3.
145. Schluth-Bolard C, Delobel B, Sanlaville D, Boute O, Cuisset J-M, Sukno S, et al. Cryptic genomic imbalances in de novo and inherited apparently balanced chromosomal rearrangements: array CGH study of 47 unrelated cases. *Eur J Med Genet.* oct 2009;52(5):291-6.
146. 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.* mars 2013;50(3):144-50.
147. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society).* 2009;172(1):137-59.
148. Frühmesser A, Blake J, Haberlandt E, Baying B, Raeder B, Runz H, et al. Disruption of EXOC6B in a patient with developmental delay, epilepsy, and a de novo balanced t(2;8) translocation. *Eur J Hum Genet.* oct 2013;21(10):1177-80.
149. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, et al. Paired-end mapping reveals extensive structural variation in the human genome. *Science.* 19 oct 2007;318(5849):420-6.
150. Home | Integrative Genomics Viewer [Internet]. [cité 30 août 2020]. Disponible sur: <http://software.broadinstitute.org/software/igv/>

151. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods*. sept 2009;6(9):677-81.
152. Zhu M, Need AC, Han Y, Ge D, Maia JM, Zhu Q, et al. Using ERDS to infer copy-number variants in high-coverage genomes. *Am J Hum Genet*. 7 sept 2012;91(3):408-21.
153. Projects · NGS / Svagga [Internet]. GitLab. <https://gitlab.inria.fr/NGS/svagga>
154. 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.
155. UCSC Genome Browser Home [Internet]. [cité 30 août 2020]. Disponible sur: <https://genome.ucsc.edu/>
156. A framework for variation discovery and genotyping using next-generation DNA sequencing data | Nature Genetics [Internet]. [cité 11 sept 2020]. Disponible sur: <https://www.nature.com/articles/ng.806>
157. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods*. déc 2001;25(4):402-8.
158. 3D Genome browser: visualize Hi-C, ChIA-PET, HiChIP, PLAC-Seq, Capture Hi-C data. [Internet]. [cité 1 nov 2020]. Disponible sur: <http://promoter.bx.psu.edu/hi-c/>
159. Krijger PHL, de Laat W. Regulation of disease-associated gene expression in the 3D genome. *Nat Rev Mol Cell Biol*. 2016;17(12):771-82.
160. Levy S, Allerston CK, Liveanu V, Habib MR, Gileadi O, Schuster G. Identification of LACTB2, a metallo- β -lactamase protein, as a human mitochondrial endoribonuclease. *Nucleic Acids Res*. 29 févr 2016;44(4):1813-32.
161. Sollars VE, McEntee BJ, Engiles JB, Rothstein JL, Buchberg AM. A novel transgenic line of mice exhibiting autosomal recessive male-specific lethality and non-alcoholic fatty liver disease. *Hum Mol Genet*. 15 oct 2002;11(22):2777-86.
162. Fliegauf M, Olbrich H, Horvath J, Wildhaber JH, Zariwala MA, Kennedy M, et al. Mislocalization of DNAH5 and DNAH9 in Respiratory Cells from Patients with Primary Ciliary Dyskinesia. *Am J Respir Crit Care Med*. 15 juin 2005;171(12):1343-9.
163. Whitfield M, Thomas L, Bequignon E, Schmitt A, Stouvenel L, Montantin G, et al. Mutations in DNAH17, Encoding a Sperm-Specific Axonemal Outer Dynein Arm Heavy Chain, Cause Isolated Male Infertility Due to Asthenozoospermia. *Am J Hum Genet*. 3 juill 2019;105(1):198-212.
164. López Herráez D, Martínez-Bueno M, Riba L, García de la Torre I, Sacnún M, Goñi M, et al. Rheumatoid arthritis in Latin Americans enriched for Amerindian ancestry is associated with loci in chromosomes 1, 12, and 13, and the HLA class II region. *Arthritis Rheum*. juin 2013;65(6):1457-67.

165. Sarkozy A, Digilio MC, Dallapiccola B. Leopard syndrome. *Orphanet J Rare Dis.* 27 mai 2008;3:13.
166. Digilio MC, Sarkozy A, de Zorzi A, Pacileo G, Limongelli G, Mingarelli R, et al. LEOPARD syndrome: clinical diagnosis in the first year of life. *Am J Med Genet A.* 1 avr 2006;140(7):740-6.
167. Shchelochkov OA, Patel A, Weissenberger GM, Chinault AC, Wiszniewska J, Fernandes PH, et al. Duplication of chromosome band 12q24.11q24.23 results in apparent Noonan syndrome. *Am J Med Genet A.* 15 avr 2008;146A(8):1042-8.
168. Geckinli BB, Aydin H, Karaman A, Delil K, Simsek H, Gokmeydan E, et al. CLINICAL REPORT OF A PATIENT WITH DE NOVO TRISOMY 12q23.1q24.33. *Genet Couns.* 2015;26(4):393-400.
169. Stanford SM, Maestre MF, Campbell AM, Bartok B, Kiosses WB, Boyle DL, et al. Protein tyrosine phosphatase expression profile of rheumatoid arthritis fibroblast-like synoviocytes: a novel role of SH2 domain-containing phosphatase 2 as a modulator of invasion and survival. *Arthritis Rheum.* mai 2013;65(5):1171-80.
170. Caubit X, Gubellini P, Andrieux J, Roubertoux PL, Metwaly M, Jacq B, et al. TSHZ3 deletion causes an autism syndrome and defects in cortical projection neurons. *Nat Genet.* 2016;48(11):1359-69.
171. Faralli H, Martin E, Coré N, Liu Q-C, Filippi P, Dilworth FJ, et al. Teashirt-3, a novel regulator of muscle differentiation, associates with BRG1-associated factor 57 (BAF57) to inhibit myogenin gene expression. *J Biol Chem.* 1 juill 2011;286(26):23498-510.
172. Meynert AM, Bicknell LS, Hurles ME, Jackson AP, Taylor MS. Quantifying single nucleotide variant detection sensitivity in exome sequencing. *BMC Bioinformatics.* 18 juin 2013;14(1):195.
173. Waggoner DJ, Raca G, Welch K, Dempsey M, Anderes E, Ostrovskaya I, et al. NSD1 analysis for Sotos syndrome: insights and perspectives from the clinical laboratory. *Genetics in Medicine: Official Journal of the American College of Medical Genetics.* oct 2005;7(8):524-33.
174. Kwaśnicka-Crawford DA, Carson AR, Scherer SW. IQCJ-SCHIP1, a novel fusion transcript encoding a calmodulin-binding IQ motif protein. *Biochem Biophys Res Commun.* 1 déc 2006;350(4):890-9.
175. Brunetti-Pierrri N, Berg JS, Scaglia F, Belmont J, Bacino CA, Sahoo T, et al. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet.* déc 2008;40(12):1466-71.
176. Chiang C, Jacobsen JC, Ernst C, Hanscom C, Heilbut A, Blumenthal I, et al. Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. *Nat Genet.* 4 mars 2012;44(4):390-S1.

177. Donnelly DE, Turnpenny P, McConnell VPM. Phenotypic variability in a three-generation Northern Irish family with Sotos syndrome. *Clin Dysmorphol.* oct 2011;20(4):175-81.
178. Search NCBI databases [Internet]. NCBI. [cité 25 oct 2020]. Disponible sur: <https://www.ncbi.nlm.nih.gov/search/all/>
179. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* 15 févr 2001;409(6822):860-921.
180. Lappalainen T, Scott AJ, Brandt M, Hall IM. Genomic analysis in the age of human genome sequencing. *Cell.* 21 mars 2019;177(1):70-84.
181. Mate Pair Sequencing [Internet]. [cité 25 oct 2020]. Disponible sur: <https://www.illumina.com/science/technology/next-generation-sequencing/mate-pair-sequencing.html>
182. Redin C, Brand H, Collins RL, Kammin T, Mitchell E, Hodge JC, et al. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat Genet.* 2017;49(1):36-45.
183. Uguen K, Jubin C, Duffourd Y, Bardel C, Malan V, Dupont J-M, et al. Genome sequencing in cytogenetics: Comparison of short-read and linked-read approaches for germline structural variant detection and characterization. *Mol Genet Genomic Med.* 2020;8(3):e1114.
184. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nature Reviews Genetics.* oct 2020;21(10):597-614.
185. Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol.* oct 2019;37(10):1155-62.
186. Mantere T, Neveling K, Pebrel-Richard C, Benoist M, van der Zande G, Kater-Baats E, et al. Next generation cytogenetics: genome-imaging enables comprehensive structural variant detection for 100 constitutional chromosomal aberrations in 85 samples [Internet]. *Genetics*; 2020 juill [cité 25 oct 2020]. Disponible sur: <http://biorxiv.org/lookup/doi/10.1101/2020.07.15.205245>
187. Higgins AW, Alkuraya FS, Bosco AF, Brown KK, Bruns GAP, Donovan DJ, et al. Characterization of apparently balanced chromosomal rearrangements from the developmental genome anatomy project. *Am J Hum Genet.* mars 2008;82(3):712-22.
188. Chromothripsis in Healthy Individuals Affects Multiple Protein-Coding Genes and Can Result in Severe Congenital Abnormalities in Offspring: *The American Journal of Human Genetics* [Internet]. [cité 10 nov 2020]. Disponible sur: [https://www.cell.com/ajhg/fulltext/S0002-9297\(15\)00058-0](https://www.cell.com/ajhg/fulltext/S0002-9297(15)00058-0)
189. 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.* 21 mai 2015;161(5):1012-25.

190. Harewood L, Kishore K, Eldridge MD, Wingett S, Pearson D, Schoenfelder S, et al. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. *Genome Biol.* 27 2017;18(1):125.
191. Dieterich K, Zouari R, Harbuz R, Vialard F, Martinez D, Bellayou H, et al. The Aurora Kinase C c.144delC mutation causes meiosis I arrest in men and is frequent in the North African population. *Hum Mol Genet.* 1 avr 2009;18(7):1301-9.